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

**Part 2:
Interpretation and application of
precision data in relation to methods
of test**

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

*Partie 2: Application des valeurs de fidélité relatives aux méthodes
d'essai*



STANDARDSISO.COM : Click to view the full PDF of ISO 4259-2:2017



COPYRIGHT PROTECTED DOCUMENT

© ISO 2017, Published in Switzerland

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

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	iv
Introduction	v
1 Scope	1
2 Normative references	1
3 Terms and definitions	1
4 Application and significance of repeatability, r, and reproducibility, R	2
4.1 General	2
4.2 Repeatability, r	2
4.2.1 General	2
4.2.2 Acceptability of results	2
4.2.3 Confidence limits calculations using results collected under repeatability conditions	3
4.3 Reproducibility, R	3
4.3.1 Acceptability of results	3
4.3.2 Confidence limits calculations using results collected under reproducibility conditions	5
4.4 Use of reproducibility to determine bias between two different test methods that purport to measure the same property	5
4.4.1 General	5
4.4.2 Process	5
5 Specifications	6
5.1 Aim of specifications	6
5.2 Construction of specifications limits in relation to scope and precision of the specified test method	6
6 Assessment of quality conformance to specification	7
6.1 General	7
6.2 Assessment of quality conformance by the supplier	8
6.3 Assessment of quality conformance by the recipient	9
6.3.1 General	9
6.3.2 Single batch of product	9
6.3.3 Multiple batches of product	9
6.3.4 Procedure for recipient to assess conformance for a single batch of product	10
7 Dispute procedure	11
7.1 Resolve dispute by negotiation	11
7.2 Use of the test method or procedure in case of dispute	11
7.3 Dispute resolution procedure	12
7.4 Dispute unresolved	12
7.5 Example of a dispute resolution	14
Annex A (informative) Explanation of formulae given in Clause 4	15
Annex B (informative) Dispute resolution for specifications based on a specified degree of criticality	18
Annex C (informative) Statistical control in the execution of test methods by a laboratory	21
Annex D (informative) General approach to bias assessment using multiple materials	23
Annex E (informative) Glossary	24
Bibliography	25

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 4259-2, together with ISO 4259-1, cancels and replaces ISO 4259, which has been technically revised. This document provides the content of Clauses 7 to 10 of ISO 4259 and connected Annexes H and I. The remaining Clauses and [Annexes A](#) to G of ISO 4259:2006 are replaced by ISO 4259-1.

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

Introduction

For purposes of setting product specifications, and to check product compliance against these specifications, standard test methods are usually referenced for specific properties of commercial petroleum and related products. 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 when setting product specifications, assessing if the differences between test results are within statistical expectation, and making specification compliance decisions based on limited test results. By using statistically-based estimates of the precision for a test method, the following can be achieved:

- an objective measure of the reliability of specification limits,
- a specification compliance decision, and
- the degree of agreement expected between two or more results obtained in specified circumstances.

This document describes the applications of the precision of test method as derived from ISO 4259-1. It is intended to be a companion document to ISO 4259-1. Additional normative and informative discussions on how to use this precision to assess the “in statistical control” status and precision capability of a specific laboratory in the execution of a test method are provided. Also, the general approach to the agreement between two different test methods that purport to measure the same property are given.

The two parts of ISO 4259 encompass both the determination of precision estimates and the application of precision data. It attempts to be aligned with ASTM D6300^[1] regarding the determination of the precision estimates and with ASTM D3244^[2] for the utilization of test data.

A glossary of the variables used in this document and ISO 4259-1 is included in ISO 4259-1:2017, Annex I.

STANDARDSISO.COM : Click to view the full PDF of ISO 4259-2:2017

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

Part 2:

Interpretation and application of precision data in relation to methods of test

1 Scope

This document specifies the methodology for the application of precision estimates of a test method derived from ISO 4259-1. In particular, it defines the procedures for setting the property specification limits based upon test method precision where the property is determined using a specific test method, and in determining the specification conformance status when there are conflicting results between supplier and receiver. Other applications of this test method precision are briefly described in principle without the associated procedures.

The procedures in this document have been designed specifically for petroleum and petroleum-related products, which are normally 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 4259-1, *Petroleum and related products — Precision of measurement methods and results — Part 1: Determination of precision data in relation to methods of test*

3 Terms and definitions

For the purposes of this document, the terms and definitions in ISO 4259-1 and the following apply.

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

- ISO Online browsing platform: available at <https://www.iso.org/obp>
- IEC Electropedia: available at <http://www.electropedia.org/>

3.1

proficiency testing program

PTP

program designed for the periodic evaluation of participating laboratories' testing capability of a Standard Test Method through the statistical analysis of their test results obtained on aliquots prepared from a single batch of homogeneous material

Note 1 to entry: The frequency of such testing varies in accordance with the program objective. Each execution of testing involves testing of a single batch of material. Materials typically vary from test to test.

Note 2 to entry: This is also commonly referred to as Inter Laboratory Cross Check Program (ILCP).

3.2
recipient

individual or organization who receives or accepts a product delivered by the supplier

3.3
supplier

individual or organization responsible for the quality of a product just before it is taken over by the recipient

4 Application and significance of repeatability, r , and reproducibility, R

4.1 General

The value of these quantities is estimated from analysis of variance (two-factor with replication) performed on the results obtained in a statistically designed inter-laboratory programme in which different laboratories each test a range of samples. Repeatability and reproducibility values estimated in accordance with ISO 4259-1 or other statistical techniques shall be included in each published test method.

NOTE See [Annex A](#) for an account of the statistical reasoning underlying the formulae in this clause.

In the following clauses, it is assumed that the result(s) are obtained from a test method that is in statistical control. For determination of “in statistical control”, see [Annex C](#).

4.2 Repeatability, r

4.2.1 General

Most laboratories do not carry out more than one test on each sample for routine quality control purposes except in some circumstances, such as in cases of dispute or if the test operator wishes to confirm that his technique is satisfactory. In such circumstances, when multiple results are obtained, it is useful to check the consistency of repeated results against the repeatability of the method. The appropriate procedure is outlined in [4.2.2](#). It is also useful to know what degree of confidence may be placed on the average results, and the method of determining this is given in [4.2.3](#).

4.2.2 Acceptability of results

When only two results are obtained under repeatability conditions and their difference is less than or equal to r , the test operator may consider his work as being under control and may take the average of the two results as the estimated value of the property being tested.

If the two results differ by more than r , both shall be considered as suspect and at least three more results obtained. Including the first two, the difference between the most divergent result and the average of the remainder shall then be calculated and this difference compared with a new value, r_1 , instead of r , given in [Formula \(1\)](#):

$$r_1 = r \sqrt{\frac{k}{2(k-1)}} \quad (1)$$

where k is the total number of results obtained.

If the difference is less than or equal to r_1 , all the results shall be accepted. If the difference exceeds r_1 , the most divergent result shall be rejected and the procedure specified in this section repeated until an acceptable set of results is obtained.

The average of the acceptable results shall be taken as the estimated value of the property. However, if two or more results from a total of not more than 20 have been rejected, the operating procedure and the apparatus shall be checked and a new series of tests made, if possible.

4.2.3 Confidence limits calculations using results collected under repeatability conditions

When a single test operator, who is working within the precision limits of the method, obtains a series of k results under repeatability conditions, giving an average, \bar{X} , and the results meet the repeatability requirement in 4.2.2, it can be assumed with 95 % confidence that the true value, μ , of the characteristic lies within the following limits:

$$\left(\bar{X} - \frac{R_1}{\sqrt{2}}\right) \leq \mu \leq \left(\bar{X} + \frac{R_1}{\sqrt{2}}\right) \quad (2)$$

where

$$R_1 = \sqrt{R^2 - r^2 \left(1 - \frac{1}{k}\right)} \quad (3)$$

When $k = 1$, use the single test result as the value for the \bar{X} term as follows:

$$\left(X - \frac{R}{\sqrt{2}}\right) \leq \mu \leq \left(X + \frac{R}{\sqrt{2}}\right) \quad (4)$$

where R is the published test method reproducibility as discussed in 4.3.

Similarly, for the single limit situation, when only one limit is fixed (upper or lower), it can be assumed with 95 % confidence that the true value, μ , of the characteristic is limited as follows:

$$\mu \leq \left(\bar{X} + 0,59R_1\right) \quad (\text{upper limit}) \quad (5)$$

or

$$\mu \geq \left(\bar{X} - 0,59R_1\right) \quad (\text{lower limit}) \quad (6)$$

The factor 0,59 is the ratio $0,84/\sqrt{2}$, where 0,84 is derived in Annex A.

When r is much smaller than R , little improvement in the precision of the average is obtained by carrying out multiple testing under repeatability conditions.

4.3 Reproducibility, R

4.3.1 Acceptability of results

The procedure specified in 4.3 is intended for judging the acceptability, with respect to the reproducibility of the test method, of results obtained by different laboratories in normal, day-to-day operations and transactions. In cases of dispute between a supplier and a recipient, the procedure specified in Clauses 5 to 7 shall be adopted.

When single results are obtained in two laboratories and their difference is less than or equal to R , the two results shall be considered as acceptable, and used to calculate the average \bar{X} . The average \bar{X} , rather than either single result separately, shall be used as the estimated value of the tested property.

The true value μ of the characteristic is contained within the following limits with a 95 % confidence:

$$\left(\bar{X} - \frac{R}{2}\right) \leq \mu \leq \left(\bar{X} + \frac{R}{2}\right) \quad (7)$$

Similarly for the single limit situation, when only one limit is fixed (upper or lower), the true value μ of the characteristic is contained with the following limits with a 95 % confidence:

$$\mu \leq (\bar{X} + 0,42R) \quad (\text{upper limit}) \quad (8)$$

or

$$\mu \geq (\bar{X} - 0,42R) \quad (\text{lower limit}) \quad (9)$$

The factor 0,42 is the ratio of $0,59/\sqrt{2}$ as it is an average of two results.

If the two results differ by more than R , both shall be considered as suspect. Each laboratory shall then obtain at least three other acceptable results (see 4.2.2).

In this case, the difference between the averages of all acceptable results of each laboratory shall be judged for conformity using a new value, R_2 , instead of R , as given by Formula (10):

$$R_2 = \sqrt{R^2 - r^2 \left(1 - \frac{1}{2k_1} - \frac{1}{2k_2}\right)} \quad (10)$$

where

R is the reproducibility of the method;

r is the repeatability of the method;

k_1 is the number of results of the first laboratory;

k_2 is the number of results of the second laboratory.

If the difference between the averages is less than or equal to R_2 , then these averages are acceptable and their overall average shall be considered as the estimated value of the tested property. If the difference between the averages is greater than R_2 , and there is a dispute on the specification conformance of the tested property, then the procedure specified in Clause 7 shall be adopted.

If circumstances arise in which there are more than two laboratories, each supplying one or more acceptable results, the difference between the most divergent laboratory average and the average of the remaining N laboratory averages shall be compared to R_3 :

where

$$R_3 = \sqrt{\frac{R_1^2}{2} + \frac{R_4^2}{2N}} \quad (11)$$

$$R_4 = \sqrt{R^2 - \frac{r^2}{N} \left(N - \frac{1}{k_1} - \frac{1}{k_2} - \dots - \frac{1}{k_N}\right)} \quad (12)$$

R_1 is given in Formula (3), and corresponds to the most divergent laboratory average.

If this difference is equal to or less than R_3 in absolute value, all results shall be regarded as acceptable and their average taken as the estimated value of the property.

If the difference is greater than R_3 , the most divergent laboratory average shall be rejected and the comparison using [Formulae \(11\)](#) and [\(12\)](#) repeated until an acceptable set of laboratory averages is obtained. The average of these laboratory averages shall be taken as the estimated value of the property. However, if two or more laboratory averages from a total of not more than 20 have been rejected, the operating procedure and the apparatus shall be checked and a new series of tests made, if possible.

4.3.2 Confidence limits calculations using results collected under reproducibility conditions

When N laboratories obtain one or more results under conditions of repeatability and reproducibility, giving an average of laboratory averages \bar{X} , the true value μ of the characteristic is contained within the following limits with 95 % confidence:

$$\bar{X} - \frac{R_4}{\sqrt{2N}} \leq \mu \leq \bar{X} + \frac{R_4}{\sqrt{2N}} \quad (13)$$

Similarly for the single limit situation, when only one limit is fixed (upper or lower), the true value μ of the characteristic is contained with the following limits with 95 % confidence:

$$\mu \leq \left(\bar{X} + 0,59 \frac{R_4}{\sqrt{N}} \right) \quad (\text{upper limit}) \quad (14)$$

or

$$\mu \geq \left(\bar{X} - 0,59 \frac{R_4}{\sqrt{N}} \right) \quad (\text{lower limit}) \quad (15)$$

4.4 Use of reproducibility to determine bias between two different test methods that purport to measure the same property

4.4.1 General

For the situation where two different test methods purport to measure the same property, the reproducibility estimates (R) from the respective test methods shall be used in conjunction with the averages obtained from multiple laboratories for the same material to determine if a bias correction can be applied to improve statistically the agreement between the two methods for that material. For example, results collected through Proficiency Testing Programs (PTP) for different test methods using the same sample can be analysed in this fashion.

NOTE Discussion on methodology for this type of assessment for the simultaneous analysis of multiple materials / property levels that span the intersecting scope of two different test methods is beyond the scope of this document. Interested readers are encouraged to consult ASTM D6708[3] for a detailed presentation on the subject. [Annex D](#) provides a brief overview on the general statistical approach for the aforementioned situation.

4.4.2 Process

Assume that Test Method A and Test Method B are test methods that purport to measure property C.

Calculate the following statistic:

$$Z = \frac{|\bar{Y}_A - \bar{Y}_B|}{\sqrt{\frac{R_A^2}{7,683L_A} + \frac{R_B^2}{7,683L_B}}} \quad (16)$$

where

Y_A is the average from L_i results for property C for a material using Test Method A, where each result is a single result obtained under reproducibility conditions;

L_A is the total number of laboratories (results) for Test Method A and should be >20 ;

R_A is the reproducibility of Test Method A;

Y_B is the average from L_B results for property A using Test Method B on the same material tested by Test Method A, where each result is a single result obtained under reproducibility conditions;

L_B is the total number of laboratories (results) for Test Method B, and should be >20 ;

R_B is the reproducibility of Test Method B.

If $Z > 2$, it shall be concluded, with 95 % confidence, that a constant bias correction statistically improves the degree of agreement between Test Method 1 and Test Method 2 for property C for this material.

5 Specifications

5.1 Aim of specifications

The purpose of a specification is to specify an acceptable limit or limits to the true value, μ , of the property as determined by a specified test method. In practice, however, this true value can never be established exactly since the results obtained by applying the specified test method in a single or multiple laboratories can show acceptable scattering as defined by the repeatability and reproducibility. There is, therefore, always some uncertainty as to the true value of the tested property determined from a finite number of test results.

Petroleum product compliance with specifications is assessed in accordance with [Clauses 6](#) and [7](#). By prior agreement a supplier and recipient may use the alternative procedures described in [Annex B](#).

It is important that the test method specified for the determination of the property governed by the specification limit(s) is sufficiently precise to reliably determine whether or not the product meets the specifications.

5.2 Construction of specifications limits in relation to scope and precision of the specified test method

The specification limits shall not be outside the method scope limits as determined in ISO 4259-1.

The lower specification limit shall not be less than the lower scope limit of the test method, and the upper specification limit shall not be greater than the upper scope limit of the test method (see §6.5 in ISO 4259-1).

In addition, the distance between lower and upper specification limit shall also satisfy the following condition: upper specification limit minus lower specification limit shall not be less than the quantity $2R$ evaluated at lower method scope limit plus $2R$ evaluated at upper method scope limit. See [Figure 1](#) for an illustration of this concept.

Usually, specifications deal with limits for the values of the properties. To avoid uncertainty, such limits are normally expressed as “not less than” or “not greater than”. Limits are of two types:

- a double limit, upper and lower, for example viscosity not less than 5 mm²/s and not greater than 16 mm²/s; boiling point 100 °C ± 0,5 °C;
- a single limit, upper or lower, for example water content not greater than 2 %; sulfur content not greater than 10 mg/kg; solubility of bitumen not less than 99 %.

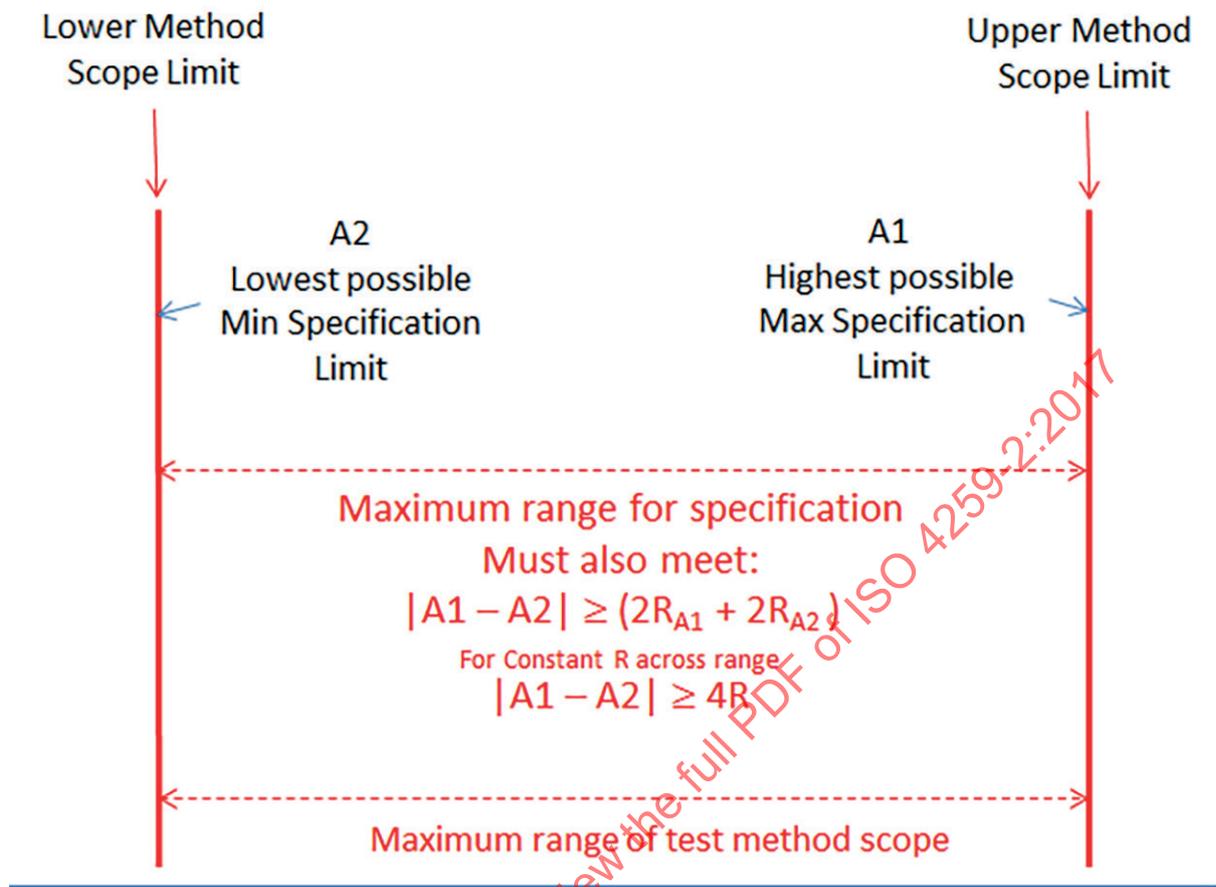


Figure 1 — Specification setting

In cases where, for practical reasons, the value of $(A_1 - A_2)$ is less than the above minimum range requirement in [Figure 1](#), the results obtained will be of doubtful significance in determining whether a sample does or does not satisfy the requirements of the specification. According to statistical reasoning, it is desirable for $(A_1 - A_2)$ to be considerably greater than the above minimum range requirement. If not, one or both of the following courses shall be adopted:

- the specification limits shall be examined to see whether they can be widened to fit in with the precision of the test method;
- the test method shall be examined to see whether the precision can be improved, or an alternative test method adopted with an improved precision, to fit in with the desired specification limits.

Conformity to this document requires specifications to be drawn up in accordance with the above principles.

6 Assessment of quality conformance to specification

6.1 General

6.1.1 [Clause 6](#) provides general information to allow the supplier or the recipient to judge the quality of a product with regard to the specification based on a single test result as obtained by the supplier or recipient. If both the supplier's and recipient's single test results are available, the estimate of the true value shall be obtained in accordance with [4.3](#). Should the recipient decide to dispute the quality

conformance to specification after examining his single result, or the estimate from 4.3, the procedure specified in Clause 7 shall be adopted.

As a prerequisite for acceptance for laboratory test results to be used in 6.2, 6.3, Clause 7 and Annex B, the conditions in 6.1.2 to 6.1.4 shall be satisfied.

6.1.2 Each laboratory's test result shall be obtained from a test method that is in statistical control in terms of precision and bias, as substantiated by in-house SQC charts or other equivalent statistical techniques.

NOTE For a general discussion on the topic of statistical control and control charts, see Annex C.

6.1.3 The standard deviation from the control charts (or equivalent statistical techniques) in 6.1.2, as calculated from at least 30 most recent results obtained over at least 15 days, with results that are separated by at least 6 h, shall not exceed the published test method standard deviation ($R/2,77$).

If evidence exists from the published results of multiple PTP, that the R for a published test method is statistically inconsistent with the R actually achieved, the latter may be used in lieu of the published R to judge conformance to this clause, provided all of the following conditions are met:

- if legally permissible, and
- if the R calculated from multiple PTP have sufficient degrees of freedom (>30) using results that have been properly screened for outliers in accordance with GESD protocol in ISO 4259-1 or other equivalent statistical technique, and
- upon mutual agreement between disputing parties.

6.1.4 Each laboratory shall be able to demonstrate, by way of results from participation in PTPs, if available, a sustained testing proficiency and a lack of bias relative to PTP averages for the appropriate test method(s). In the event that a suitable PTP is not available, proficiency shall be demonstrated by way of testing certified reference materials (CRM) and in-house control charts on quality control (QC) samples, or by other method validation techniques acceptable to both parties.

6.2 Assessment of quality conformance by the supplier

A supplier who has no other source of information on the true value of a characteristic than a single result shall consider, with 95 % confidence, that the product meets the specification limit, only if the result, X_s , is such that

- in the case of a single upper limit, A_1 :

$$X_s \leq A_1 - 0,59R \quad (17)$$

- in the case of a single lower limit, A_2 :

$$X_s \geq A_2 + 0,59R \quad (18)$$

- in the case of a double limit (A_1 and A_2), both these conditions are satisfied (see 4.2.3).

The 95 % confidence decision limits as calculated using Formulae (17) and (18) are for the guidance of the supplier, and are not to be interpreted as an obligation. A reported value between the specification value and the limit from Formula (17) or (18) is not proof of non-compliance, but is an indication that the confidence for the product to meet the specification limit is less than 95 %. If the result is exactly at the specification limit, the probability of a re-test result meeting specification, by either the supplier or the recipient, is 50 %. A direct consequence of releasing product with a low confidence is that the probability of the receiver obtaining an off-specification result will be high.

The supplier shall only release the product if their test result meets specification or by mutual agreement with the customer.

If multiple results are obtained by the supplier under repeatability conditions, the average of the acceptable results and R_1 as determined in 4.2.2 and 4.2.3 shall be used by the supplier as the basis for determination of specification conformance.

In the event of a dispute with the recipient, procedures in [Clause 7](#) shall be followed.

6.3 Assessment of quality conformance by the recipient

6.3.1 General

[Figure 2](#) is a procedure flowchart that describes steps outlined in this subclause, taking into account all available data and requirements of this document.

6.3.2 Single batch of product

A recipient who has no other source of information on the true value of a characteristic than a single result shall consider that the product fails the specification limit with 95 % confidence, only if the result, x , is such that:

- in the case of an upper limit of the specification A_1 ,

$$x > A_1 + 0,59R \quad (19)$$

- in the case of a lower limit of the specification A_2 ,

$$x < A_2 - 0,59R \quad (20)$$

- in the case of a double limit (A_1 and A_2), either of these conditions applies.

The 95 % confidence decision limits as calculated using [Formulae \(19\)](#) and [\(20\)](#) is for the guidance of the recipient, and is not to be interpreted as an obligation. A reported value between the specification value and the limit from [Formula \(19\)](#) or [\(20\)](#) is not proof of non-conformance, but is an indication that the confidence for the product failing specification limit is less than 95 %. If the result is exactly at the specification limit, the probability of a re-test meeting specification, by either the supplier or recipient, is 50 %.

If multiple results are obtained by the recipient under repeatability conditions, the average of the acceptable results and R_1 as determined in 4.2.3 shall be used as the basis for determination of specification conformance.

Should the recipient decide to dispute the specification conformance status for the batch in question, regardless of the recipient's result that was used as a basis for the decision to dispute, the procedures in [Clause 7](#) shall be adhered to.

6.3.3 Multiple batches of product

Persistent results over multiple batches that fail to meet the specification limit, but by an amount not greater than $0,59R$, is a strong indication that the product release confidence by the supplier is less than 95 %. If the latter is not acceptable by the recipient, it is recommended that the recipient contact the supplier and arrive at a mutually satisfactory resolution.

NOTE Five results in a row that fail to meet the specification limit constitute compelling evidence (greater than 95 % confidence) that at least one of the batches does not meet specification.

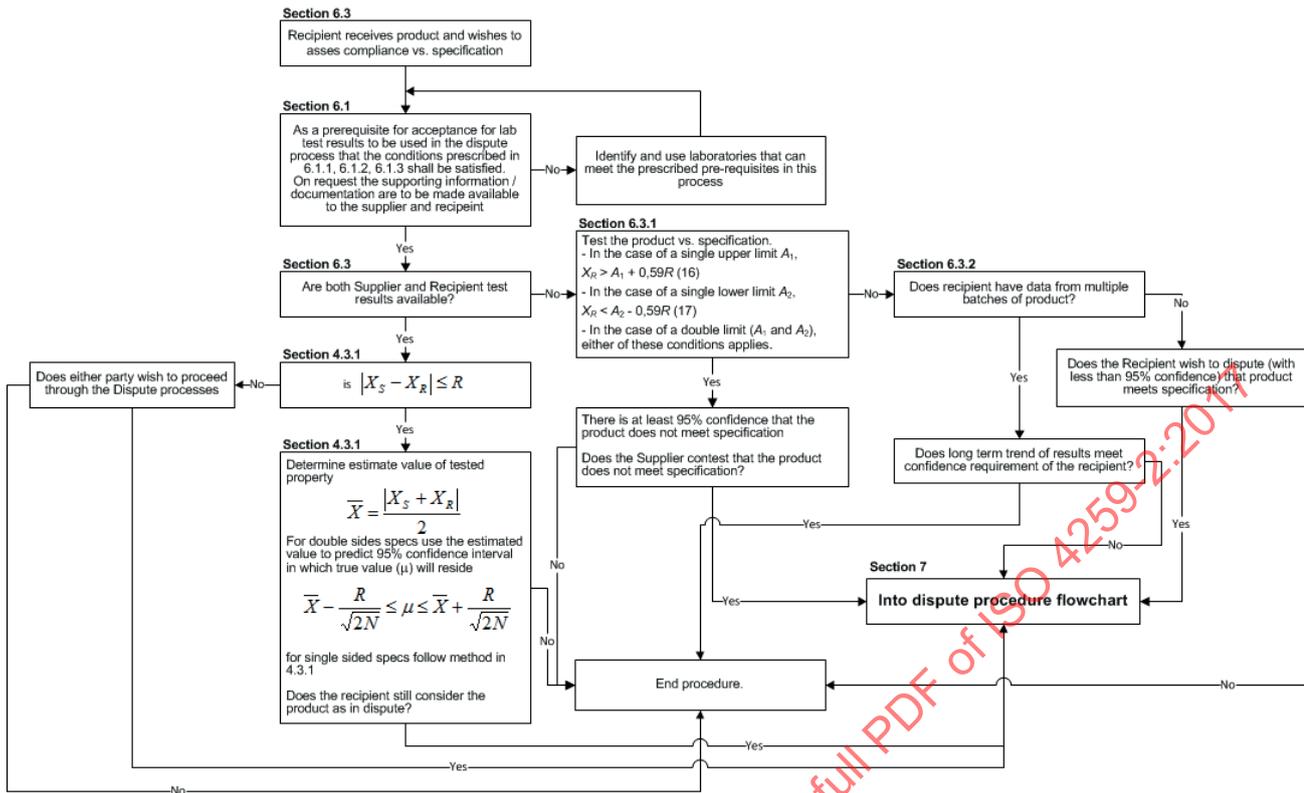


Figure 2 — Flowchart for assessment of specification conformance by recipient

6.3.4 Procedure for recipient to assess conformance for a single batch of product

In the case of assessing the conformance to specification of a single batch of product, the following example shows how to evaluate the process. The Research Octane Number (RON) specification within EN 228[6] tested by ISO 5164[Z]: is used as an example. The reproducibility of RON using this method is 0,7 at the EN 228 specification of 95,0 ON (octane number) and the repeatability at the same level is 0,2 ON.

In this example, a supplier sells a batch of gasoline as compliant with EN 228 after certifying the product at their chosen laboratory as meeting specification. The supplier RON result is 95,1 ON against the specification of 95,0 ON and the batch is sold Free On Board (FOB). The sales contract specifies the representative sample as the shore tank sample which is stored appropriately and in sufficient volume for any follow-up testing.

The recipient purchases the batch and takes a sample to check for quality and the result from their chosen laboratory is determined as 94,7, which is off-specification but within 0,59R. Subclause 6.3.1 indicates that the product fails the specification limit with 95 % confidence only if it exceeds the specification by an amount greater than 0,59R; this is not the case for this example. It is common for the test against 0,59R to be the end of the procedure unless data from multiple batches is available.

In this case, despite the guidance in 6.3.1, the recipient is unhappy with their result and contacts the supplier regarding their result. Both are able to confirm and demonstrate that the results came from laboratories that have in-house SQC programs that are in control with respect to precision and bias for testing using ISO 5164 and that they regularly participate in industry PTP schemes to confirm a lack of bias versus industry averages.

Under this scenario, 4.3.1 is applicable since both the supplier and the recipient results are available.

- a) Acceptability of results states that “when single results are obtained in two laboratories and their difference is less than or equal to R, the two results shall be considered as acceptable, and used to

calculate the average \bar{X} . The average \bar{X} , rather than either single result separately, shall be used as the estimated value of the tested property”.

- b) First test the difference of the two values versus reproducibility to determine if the two results can be considered part of the same population of results.

$|X_S - X_R| \leq R = |95,1 - 94,7| \leq R = 0,4 \leq 0,7$ so in this example both results are considered as valid. If this had not been the case then [4.3.1](#) gives the process to follow rejecting the suspicious results and how to gain acceptable results for comparison.

- c) After passing the comparison versus reproducibility, we then calculate the average result:

$$\bar{X} = (95,1 + 94,7)/2 = 94,9 \text{ ON}$$

- d) Now it is possible to calculate the 95 % confidence limits in which the true value μ lies using [Formula \(7\)](#), [\(8\)](#) or [\(9\)](#) as appropriate.

For a single lower limit, [Formula \(9\)](#) gives:

$$\mu \geq (\bar{X} - 0,42 R), \text{ that is } \mu \geq (94,9 - 0,294) \text{ or } \mu \geq 94,6 \text{ ON}$$

- e) Based on the outcome from d), it can be concluded, with 95 % confidence, that the true value is no worse than 94,6 ON. If the recipient still considers the product as in dispute, see [Clause 7](#).

If only the recipient result is available then [6.3.1](#) guide the comparison on conformance to specification and the confidence limit from [Formula \(9\)](#) in this examples simplifies to $\mu \geq (X - 0,59R)$.

If the recipient had a long-term relationship with the supplier and had data from multiple batches of product, [6.3.2](#) gives guidance on how to assess whether the confidence of product meeting specification is acceptable.

7 Dispute procedure

7.1 Resolve dispute by negotiation

Given the complexity of the procedures outlined in [Clause 7](#) and [Annex B](#), it is highly recommended that the supplier and recipient attempt to resolve the dispute through negotiation to arrive at mutually acceptable terms for settlement. [Figure 3](#) is a flowchart summary of the steps outlined in [Clause 7](#).

In order to engage in this procedure a pre-requisite is that there is sufficient amount of agreed-upon adjudication sample available.

7.2 Use of the test method or procedure in case of dispute

For a specification document that lists multiple test methods for the same property, or, a test method standard where multiple procedures for the same property are given, the designated referee method or procedure shall be given.

NOTE It is the responsibility of the standardization committee supervising either the specification or the test method standard to ensure a referee method or procedure for adjudication purposes is designated in the document.

In the event that a referee method or procedure is not specified, supplier and receiver shall agree on a specific method or procedure prior to execution of the procedures in this clause.

7.3 Dispute resolution procedure

7.3.1 If the supplier and the recipient cannot reach agreement regarding the specification conformance for the quality of the product on the basis of their existing results (see 6.3.3), then the dispute resolution procedures given in either this clause or Annex B shall be adopted. Note that in order to use Annex B, the degree of criticality, p_c , is to be agreed upon in advance between supplier and recipient.

7.3.2 Each laboratory shall reject its original results and obtain at least three other acceptable results on their own check sample to ensure that the work has been carried out correctly under repeatability conditions. The average of the acceptable results in each laboratory shall then be computed, divergent results being discarded as indicated in 4.2.2. If the re-testing does not resolve the dispute, then continue as given below.

Let

\bar{X}_S be the average of the supplier;

\bar{X}_R be the average of recipient;

A_1 be the upper limit of the specification;

A_2 be the lower limit of the specification.

\bar{X}_S and \bar{X}_R shall be compared as follows with A_1 and A_2 :

If $\frac{\bar{X}_S + \bar{X}_R}{2} \leq A_1$ or $\geq A_2$

- product meets specification if $|\bar{X}_S - \bar{X}_R| \leq 0,84R_2$ (for R_2 , see 4.3.1);
- it cannot be stated with confidence whether the product does or does not comply with the specification limit if $|\bar{X}_S - \bar{X}_R| > 0,84R_2$.

In the latter case, since it cannot be stated with confidence whether the product does or does not comply with the specification limit, resolution of the dispute may be achieved by negotiation.

If $\frac{\bar{X}_S + \bar{X}_R}{2} > A_1$ or $< A_2$ dispute whatever the difference $\bar{X}_S - \bar{X}_R$.

7.4 Dispute unresolved

7.4.1 If $|\bar{X}_S - \bar{X}_R| > 0,84R_2$, and the dispute is unresolved, the two laboratories shall contact each other and compare their operating procedures and apparatus.

7.4.2 If the disagreement on product conformance status still remains and resolution of the dispute cannot be reached by negotiation, at least one additional third-party laboratory that is neutral, expert and accepted by the two parties in dispute shall be invited to assist in the dispute resolution. All parties (supplier, recipient, and third-party laboratory) shall first agree on a common adjudication sample. Subject to sample volume adequacy, each party shall obtain their own average to be used for adjudication purposes using at least three acceptable results for the adjudication sample in accordance with 4.2.2. If there is an insufficient amount of adjudication sample available, all parties shall agree to the number of test results that each party is to use to calculate their average to be used for adjudication purposes.

Suppose one additional third-party laboratory is used, and \bar{X}_E is the adjudication result from the third-party laboratory. If the difference between the most divergent laboratory adjudication result and the

average of the two other laboratory adjudication results is less than or equal to R_3 (see 4.3.1) the following procedure shall be adopted.

- If $\frac{\bar{X}_S + \bar{X}_R + \bar{X}_E}{3} \leq A_1$ or $\geq A_2$, product meets specification.
- If $\frac{\bar{X}_S + \bar{X}_R + \bar{X}_E}{3} > A_1$ or $< A_2$, product fails specification.

If more than one additional third-party laboratory is involved, and the difference between the most divergent laboratory adjudication result versus the average of all other laboratories' adjudication results is less than or equal to R_3 (see 4.3.1), then the grand average from all adjudication results shall be used to settle the dispute as follows.

Let

$X_{\text{GRAND AVG}}$ be the grand average calculated from all adjudication results.

If $X_{\text{GRAND AVG}} \leq A_1$ or $\geq A_2$, product meets specification, else product fails specification.

7.4.3 If the difference between the most divergent laboratory adjudication result and the average, \bar{X} , of the other adjudication results is more than R_3 , the following procedure shall be adopted.

- If $\bar{X} \leq A_1$ or $\geq A_2$, product meets specification.
- If $\bar{X} > A_1$ or $< A_2$, product fails specification.

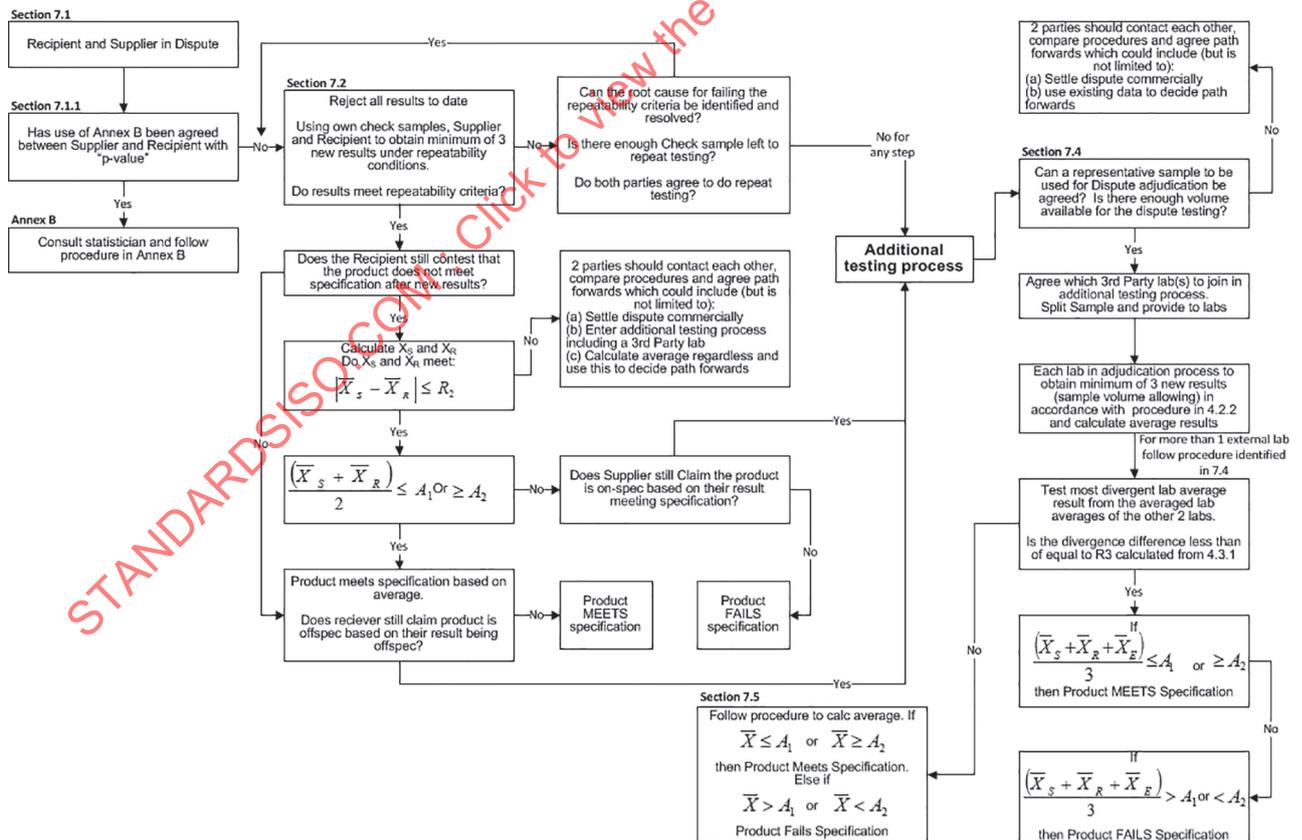


Figure 3 — Flowchart for dispute procedure

7.5 Example of a dispute resolution

In this subclause it is shown how to evaluate the procedure described in 7.1, continuing the earlier example in 6.3.4. That example uses the RON specification within EN 228[6] tested by ISO 5164[7]. In this dispute procedure example, it is assumed that the use of 6.3.4 failed to bring the discussion between supplier and recipient to closure and that the contract did not specify a designated criticality level requiring the use of Annex B.

Having already demonstrated the ability of both laboratories to test, be in statistical control and a lack of persistent bias versus industry as part of 6.3.4, the supplier and recipient enter into the dispute procedure outlined in Clause 7.

All results taken to date as part of the process in 6.3.4 are rejected and assuming there is sufficient volume of the supplier and recipient check samples available, each then takes a minimum of 3 new results under repeatability conditions. Each pair of results taken are compared against method *r*.

The results for the supplier in their testing are 94,9 ON; 95,1 ON and 95,2 giving an average when rounded to the reporting precision plus 1 digit of 95,07 ON (this becomes the \bar{X}_S result).

The results for the recipient in their testing are 94,8 ON; 95,0 ON and 94,9 giving an average when rounded to the reporting precision plus one digit of 94,90 ON (this becomes the \bar{X}_R result).

The differences between the \bar{X}_S and \bar{X}_R averaged results shall meet the reduced reproducibility R_2 which is defined in (21) and shown below.

$$R_2 = \sqrt{R^2 - r^2 \left(1 - \frac{1}{2k_1} - \frac{1}{2k_2} \right)} \tag{21}$$

$$R_2 = \sqrt{0,7^2 - 0,2^2 \left(1 - \frac{1}{2 \times 3} - \frac{1}{2 \times 3} \right)} = 0,68 \text{ ON when rounded to an extra significant figure.}$$

Clearly in this example the difference of 0,17 ON between supplier and recipient averages meets the R_2 criteria and thus the results are accepted in the dispute process. Had this not been the case then the 3 paths forward are outlined in the dispute flow chart.

To determine whether the product met specification, the \bar{X}_S and \bar{X}_R averaged results are averaged and compared against specification. The average of 95,07 ON and 94,90 ON is 94,985 ON which is rounded to the reporting precision of 95,0 ON. This is compared to the specification of 95,0 ON thus passing the test.

$$\frac{(\bar{X}_S + \bar{X}_R)}{2} \leq A_1 \text{ or } \geq A_2$$

In this case the dispute procedure outcome is that there is no dispute as the product meets specification (95,0 ON \geq 95,0 ON).

Had this not been the case, and the receiver refuses to recognize that the product meets specification, then 7.4 outlines the path forward. This would involve agreeing a representative retained sample for additional testing, agreeing an external laboratory (E) to engage in the additional testing process alongside the supplier and recipient nominated laboratories and then repeating the multi testing process on the new sample. The divergence of averaged lab results is tested against a reduced reproducibility R_3 defined in Formula (11).

Annex A (informative)

Explanation of formulae given in [Clause 4](#)

Let σ_0^2 be the component of variance of results obtained under repeatability conditions.

Let σ_1^2 be the component of variance due to interaction between laboratories and samples (errors which contribute to reproducibility errors). Then $(\sigma_0^2 + \sigma_1^2)$ is the variance of results obtained under reproducibility conditions.

r is calculated as

$$Z\sqrt{2}\sigma_0 \tag{A.1}$$

and R is calculated as

$$Z\sqrt{2}\sqrt{\sigma_0^2 + \sigma_1^2} \tag{A.2}$$

where Z is the factor^[4] for converting a standard deviation to a confidence limit (see [Table A.1](#)), and which corresponds in this case to a double-sided 95 % probability level, having a value of 1,96.

Table A.1 — Critical values of the normal distribution

p	0,70	0,80	0,90	0,95	0,975	0,99	0,995
Z	0,524	0,842	1,282	1,645	1,960	2,326	2,576
$2(1 - p)$	0,60	0,40	0,20	0,10	0,05	0,02	0,01
NOTE This table is reproduced from ISO 4259-1.							

Furthermore, the variance of means of k results obtained under repeatability conditions is σ_0^2/k .

In a set of k such results, therefore, the variance of the difference between a single result and the mean of the remainder is:

$$\sigma_0^2 + \frac{\sigma_0^2}{(k-1)} = \frac{r^2}{1,96^2} \times \frac{k}{2(k-1)} \tag{A.3}$$

and the 95 % confidence limit for the absolute difference is:

$$r_1 = 1,96 \sqrt{\sigma_0^2 + \frac{\sigma_0^2}{k-1}} = r \sqrt{\frac{k}{2(k-1)}} \tag{A.4}$$

If the mean of k results is obtained in each of several laboratories, these laboratory means have a variance:

$$\frac{\sigma_0^2}{k} + \sigma_1^2 = \frac{1}{2} \times \frac{1}{1,96^2} \left[R^2 - r^2 \left(1 - \frac{1}{k} \right) \right] \tag{A.5}$$

Let

$$R_1 = \sqrt{R^2 - r^2 \left(1 - \frac{1}{k} \right)} \tag{A.6}$$

The double-sided 95 % confidence limits for such means are:

$$\text{mean} \pm 1,96 \sqrt{\frac{\sigma_0^2}{k} + \sigma_1^2} = \text{mean} \pm \frac{R_1}{\sqrt{2}} \tag{A.7}$$

Confidence limits for probability levels other than 95 % may be calculated by selecting the appropriate Z value from [Table A.1](#) (with single- or double-sided probability as required) and multiplying by the conversion factor $Z/1,96$. For a single-sided probability of 95 %, $Z = 1,64$ and the conversion factor is 0,84.

In general, N laboratories obtain average results from k_1, k_2, \dots, k_N results, respectively. The variance of the average of N such laboratory averages is:

$$\frac{1}{N^2} \left[\left(\frac{\sigma_0^2}{k_1} + \sigma_1^2 \right) + \dots + \left(\frac{\sigma_0^2}{k_N} + \sigma_1^2 \right) \right] = \frac{1}{2N1,96^2} \left[R^2 - \frac{r^2}{N} \left(N - \frac{1}{k_1} - \dots - \frac{1}{k_N} \right) \right] \tag{A.8}$$

Let

$$R_4 = \sqrt{R^2 - \frac{r^2}{N} \left(N - \frac{1}{k_1} - \dots - \frac{1}{k_N} \right)} \tag{A.9}$$

The double-sided 95 % confidence limits for such means then becomes:

$$\text{mean} \pm \frac{R_4}{\sqrt{2N}} \tag{A.10}$$

Confidence limits for probability levels other than 95 %, single- or double-sided as required, can be calculated by selecting the appropriate value Z from [Table A.1](#), and multiplying by the conversion factor $Z/1,96$.

In a set of $N + 1$ such averages, therefore, the variance of the difference between a single average of k results and the mean of the remaining N averages is:

$$\frac{R_1^2}{2 \times 1,96^2} + \frac{R_4^2}{2N1,96^2} \quad (\text{A.11})$$

The 95 % confidence limit, R_3 , for the absolute difference is therefore:

$$R_3 = \sqrt{\frac{R_1^2}{2} + \frac{R_4^2}{2N}} \quad (\text{A.12})$$

In the case of only two laboratory averages (when $N = 1$) this formula reduces to:

$$R_2 = \sqrt{R^2 - r^2 \left(1 - \frac{1}{2k_1} - \frac{1}{2k_2} \right)} \quad (\text{A.13})$$

STANDARDSISO.COM : Click to view the full PDF of ISO 4259-2:2017

Annex B (informative)

Dispute resolution for specifications based on a specified degree of criticality

B.1 General

This annex presents an alternate procedure for dispute resolution between the supplier and recipient using a specified degree of specification criticality in view of the procedure in [7.3](#).

B.2 Criticality of specifications

Some specifications, because of the product characteristic or the end use of the product, or both, require that the recipient has a high degree of assurance that the product meets or exceeds the quality level indicated by the specification level(s). For the purpose of this document, such specifications are called “critical specifications”.

Specifications that require assurance only that the quality is not substantially poorer than is indicated by the specification are called “non-critical specifications” for the purposes of this document.

For this document, the degree of criticality of a specification, denoted by p_c , is defined to be the maximum probability (risk) that the recipient can tolerate of accepting a shipment which fails specifications. This p_c is set as the maximum probability of accepting a shipment based on the limiting case where property in question is exactly at the specification limit. The risk borne by the supplier, that is, the recipient will reject a shipment which marginally meets or marginally does not meet the specification, will thus be $1 - p_c$. In order to use the procedure in this annex, p_c is subject to prior agreement among the parties, and, like the specification limits and the test method, shall in that case be considered as an integral part of the specification.

This document considers specifications with a p_c value of $< 0,5$ as critical specifications, while those with p_c value $\geq 0,5$ are considered to be non-critical specifications.

Execution of procedures in this annex assumes the parties in dispute (supplier and recipient) have agreed a priori to a p_c as defined above.

While p_c may be selected at any value that is agreed upon between supplier and recipient, this document recommends the maximum and minimum values for p_c to be 0,95 (non-critical specification) and 0,05 (critical specification).

B.3 Construction of specifications

This step is described in [5.2](#).

B.4 Product acceptance or rejection to specifications for a pre-specified p_c

This subclause provides general procedures to allow the parties in dispute (supplier and recipient) to accept or reject the product with regard to the specification when one or more acceptable results are available from one or more laboratories. Results are deemed acceptable if [6.1.2](#) to [6.1.4](#) are satisfied. If it is necessary for the recipient to take action after examining these results, the procedure specified in [B.5](#) shall be adopted.

The procedures of [B.5](#) assume a test method which follows the normal probability distribution, with no bias, and with repeatability, r , and reproducibility, R . They also assume that a degree of criticality, p_c , has been agreed to in advance by supplier and recipient.

B.5 Dispute procedure

B.5.1 When there is no other source of information on the true value of a characteristic other than a single result, the product shall be considered to meet the specification limit, with confidence $100(1 - p_c)\%$, if the single result, X , is such that:

in the case of the upper limit of the specification A_1 ,

$$X \leq A_1 + 0,361(Z \times R) \quad (\text{B.1})$$

in the case of a the lower limit of the specification A_2 ,

$$X \geq A_2 - 0,361(Z \times R) \quad (\text{B.2})$$

and, in the case of a double limit (A_1 and A_2), both these conditions are satisfied.

The factor Z in these formulae is the value of the standard normal distribution corresponding to a probability, p (see [Table A.1](#)). Note that for critical specifications ($p_c < 0,5$), Z has a negative value, and that the confidence $100(1 - p_c)\%$ that the product meets the specification limit is greater than for non-critical specifications. The factor 0,361 is the reciprocal of $1,96\sqrt{2}$ (see [Formula \(A.2\)](#)) used to convert reproducibility to a standard deviation.

If the reproducibility, R , is a function of the true value of the property in question, the value of R applied in [Formula \(B.1\)](#) is that which is appropriate for a true value of A_1 , whereas for [Formula \(B.2\)](#), R shall be computed assuming the true value is A_2 .

B.5.2 If it is not possible for the supplier and the recipient to reach agreement about the quality of the product on the basis of their existing results, then the procedures given in [B.5.3](#) to [B.5.11](#) shall be adopted.

To engage in this procedure a pre-requisite is that there is a sufficient amount of agreed-upon adjudication sample available.

B.5.3 Each laboratory shall reject its original results and obtain at least three other acceptable results on the adjudication sample to ensure that the work has been carried out under repeatability conditions. Result(s) from each laboratory is considered to be acceptable if [6.1.2](#) to [6.1.4](#) are met, and, in the case of more than one result, if [4.2.2](#) is met. The average of the acceptable results in each laboratory shall then be computed, divergent results being discarded as indicated in [4.2.2](#). The supplier's and recipient's averages will be denoted \bar{X}_S and \bar{X}_R , respectively.

B.5.4 If \bar{X}_S and \bar{X}_R are acceptable in terms of reproducibility (see [4.3.1](#)), and if $\bar{X} = (\bar{X}_S + \bar{X}_R)/2$ satisfies [Formula \(B.1\)](#) or [Formula \(B.2\)](#) or both, as appropriate, with R replaced by R_2 of [4.3.1](#) if necessary, then the product meets the specification.

B.5.5 If \bar{X}_S and \bar{X}_R are acceptable in terms of reproducibility (see [4.3.1](#)) and if $\bar{X} = (\bar{X}_S + \bar{X}_R)/2$ fails either [Formula \(B.1\)](#) or [Formula \(B.2\)](#), as appropriate, with R replaced by R_2 of [4.3.1](#) if necessary, then the product fails the specification.