TECHNICAL REPORT

ISO/TR 9272

Second edition
2005-07-15

Rubber and rubber products — Determination of precision for test method standards

Caoutchouc et produits en caoutchouc — Évaluation de la fidélité des méthodes d'essai normalisées

Reference number
ISO/TR 9272:2005(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies

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.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. 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.

In exceptional circumstances, when a technical committee has collected data of a different kind from that which is normally published as an International Standard ("state of the art", for example), it may decide by a simple majority vote of its participating members to publish a Technical Report. A Technical Report is entirely informative in nature and does not have to be reviewed until the data it provides are considered to be no longer valid or useful.

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.

ISO/TR 9272 was prepared by Technical Committee ISO/TC 45, *Rubber and rubber products*, Subcommittee SC 2, *Testing and analysis*.

This second edition cancels and replaces the first edition (ISO/TR 9272:1986), which has been technically revised.

Introduction

The primary precision standard for ISO test method standards is ISO 5725, a generic standard that presents the fundamental statistical approach and calculation algorithms for determining repeatability and reproducibility precision as well as accuracy and a concept related to bias called trueness. However there are certain parts of ISO 5725 that are not compatible with precision determination in the rubber manufacturing and carbon black industries over the past four decades.

two major problems exist:

a) strict adherence to ISO 5725 conflicts with the operational procedures and the past history of testing as conducted in these two industries and

b) ISO 5725 does not address certain requirements that are unique to rubber and carbon black testing.

Thus although ISO 5725 is necessary as a foundation document for this Technical Report and is used as such, it is not sufficient for the needs of TC 45.

This Technical Report replaces ISO/TR 9272, an interim document that has been used for guidance on precision determination since 1986. This new edition of the Technical Report has a more comprehensive approach to the overriding issue with precision determination over the past several decades — the discovery that the reproducibility (between-laboratory variation) of many test methods is quite large. The existence of very poor between-laboratory agreement for many fundamental test methods in the industry has been the subject of much discussion and consternation. Experience has shown that poor reproducibility is most often caused by only a small number (percentage) of the laboratories that may be designated outlier laboratories. This new edition of ISO/TR 9272 describes a "robust" analysis approach that eliminates or substantially reduces the influence of outliers. See Annex E for a more detailed discussion of these issues and additional background on ISO 5725.

Five annexes are presented. These serve as supplements to the main body of the Technical Report. They are in addition to the terminology section proper.

— Annex A defines the Mandel h and k statistics, illustrates how they are calculated and gives tables of critical h and k values.

— Annex B lists the calculation formulae for repeatability and reproducibility. It also describes how to generate and use six tables that are required for a spreadsheet precision analysis.

— Annex C outlines the procedure for calculating replacement values for outliers that have been rejected by h and k value analysis. Outlier replacement rather than deletion is an option that may be used for precision determination with a minimum number of laboratories and/or materials.

— Annex D is an example of a typical general precision determination programme for Mooney viscosity testing. It shows how a precision database is reviewed for outliers, using both the h and the k statistics, and illustrates some of the problems with outlier identification and removal as described in ISO 5725-2.

— Annex E presents some background on ISO 5725, robust analysis and other issues related to precision determination.

Annex E is given mainly as background information that is important for a full understanding of precision determination. Annexes A, B, and C contain detailed instructions and procedures needed to perform the operations called for in various parts of this Technical Report. The use of these annexes in this capacity avoids long sections of involved instruction in the main body of the Technical Report, thus allowing better understanding of the concepts involved in the determination of precision.

Rubber and rubber products — Determination of precision for test method standards

1 Scope

This Technical Report presents guidelines for determining, by means of interlaboratory test programmes (ITPs), precision for test method standards used in the rubber manufacturing and the carbon black industries. It uses the basic one-way analysis of variance calculation algorithms of ISO 5725 and as many of the terms and definitions of ISO 5725 as possible that do not conflict with the past history and procedures for precision determination in these two industries. Although bias is not determined in this Technical Report, it is an essential concept in understanding precision determination. The ISO 5725 concepts of accuracy and trueness are not determined in this Technical Report.

Two precision determination methods are given that are described as "robust" statistical procedures that attempt to eliminate or substantially decrease the influence of outliers. The first is a "level 1 precision" procedure intended for all test methods in the rubber manufacturing industry and the second is a specific variation of the general precision procedure, designated "level 2 precision", that applies to carbon black testing. Both of these use the same uniform level experimental design and the Mandel h and k statistics to review the precision database for potential outliers. However, they use slight modifications in the procedure for rejecting incompatible data values as outliers. The "level 2 precision" procedure is specific as to the number of replicates per database cell or material-laboratory combination.

2 Normative references

The following referenced documents are indispensable for the application 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 3534-1, *Statistics — Vocabulary and symbols — Part 1: Probability and general statistical terms*

ISO 5725 (all parts), *Accuracy (trueness and precision) of measurement methods and results*

3 Terms and definitions

3.1 General

For the purposes of this document, the terms and definitions given in 3.3 apply, together with those in ISO 5725 with modifications in 3.2.

Additional terms concerning certain types of precision can be found in 5.3. Better understanding can be gained by giving these definitions, which relate to the nature of the material to be tested, in that subclause.

3.2 ISO 5725 terms

Terms defined in ISO 5725, usually those from ISO 3534-1, are used when:

a) their definition does not conflict with the procedures required for a comprehensive treatment of precision determination for TC 45 test method standards, and

b) when they are adequate to the task of giving definitions that are informative and promote understanding.

In this subclause, some additional notes have been added to the ISO 5725 term definitions to give greater insight into precision determination for TC 45 test methods.

3.2.1
accepted reference value
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 (measured) quantity, i.e. the mean of a specified population of measurements.

3.2.2
test result
value of a characteristic obtained by carrying out a specified test method

NOTE The test method should specify that one or a number of individual measurements, determinations or observations be made and their average or another appropriate function (median or other) be reported as the *test result*. It may also require standard corrections to be applied, such as correction of gas volumes, etc.

3.2.3
accuracy
closeness of agreement between a test result and the accepted reference value

NOTE 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.

3.2.4
bias
difference between the expectation of the test results and an accepted reference value

NOTE Bias is the total systematic error (deviation) as contrasted to random error. There may be one or more systematic error components contributing to bias. A larger systematic difference from the accepted reference value is reflected by a larger bias.

3.2.5
laboratory bias
difference between the expectation of the test results from a particular laboratory and an accepted reference value

3.2.6
precision
closeness of agreement between independent test results obtained under stipulated conditions

NOTE 1 Precision (for within-laboratory conditions or repeatability) depends on the distribution of random errors and does not relate to the true value (accepted reference value) or the specified value. For a global testing domain (between-laboratory conditions), see 3.3.1 below, the between-laboratory precision (reproducibility) is influenced by laboratory bias as well as the random variations inherent in such a global testing domain.

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

NOTE 3 The term "independent test results" is defined as a set of results where the measurement of each value (of the set) has no influence on the magnitude of any other test result in the set.

NOTE 4 Quantitative measures of precision depend critically on the stipulated conditions (the type of test domain). Repeatability and reproducibility conditions are particular sets of extreme conditions.

NOTE 5 Alternatively, precision may be defined as a "figure of merit" concept. It is proportional to the inverse of the dispersion of independent replicate (test or observed) values, as estimated by the standard deviation, for a specified testing domain.

3.2.7
repeatability conditions
conditions where independent test results are obtained with the same method on identical test items (or elements) in the same laboratory by the same operator using the same equipment within short intervals of time

NOTE As defined in 3.3.1, a "local test domain" is the locale or environment (in a particular laboratory) under which repeatability tests are conducted. The word "identical" should be interpreted as "nominally identical", i.e. no intentional differences among the items. The "intervals of time" between repeat measurement of test results may be selected by the consensus of a particular testing community. For TC 45 and the international rubber manufacturing industry, the time interval between repeat tests is of the order of one to seven days.

3.2.8
repeatability
precision under repeatability conditions

NOTE 1 Repeatability, defined by the symbol r, is expressed in terms of an interval or range that is a multiple of the standard deviation; this interval should (on the basis of a 95 % probability) encompass duplicate independent test results obtained under the defined local testing domain.

NOTE 2 Relative repeatability, (r), is expressed in terms of an interval (a multiple of the standard deviation) that is a percentage of the mean level of the measured property; this interval should (on the basis of a 95 % probability) encompass duplicate independent test results (on a percentage basis) obtained for a defined local testing domain.

NOTE 3 Repeatability may be dependent on the magnitude or level of the measured property and is usually reported for particular property levels or materials or element classes (that determine the level).

NOTE 4 Although repeatability as defined above applies to a local testing domain, it can be obtained in two different ways and the term repeatability can be used in two different contexts. It can pertain to a common community value, obtained as an average (or pooled) value from all laboratories in an ITP among N different laboratories. This can be referred to as a universal or *global* repeatability, that applies to a "typical laboratory", that stands as a representative of all laboratories that are part of a global testing domain. It can also pertain to the long-term or established value for a "particular laboratory" as derived from ongoing testing in that laboratory, not related to any ITP. The second use can be referred to as a *local* repeatability, i.e. repeatability obtained in and for one laboratory.

3.2.9
reproducibility conditions
conditions where test results are obtained with the same method on identical test items (or elements) in different laboratories with different operators using different equipment

NOTE 1 Each laboratory (or location) in the global testing domain, see 3.3.1.5, conducts n repeatability tests on a material (target material) and reproducibility is determined based on the mean values (of the n local domain tests) for the N laboratories for that material. Reproducibility may also depend on the level of the measured property or on the materials tested and it is also usually reported for particular levels or materials.

NOTE 2 The term "different equipment" should be interpreted as different realizations of an accepted and standard test device, i.e. all of the test devices are nominally identical but they are located in different laboratories.

3.2.10
reproducibility
precision obtained under reproducibility conditions

NOTE 1 Reproducibility, R, (for a defined global testing domain) is obtained by way of independent tests conducted in N laboratories (with n replicates each) on nominally identical test items or elements, expressed in terms of an interval or range that is a multiple of the standard deviation; this interval should (on basis of a 95 % probability) encompass duplicate test results, each obtained in different laboratories for a defined global testing domain.

NOTE 2 Relative reproducibility, (R), is expressed in terms of an interval (a multiple of the standard deviation) that is a percentage of the mean level of the measured property; this interval should (on the basis of a 95 % probability) encompass duplicate independent test results (on a percentage basis) each obtained in different laboratories for a defined global testing domain.

NOTE 3 Reproducibility may also depend on the level of the measured property or on the materials tested and it is also usually reported for particular levels or materials. Reproducibility usually does not have the dual interpretation or use as discussed above for repeatability, since it is a "group characteristic" that only applies across a number of laboratories in a global testing domain.

NOTE 4 As indicated in Note 1 in the definition of precision above, reproducibility is determined by the magnitude of random variations in the global testing domain as well as the distribution of bias components in this same global domain. Laboratories that have good agreement with either a reference value or an overall mean value for the ITP, have either zero or a very small bias. Laboratories that do not have good mean value agreement have substantial biases and, although the bias magnitude is relatively constant for each laboratory, it differs among the biased laboratories, i.e. it has the characteristics of a distribution.

3.2.11
outlier
member of a set of values which is inconsistent with the other members of that set

NOTE This TC 45 standard defines a "set" as a "class of elements" that are subjected to measurement. See *element* and *element class* defined in 3.3.1 below.

3.3 Required terms not in ISO 5725

A number of specialized terms are defined here in a systematic sequential order, from simple terms to complex terms. This approach allows the simple terms to be used in the definition of the more complex terms; it generates the most succinct and unambiguous definitions.

3.3.1 Basic testing terms

3.3.1.1
element
entity that is tested or observed to determine a property or characteristic; it may be a single object among a group of objects (test pieces, etc.) or an increment or portion of a mass (or volume) of a material

NOTE The generic term *element* has a number of synonyms: item, test piece, test specimen, portion, aliquot part, sub-sample, laboratory sample.

3.3.1.2
element class
class of elements)
category or descriptive name for a group of elements that have a common origin or have nominally identical properties

NOTE The term “nominally identical” implies that the elements come from a source that is as homogeneous as possible with regard to the property being measured.

3.3.1.3
testing domain
location and operational conditions under which a test is conducted; it includes a description of the element preparation (test sample or test piece), the instrument(s) used (calibration, adjustments, settings), the selected test technicians and the surrounding environment

3.3.1.4
local testing domain
domain comprised of one location or laboratory as typically used for quality control and internal development or evaluation programmes

3.3.1.5
global testing domain
domain that encompasses two or more locations or laboratories, domestic or international, typically used for producer-user testing, product acceptance and interlaboratory test programmes

3.3.1.6
balanced uniform level design
plan for an interlaboratory test programme (ITP) for precision, where all laboratories test all the materials selected for the programme and each laboratory conducts the same number of repeated tests, n, on each material.

3.3.2 Material and sampling terms

3.3.2.1
material
specific entity or element class to be tested; it usually exists in bulk form (solid, powder, liquid)

NOTE Material is used as a generic term to describe the “class of elements” that is tested, i.e. a material may be a rubber, a rubber compound, a carbon black, a rubber chemical, etc. A material may or may not be homogeneous. In product testing, the term material may be used to describe the “class of elements” or type of rubber product such as O-rings, hose assemblies, motor mounts, etc. See also the definition of “target material” in 5.3.

3.3.2.2
lot
specified mass or volume of material or number of objects; usually generated by an identifiable process, frequently with a recognized composition or property range

NOTE A lot may be generated by a common production (or natural) process in a restricted time period and usually consists of a finite size or number. A lot may be a fractional part of a population. A recognized property range implies that some rough approximation is available.

3.3.2.3
sample
〈physical sample〉 number of elements or the specified mass of a material, selected according to a particular procedure, used to determine material, lot or population characteristics

NOTE The term “sample” should not be used as a synonym for “material”, or “target material”, see 5.3. Ideally several “materials” are tested in any ITP with each material being different (chemically, structurally, property wise). From each material, some number of “samples” (all nominally identical) may be taken for testing.

3.3.2.4
sample
⟨data⟩ number of test or observation values (n = 1, 2, 3, etc.) obtained from (one or more) physical samples by the application of a specific test (observation) method

3.3.2.5
test sample
part of a (physical) sample of any type taken for chemical or other analytical testing, usually with a prescribed blending or other protocol

NOTE A test sample is usually a mass or volume that is some very small fractional part of a bulk material.

3.3.2.6
test piece
object (appropriately shaped and prepared) taken from a sample (or lot) for physical or mechanical testing

NOTE The term "test specimen" is a synonym for test piece.

3.3.3 Additional statistical terms relating to precision

3.3.3.1
replicate
one of a selected number of independent fractional parts or independent number of elements, taken from a sample; each fractional part or element is tested.

NOTE The word replicate as defined above refers to a physical object (element). It can also be used in reference to a data set, where it refers to one of a number of independent data values.

3.3.3.2
true value
measured or observed value for an element, that would be obtained for a testing domain in the absence of errors, deviations or variations of any sort, i.e. where there is no "system-of-causes" variation

NOTE The true value is also defined as the mean that would be obtained by testing all members of any population. Typical "systems of causes" are the unavoidable fluctuations in temperature, humidity, operator technique, fidelity of calibration, etc. in a controlled testing domain.

3.3.3.3
uncertainty
quantity that characterizes, in an inverse manner, the "figure of merit" for a measurement or observation; for a given local domain, it is the magnitude of the difference between the measured element value and an accepted reference value and includes both random and bias deviations

NOTE The definition of "uncertainty" given here attempts to capture the general nature of the concept. It has been defined equivalently but using different words, by a number of organizations addressing this concept. The word "uncertainty" as defined here is distinguished from the ordinary use of the word. As indicated, "goodness" or "merit" and "uncertainty" (doubt about the measurement) are inversely related. Uncertainty is a characteristic of a *local test domain*; each local domain for any defined test may have a different uncertainty value. Precision as determined by a typical ITP (both repeatability and reproducibility) is a characteristic of a *global test domain*; the precision values obtained in any ITP are intended for universal application, i.e. to a number of laboratories as a group.

4 Field of application

4.1 General background

This Technical Report applies to test methods that have test results expressed in terms of a quantitative continuous variable. It is in general limited to test methods that are fully developed and in routine use in a number of laboratories.

Tests are conducted using standard test methods to generate test data that are used to make technical and other decisions for commercial, technical and scientific purposes. Therefore the precision of a particular test method is an important quality characteristic or figure of merit for a test method and a decision process.

A determination of the precision of a test method is normally conducted with (1) some selected group of materials typically used with that method and (2) with a group of volunteer laboratories that have experience with the test method. The determination represents an "event in time" for the test method for these materials and laboratories. Another ITP precision determination with somewhat different materials or even with the same materials with the same laboratories at a different time may generate precision results that differ from the initial ITP.

The confidence intervals for the estimated values for repeatability and reproducibility standard deviations is addressed in ISO 5725 and is not part of this Technical Report. The treatment of precision parameter confidence intervals in ISO 5725 assumes the inherent variation in individual values for both repeatability and reproducibility standard deviations (in a long run series of evaluation programmes), is attributable to random test data variations with a normal distribution. Experience as indicated in References [1], [2], [3] and [4] and elsewhere has shown, however, that the poor reproducibility among the laboratories of a typical ITP is due to interlaboratory bias. Certain laboratories are almost always low or high compared to a reference as well as other laboratories in all tests. This offset or bias is typically different for each laboratory that has such a bias. This is in distinction to random deviations compared to a reference as required by a normal distribution. Thus any confidence intervals calculated for the important precision parameter reproducibility, based on a random model, are not valid.

Caution is urged in applying precision results of a particular test method to product testing for consumer-producer product acceptance. Product acceptance procedures should be developed on the basis of precision data obtained in special programmes that are specific to the commercial products and to the laboratories of the interested parties for this type of testing.

An additional concept related to test method technical merit is "test sensitivity". Test sensitivity is defined as the ratio of the test discrimination power for the fundamental property measured, to the property measurement error or standard deviation.

4.2 Defining repeatability and reproducibility

Repeatability and reproducibility are each equal to a range or interval that is a special multiple of the respective standard deviation. The repeatability, designated r, is given by:

$$\text{Repeatability} = r = \phi 2^{1/2} s_r \quad (1)$$

where s_r = the pooled (across all laboratories) "within-laboratory" standard deviation,

and reproducibility, designated R, is given as:

$$\text{Reproducibility} = R = \phi (2)^{1/2} s_R \quad (2)$$

where s_R = the square root (or standard deviation) of the sum of (1) the between-laboratory variance (using the mean of n values for each laboratory for the calculation) and (2) the pooled within-laboratory variance (variance for the n values in each laboratory).

The term $(2)^{1/2}$ is required since r and R are defined as the maximum difference between two single test results that can be expected on the basis of a chance or random occurrence alone at the 5 % probability level or 95 % confidence level. The variance of the difference $(x_1 - x_2)$ for two values taken at random from a population is equal to the sum of the variances for values (of x) taken one at a time from the same population. Since there are two x values, the sum of the variances is simply the variance of x values times two and the square root places this term on a standard deviation basis. In this context each x value represents a "test result" as defined in any particular test method standard.

Thus $(2)^{1/2} s_R$ is the standard deviation of differences. The factor ϕ depends on both the total degrees of freedom in the estimation for either of the standard deviations and on the shape of the distribution of the

variable bias terms and the random error terms. The normal assumptions for these are (1) the distributions are unimodal, (2) the number of test results is sufficient (approximately 20) and (3) a probability level of p = 0,05 or confidence level of 95 % is chosen. Under these assumptions the value of ϕ is similar to a t-value or approximately 2,0 and therefore the simplified expressions for r and R are

$$\text{Repeatability} = r = 2{,}83 s_r \tag{3}$$

$$\text{Reproducibility} = R = 2{,}83 s_R \tag{4}$$

For more details, see the discussion notes in the definitions for repeatability and reproducibility in 3.1.

5 Precision determination: Level 1 precision and level 2 precision

5.1 Level 1 precision

Two precision categories are described: level 1 precision and level 2 precision. Level 1 precision is discussed first. Level 1 precision determination follows established procedures used in the rubber manufacturing industry on an international basis for the past two decades. The determination is conducted using a balanced uniform level design ITP with three or more materials sent to each of the participating laboratories with tests conducted to generate an independent "test result", on each of two test days. The ITP database is reviewed for outliers by the Mandel h and k consistency statistics (see Annex A).

a) *Options for outliers* — If no outliers are found, the original database is used to develop a table of precision results. If outliers are identified in any ITP database, there are two options for outlier treatment. Option 1, outlier deletion, is the first choice. Option 2, outlier replacement, is chosen for an ITP with a minimum number of laboratories (ca. six). Issues such as the number of replicate values on each test day and/or the number of technicians or operators used to obtain a test result, which are characteristic of the particular test, are considered on a case-by-case basis by the ITP organizing committee. Outlier treatment is discussed in greater detail in Annexes A, C, D and E.

b) *Types of test method* — Level 1 precision has been successfully used for the broad range of test methods characteristic of the rubber manufacturing industry; from simple "bench type" tests, conducted in few minutes (hardness and pH tests) to a complex multi-step test method, such as an ageing test. Such a test requires preliminary property measurement, a substantial ageing period (days) followed by property measurement after ageing to obtain a final calculated test result or performance index. For such complex tests, any realistic precision determination must include all of the procedural steps in arriving at the test result, the basic datum used in precision analysis and determination. The procedures required for general precision are described in Clauses 8, 9 and 10.

5.2 Level 2 precision

The carbon black industry has adopted a slightly revised precision determination procedure designated "level 2 precision". The number of replicates in each cell of a uniform level design ITP is specified as four, two by each of two test technicians. The outliers are reviewed by a special procedure that depends on the number of laboratories in the ITP and the precision, absolute or relative, is expressed by a specified procedure. The procedures for this precision are listed in Clause 11.

5.3 Types of level 1 and level 2 precision

In addition to the ageing tests cited above, other tests also require a more complex total sequence of operations to generate a final test result. One important test of this type is a "performance-in-rubber" test; the evaluation of various rubbers, reinforcement fillers or other compounding materials in standardized formulations. The typical stress-strain evaluation of a lot of a specified rubber will require:

a) a representative sample of the rubber;

b) a standardized formulation and mixing operation to prepare a compound using standard materials;

c) processing of this compound to prepare cured moulded sheets for a selected time and temperature;

d) cutting and gauging of dumbbell (or other) test pieces;

e) the testing of these to obtain the final test results for modulus, elongation and tensile strength properties.

To permit realistic precision determination for performance-in-rubber testing, it is necessary that *all the steps* in the operation be replicated, starting from the raw materials to the final test result. Each of these steps has a potential component of variance and the sum of all variance components establishes the overall test variance and standard deviation. To address this, two types of precision are defined. The two types are characterized by the relationship between the material (or element class) *tested* and the material *directly* evaluated for precision. To explain this, it is necessary to introduce and define a new term:

— *target material*: the material (or class of elements) that is the primary *focus of attention* for a precision determination programme; however it may not be tested in its usual or ordinary physical state.

Using the term “target material”, two types of precision may be defined:

— *Type 1 precision* — A precision determined *directly* on, a target material; prepared test pieces or test portions of the target material (class of elements) drawn from a homogeneous source are tested, with no processing or other operations required prior to testing.

NOTE 1 An example is a lot comprised of died-out, gauged dumbbells for stress-strain testing.

— *Type 2 precision* — A precision determined *indirectly* for a target material; the target material is usually combined with a number of homogeneous ancillary materials to form a composite material and testing is conducted on samples of this and the property response of the target material is determined.

NOTE 2 The properties of the composite material are directly related to the quality or properties of the target material. An example: To determine the quality of a grade of SBR, a sample of the rubber, plus curatives, fillers, antioxidants, etc., are mixed and cured, test pieces are prepared and the resulting compound tested for specified quality properties.

NOTE 3 It is possible that a type 1 precision programme might be conducted on test pieces or portions that require some minimum processing or other simple operations prior to actual testing. This is, in a strict sense, an intermediate level of precision. However, to avoid unnecessary complications, this will be designated a type 1 precision.

6 Steps in organizing an interlaboratory test programme

The steps required to organize an ITP, with a discussion for each procedural step, are as follows:

a) *Organization committee* — An organization committee or task group and a programme co-ordinator should be selected. One member of the committee or group should be a statistician familiar with the technology of the test method as well as the content of this Technical Report. Most ITPs are organized on the basis of a *balanced uniform level design* for the precision programme. For more advanced designs, see ISO 5725.

b) *Category and type of precision* — For all programmes except for carbon black testing, a level 1 precision ITP is organized. For carbon black testing a level 2 precision ITP is organized. The type of precision to be determined shall be selected (see 5.3). Type 1 precision is the most frequently determined. For some test methods, such as rubber or polymer or other performance-in-rubber evaluations using standard formulations, a type 2 precision is required.

c) *Test operator or technician selection* — For simple level 1 precision testing requiring only one operator or technician, all replicate tests should be conducted by the same technician unless the effect of different technicians is part of the intended programme. For more complex tests where several operators or technicians are required to perform a sequence of different steps to arrive at a test result, the same “operator team” should conduct testing for all replicates. For level 2 precision testing, follow the procedure of using two technicians on each of two test days (see Clause 11).

d) *Test result and number of replicates* — Each test method has a final value for the property under evaluation, defined as a *test result*. A test result may be a mean or median value of a number of individual determinations as specified by the test method. For the purposes of this Technical Report, a replicate is defined as a *test result*. The number of replicate test results, n, within each laboratory on any material should be specified. In most ITPs, this is two (2). For some tests, three (3) or four (4) replicates, as in level 2 precision, may be selected. All analysis is conducted on test results.

e) *Time period for repeatability* — The time period between replicate tests within any laboratory should be selected. This time period is usually in the range of 1 day to 7 days. See Annex E for more discussion on repeatability time periods. For special tests (long ageing periods), replicate tests may require a longer time span. For other special testing operations, shorter time periods (minutes, hours) may be selected. The primary consideration is how the test method is typically used in the industry. The selected time period shall be reported in the precision clause of the test method standard.

f) *Number of target materials* — The number of target materials or classes of objects (or manufactured products) to be tested should be selected. Ideally this should be three or four with substantially different property levels. The target materials should represent typical industry materials as normally used and subjected to test. See 5.3 for details.

g) *Preparation of homogeneous target materials* — A homogeneous lot of each of the target materials should be prepared, with sufficient reserve quantity so that re-tests can be made if needed. If the material lends itself to a blending operation to ensure homogeneity, blending should be done. If blending is not possible, special procedures should be conducted to obtain the most homogeneous material (or collection of elements) that is possible by way of closely monitored laboratory or other preparation operations. Documentation should be provided to ascertain the homogeneity. If any ancillary materials are required as for type 2 precision, these lots should be either standard reference materials or special documented homogeneous lots.

h) *Number of laboratories* — For a reliable estimate of precision, at least six (6) laboratories skilled in the test method are required for the final database (after outlier treatment) in the ITP. For the more important industry test methods, 12 to 18 laboratories should participate. If six or more laboratories are not in the final database, an analysis can be conducted with fewer laboratories but the estimates of precision, especially reproducibility, are seriously compromised and only represent very rough estimates.

i) *Packaging and delivery of materials* — All the materials required for any ITP should be appropriately packaged to prevent any change with time or storage in the properties to be measured. Appropriate storage conditions in each participating laboratory prior to test need to be specified. The shipment of all materials should be co-ordinated with the test schedule (discussed below) so that all materials are available for the scheduled test dates.

j) *Testing instructions* — Although all ITPs are usually conducted for a standard test method that includes the complete set of instructions for the test, some supplementary instructions are required. One important supplementary instruction is the schedule for the testing. All tests should be performed on specified days and all participating laboratories should conduct the test as specified by the standard. The schedule should allow for adequate material delivery time. Any special modifications of the standard method should be clearly described as well as special instructions as to operators or technicians (one, two or more) vs replicate testing. If an ITP is to be conducted for a test method at some intermediate development level, it is essential to give all participating laboratories instructions for conducting the test as well as all the required ITP instructions.

k) *ITP test data report* — A “test report data form” should be prepared by the ITP co-ordinator and a copy sent to each participating laboratory along with the test materials and instructions. This form should contain locations to report the following: the name of the laboratory; the test dates actually used and for each target material tested, and the test value (test result) for each replicate test (day), reported if possible to one more significant figure than is normally used (i.e. do not truncate). The test report form should also ask for a description of the test equipment or machines used (model No., condition), comments about any unintended deviations from standard test procedure and disclosure of any mishaps or other pertinent information. The completed test report should be returned to the ITP co-ordinator.

7 Overview of level 1 precision analysis procedure

7.1 Analysis operation sequence

This clause gives a quick overview of the procedures for the analysis of the ITP database and provides the user with a better appreciation of the complete analysis process. Some background on outliers is also presented in this clause. The level 1 precision procedure may require as many as three analysis operations or overall steps. The actual number will be determined by the uniformity of the data in the database. If there are no outliers, only analysis step 1 is used.

If outliers are present, analysis steps 2 and 3 may be required, depending on the extent of outliers in the database. Annex B contains instructions for all three analysis operations and also gives the details on how to lay out the required computer spreadsheet tables and their interlinking that enables the automatic recalculation of the final precision parameters, r and R, when outliers are deleted or replacement values are substituted into the Table 1 format basic data. Figure 1, is a decision tree or flow chart diagram that outlines the steps in the complete analysis process.

a) *Preliminary data review* — A quick numerical review of any database is important to gain a first impression of the results of any ITP. This is conducted after cell averages and cell standard deviations (or cell ranges) have been calculated. Part of this review is the generation of special plots of cell averages and cell standard deviations or cell ranges vs laboratory number. These plots, described in 8.1, clearly show potential outlier values.

b) *Analysis step 1* — The original database is analysed to generate values for repeatability and reproducibility for each material (or target material) and the h and k statistics calculated. See Annex A. Annex B gives the instructions for generating six tables that yield values for the h and k statistics and the precision results for each material. The calculated h and k values are compared to the 5 % significance level critical h and k values to determine if there are any significant outlier values. If there are none, the analysis is complete and the values found for repeatability and reproducibility are used to generate a table of precision results for the test method. If there are any significant outliers, analysis step 2 is required.

c) *Analysis step 2* — If there are any outliers at the 5 % significance level, the outlying values are

   1) either deleted using option 1 as described in 5.1;

   2) or replaced (see Annex C) using option 2.

On the basis of either option, the resulting revised database, designated revision 1 (or R1), is analysed to generate new values for repeatability and reproducibility, designated revision 1 precision values. This analysis produces a new set of calculated h and k values that are compared to 2 % significance level critical h and k values to determine if there any significant outlier values at this level. If there are none, the analysis is complete and the values found for repeatability and reproducibility are used to generate a table of revision 1 precision results for the test method. If there are any significant outliers, analysis step 3 is required.

d) *Analysis step 3* — If any of the revision 1 calculated h and k values *exceed* the 2 % significance level critical h and k values, the outlying values are

   1) either deleted using option 1;

   2) or replaced using option 2.

e) On the basis of either option, the resulting revised or revision 2 (or R2) database is analysed to generate new values for repeatability and reproducibility, designated revision 2 precision values. This completes the analysis sequence and the values found for repeatability and reproducibility for each material are used to prepare a table of precision results for the test method.

The level 1 precision part of this Technical Report does not address the issue of attempting to find a relationship between r, R, (r) or (R) and the property (level) for any ITP for two reasons. First, most ITPs do

not have a sufficient number of materials to produce any meaningful functionality of precision vs material level; the degrees of freedom for any obtained fit are small. Second, experience has shown that, even when there are several materials in an ITP, a well-fitting linear or other relationship is not obtained. It should be remembered that any ITP is an "event in time" that gives an indication of the general level of precision for three or four materials in a selected number of laboratories. With some occasional exceptions, the precision found is usually quite different for each material with no detectable pattern or functionality.

7.2 Background on outliers

The recognition and removal of the incompatible test values in any precision database is a subject of some controversy. If true outliers are not removed and their magnitude is substantial, seriously inflated values may be obtained for both precision parameters. This can result from only a few of the participating laboratories. However, caution must be exercised to ensure that high (or low) magnitude, but *bona fide*, values are not deleted. If such values are removed, the precision estimates will be too optimistic. The procedures presented in this Technical Report attempt to find a middle-ground position, designated a "robust analysis". Although objective probability-based techniques are used to declare incompatible values as outliers, all outlier rejection operations have a substantial conditional character and require some input and experience from the analyst.

7.3 Outlier appearance patterns

Outliers frequently occur in two general appearance patterns:

a) *None or infrequent* — There are no outliers or there are only a few outliers; one or two for every 20 data cells in a Table 1 format.

b) *Extensive* — Outliers occur in greater numbers, three, four or more for every 20 data cells and frequently in several of the cells for any laboratory.

When outliers are extensive, they may frequently be of substantial magnitude. There are of course some intermediate cases between these two extremes.

a) *Rationale 1 for outlier rejection* — There are two points of view on what significance level should be adopted for outlier rejection. The extremely conservative approach maintains that outliers should rarely be eliminated in any ITP. This is based in part on the concept that, in the preliminary stages of test method development, outlier rejection will lead to an overly optimistic impression of the quality of the method. This approach usually adopts a probability significance level of 0,5 % (p = 0,005) for outlier rejection. This approach has some limited merit for the initial stages of development for any test method especially when only a few laboratories participate in an ITP. However, this approach has some serious limitations as described below.

b) *Rationale 2 for outlier rejection* — For well-established test methods and any group of laboratories, experience has taught that there is a distribution of skill and testing competence, from poor to good. This capability range argues for a more realistic approach to the outlier issue; the use of a 5 % significance level, p = 0,05 (or a 95 % confidence level) for the declaration of incompatible values as outliers. This is the usual level for most statistical significance tests and will in general reject the results of laboratories that have poor quality control for internal testing and are in need of improved testing procedures. Allowing a few "poor" laboratories to inflate the determined precision gives a false negative impression of the true precision defined by laboratories with good control of testing operations. The precision of the "good" laboratories (the majority of those participating) should be the benchmark for industry-wide precision level for any test method. The use of the robust level 1 and level 2 precision procedures to identify these poor quality control laboratories can lead to a general industry-wide improvement for any test method, provided that feedback is employed to encourage the poorly performing laboratories to improve testing operations.

7.4 Sequential review of outliers

Experience in outlier review at the 5 % significance level raises the issue of a subsequent review of the database once the 5 % outliers are deleted. To properly frame this operation, recall that the h and k statistics represent ratios of either individual cell averages or cell standard deviations to the "across all laboratory"

standard deviation for each parameter. The influence of any outlier extends to both the outlier value itself (the numerator for h and k), as well as the standard deviation for all laboratories (the denominator for h and k).

The removal of 5 % significance outliers has now generated a second (or revision 1) database with substantially reduced "across all laboratories" or denominator standard deviation for either the h or the k statistic, or both. When outliers are deleted, the resulting revised database is one that might have been obtained had the outlying laboratories not volunteered for the ITP. The question now presents itself: Can this revision 1 database be reviewed again for h and k outliers using the newly calculated "across all laboratory" h and k standard deviations?

For any ITP that contains originally six or more laboratories, the answer to this question is "yes" and the second or revised database should be reviewed for any potential outliers. However, to guard against the generation of an excessively optimistic precision, the significance level for this second review should be more rigorous than for the initial review and should be conducted at the 2 % significance level. For any ITP that contains less than six laboratories, the decision to conduct a second review is left to the judgement of the analyst.

8 Level 1 precision: Analysis step 1

8.1 Preliminary numerical and graphical data review

Prior to the detailed calculations of analysis step 1, it is important to review the data by a graphical technique that indicates the uniformity of the database. The most frequently used precision determination is a uniform level design; all laboratories test the same number of replicates and test all materials. Table 1 indicates the layout for this uniform level design and gives the format for tabulating the basic data. There are a total of p laboratories and a total of q materials or element classes and a total of pq cells in the table. Each cell of the table, which constitutes a laboratory/material combination, contains n replicates; each test result replicate is designated as a Y_{ijk} value. The most frequently used design has two replicates per cell or n = 2.

A table in the format of Table 2 is prepared, for calculating cell averages, cell ranges or standard deviations, by calculating the average of the n replicates per cell as given in Table 1. After *cell averages* have been calculated they should be reviewed for any apparent outlier values as described in 8.1 and these noted for determination as given in the formal step 1 outlier rejection procedure as described in 8.3 and 8.4. See also Annex A.

A table in the format of Table 3 is prepared by calculating, for all cells, the standard deviation for the n replicates per cell. Alternatively cell ranges, denoted by w, the absolute difference between the maximum and minimum values in each cell, may be calculated. Both the *cell ranges* and the *cell standard deviations* should also be reviewed for any apparent outlier values and these noted for determination as given in the formal step 1 outlier rejection procedure as described in 8.3 and 8.4. See Annex A.

8.2 Graphical review of cell values

The general distribution of the data to disclose any potential outliers is reviewed with special plots of the cell averages and the cell ranges or standard deviations, using a typical spreadsheet programme. Prepare two new tables, one for cell averages, one for cell ranges or standard deviations. Cell ranges are used here because they facilitate certain calculation options that will be employed later in treating outliers, i.e. either deletion or replacement. For the *cell average* table and for the first material, generate two columns in the table, the first column containing the laboratory number, 1 to N, the second containing the corresponding cell average. Repeat this two-column "laboratory number/cell average" sequence for all materials. Prepare a table for *cell ranges* (or standard deviations) in the same manner as for cell averages with the "laboratory number/cell range" dual-column scheme.

a) Using the prepared tables, for each laboratory/material pair of columns, sort the cell averages (or cell ranges) in ascending order (across all laboratories), retaining the laboratory number with the cell value in the sorting operation. For each parameter (cell average or cell range), plot the parameter value vs the

laboratory number in ascending laboratory number order, using a line plot procedure. This is designated as an “ascending order trend”, or AOT, plot.

b) For an ITP with no outliers, the cell average plot is typically a positive-slope straight line with some reasonable degree of point scatter. If any outliers are present, they will be at the opposite ends of the plot, and will show substantial departure from the straight line of the central data point region. The cell range plot may contain more curvature from the low end (which may contain zero values) toward the central point region, but it will also indicate the outliers at the high-value end of the plot. Ascending-order plots will be used in the operation to replace outlier values with “replacement values” as outlined in Annex C.

8.3 Calculation of precision for original database

Comprehensive and specific instructions for this are given in Annex B. The test result values for the original database are entered into a table, designated Table B.1. This tabular format is also described as Table 1 in the main body of the Technical Report. However, to preserve continuity between Annex B and the following instructions, the table identification terminology of Annex B will be used.

NOTE There are no actual tables (with data or other actual table layout details) designated Tables B.1 to B.6 in this Technical Report. Annex B simply gives the instructions for the analyst to construct tables of the Table B.1 to B.6 format in a computer spreadsheet programme to be able to conduct an analysis. See however Annex D, the Mooney test example, which does give actual data tables in the format of Tables B.1 to B.6.

The next step is to set up a tabular format designated Table B.2 for cell averages and cell averages squared. The corresponding values in Table B.1 are the argument values for Table B.2.

Table B.3 is generated next: cell average deviations, denoted by d, and the calculated h-values. The corresponding values in Table B.2 are used as the arguments for Table B.3. Refer to Annex A for cell deviation d and h-value calculations.

Table B.4R for cell ranges and cell ranges squared and Table B.4S for cell standard deviations and cell variances (standard deviations squared) both address the same issue; the within-cell variation. It is recommended that both tables be generated in the analysis.

Table B.5 is used to calculate k-values for each cell in the database. The corresponding values in Table B.4S are used as the arguments to calculate k-values in Table B.5. Refer to Annex A for k-value calculations.

Table B.6 is used to calculate the precision parameters r and R. Values for T_1, T_2, T_4 and n and p are required to calculate r and R. See the embedded calculation algorithms 1 to 5 in Table B.6 and also Annex B for the details of these calculations.

8.4 Detection of outliers at the 5 % significance level using h and k statistics

The calculated values of h in Table B.3 and the calculated values of k in Table B.5 are reviewed for potential outlier values.

a) If the Table B.3 h-value for any cell equals or exceeds the 5 % significance level critical h-value given in Annex A, Table A.1, that particular cell value is declared an outlier.

b) If the Table B.5 k-value for any cell equals or exceeds the 5 % significance level critical k-value given in Table A.1, that particular cell value is declared an outlier.

c) If outliers are detected, a summary of the outliers detected is presented in the form of a sub-table at the bottom of Table B.6 showing the laboratory numbers that had 5 % significance outliers for both h and k for each material. See Table D.6 in Annex D for an example. When outliers are present, a revised database is generated by the use of either option 1, outlier deletion, or option 2, outlier replacement.

d) If there are no outliers for either cell averages or cell standard deviations (or ranges), the precision analysis is complete and the resulting values for r and R may be used to prepare a precision table for the test method.

8.5 Generation of revision 1 database using outlier option 1 or 2

If outliers are detected, the database is revised using either option 1 or 2.

a) Option 1 is the deletion of the n cell values in Table B.1 that are indicated as outliers and the correction of ERR indications in certain cells in Tables B.2 to B.6 that result from the deletion process described in Annex B. The deletion applies both to cell averages indicated by greater than 5 % critical h-values and to cell standard deviations (or ranges) indicated by greater than 5 % critical k-values. Once all ERR corrections have been made, the database is designated a revision 1 (R1) database. Each revision 1 table designation contains the appended symbols R1-OD (OD = outliers deleted). This revised OD database will be reviewed again for outliers at the more critical 2 % significance level as described in analysis step 2.

b) Option 2 is the replacement of the n cell values in Table B.1 that are indicated as outliers. The replacement applies to both cell averages and to cell standard deviations (or ranges) as indicated by greater than 5 % critical values. For either the h or the k values, the replacement is a two-sequence, one- or two-stage process. All of the details for this are fully described in Annex C. Once data replacements have been generated by the Annex C procedure, they are inserted into the database, replacing the outlier values to produce an R1 database using the table identification symbol R1-OR (OR = outliers replaced). This revised OR database will be reviewed again for outliers at the more critical 2 % significance level as described in analysis step 2.

8.6 Revision 1 (R1) database tables

A second set of tables in the format of Tables B.1 to B.6 is prepared for the step 2 analysis. As noted above, this second set should be:

a) tables designated B.1-R1-OD to B.6-R1-OD for the selection of option 1, outlier deletion, or

b) tables designated B.1-R1-OR to B.6-R1-OR for option 2, outlier replacement.

Once the deletions or the replacements have been made in accordance with the instructions in Annex B, the new set of precision values will appear in Table B.6-R1-OD or Table B.6-R1-OR depending on the option chosen.

9 Level 1 precision: Analysis step 2

9.1 Detection of outliers at the 2 % significance level using h and k statistics

The calculated values for h in Table B.3-R1-OD or Table B.3-R1-OR and the calculated values of k in Table B.5-R1-OD or Table B.5-R1-OR are reviewed for potential outlier values at the 2 % significance level. The calculated h and k values must be *greater than* the 2 % significance level for outliers to be rejected. For each of these tables, a sub-table is generated at the bottom of either table to summarize the results of the h and k comparisons of calculated values vs critical values. See Annex D for an example. If outliers are detected, the database is revised using either outlier option 1 or 2. The revision procedure is described in Annex B.

9.2 Generation of revision 2 database using outlier option 1 or 2

Outlier option 1 is the deletion of the n cell values in Table B.1-R1-OD that are indicated as outliers and the correction, as noted above, of ERR indications in certain cells in Tables B.2-R1-OD to B.6-R1-OD that result from the deletion process. Once all ERR corrections have been made, the database is designated a revision 2, or R2-OD, database. This revised OD database will be used for the operations of analysis step 3.

Outlier option 2 is the replacement of the n cell values in Table B.1-R1-OR that are indicated as outliers. The replacement applies to both cell averages as indicated by *greater than* 2 % critical values for either h or k. All of the details for this are fully described in Annex C. Once data replacements have been generated, they are inserted into the database to produce a revision 2, or R2-OR, database. This revised OR database will be used for the operations of analysis step 3.

10 Level 1 precision: Analysis step 3 — Final precision results

Although the Figure 1 decision tree diagram or flow sheet implies that analysis step 3 involves an analysis operation, the analysis has already been automatically conducted with the outlier treatment described in step 2. Step 3 is really a review of the precision results that have been previously obtained from the revision 2 database. The automatic calculation procedure of the interlinked Tables B.1 to B.6 produces the new precision results once either outlier option 1 (deletion) or option 2 (replacement) have been selected and the deletion and replacement operations completed. Analysis step 3 is the end of the precision calculations when outliers have been found at both the 5 % and 2 % significance levels. The results for either Table B.6-R2-OD or Table B.6-R2-OR are used to generate a precision table for the test method under review. Refer to Clause 12 for the appropriate format for a precision table and the appropriate text for the precision clause.

11 Level 2 precision: Analysis of results obtained when testing carbon blacks

11.1 Background on level 2 precision

The evaluation of test methods for the carbon black manufacturing industry shall be conducted by the procedures described in this clause for the typical uniform level experimental design. These procedures differ from the requirements set forth in the level 1 precision procedure as follows:

a) the number of replicates in each cell of the Table 1 format is specified as four;

b) the cell averages and cell standard deviations are reviewed for potential outliers by a procedure that differs from the procedure specified for level 1 precision;

c) special calculations are made to select the mode of precision expression (absolute or relative) that is most free of influence of the level (magnitude) of the measured property on the reported precision value.

The terminology set forth in Clause 3 of this Technical Report shall apply to the procedures for this level 2 precision. Frequently in the carbon black industry and elsewhere, the word “sample” is used as a synonym for the word “material” in the discussion of interlaboratory testing, i.e. a type or grade of carbon black used in an ITP is frequently referred to as a “sample”. This can be a source of confusion and is not consistent with the terminology of this Technical Report. To avoid confusion, the terms “material” and/or “target material” shall be used for what is tested (e.g. a series of different grades of carbon black) and in the process of organizing, reporting on and discussing interlaboratory test programmes and the precision parameters calculated from such programmes.

Selection of materials and initial data recording

The number of materials (or target materials), which will normally be different grades of carbon black, shall be selected as recommended in Clause 6. It is recommended that at least five materials be selected for any ITP. This number of materials provides at least four degrees of freedom in determining the coefficient of determination as described in 11.4.

Tests on the selected materials (or target materials) shall be conducted in accordance with the specified test method to produce two test results on each of two separate “test” days for a total of four test results. All testing shall be conducted on the same test machine or apparatus. A test result is the median or average of the number of determinations specified by the test method. For each material, the data values are recorded in an initial data format as indicated in Table 4. Each set of four values constitutes one cell of the general data tabulation as specified in the level 1 precision Table 1 format. However, for carbon black testing, a different final data tabulation is used as given by Table 5, a format that contains results for all materials in the ITP, as obtained from calculations (see 11.3) on the data for each material in the Table 4 format.

11.2 Data review and calculations

After a series of tables in Table 4 format have been prepared, one for each material and each laboratory, the next step is to use the data of each table to calculate a cell average and a cell standard deviation for each material/laboratory combination or cell. The results of these calculations are recorded in Table 5 format. On a material by material basis, the cell averages of Table 5 are reviewed for any potential outliers using the h statistic, and the cell standard deviations are reviewed for any potential outliers using the k statistic. Outliers are determined on the basis of a 5 % significance level for h(crit) and k(crit). Although both the cell average and the cell standard deviation of Table 5 each contain two undifferentiated components of variation, between tests/between days and between tests/within days, the h and k statistic procedure serves a useful purpose to detect any potential outliers on these special cell values.

The review process for carbon black or level 2 ITP testing is based on the premise that a substantial number of laboratories participate in the ITP, i.e. a number greater than 20. For each material in the Table 5 format, calculate the h-value and k-value for each cell (or laboratory) by the procedure specified in Annex A. A value for h(crit) and k(crit) at the 5 % significance level, is selected from Table A.1. The calculated h-values and k-values are reviewed to determine if any are greater than h(crit) or k(crit). The rejection process is conducted on the basis of the following rules:

a) If there are no calculated h-values or k-values greater than h(crit) or k(crit), all cell averages and/or standard deviations are retained.

b) If there is only one h-value or k-value greater than h(crit) or k(crit), reject the cell average or standard deviation.

c) If more than one h-value is greater than h(crit) and more than one k-value is greater than k(crit), the rejection process proceeds as follows:

   1) if there are 20 or fewer laboratories in the ITP, reject only one cell average or cell standard deviation per material, the greatest calculated h- or k-value,

   2) if there are greater than 20 laboratories in the ITP, and there are several h-values and/or k-values greater than the respective h(crit) and k(crit), reject cell averages and/or cell standard deviations, starting with the highest calculated h- and k-values and proceeding downward until the number of remaining laboratories is 20, and use this as the database for precision determination.

If any outliers are rejected, the issue of blank cells needs to be addressed. Refer to Annex B if the spreadsheet algorithms described in this Technical Report are used.

11.3 Expressing the precision determined for carbon black testing

Calculate the precision parameters r, R, (r) and (R) using the formulae specified in Annex B. The calculations shall be on the database after any potential outlier rejection and after applying the recommended procedures for missing cell values as discussed in Annex B. Plot the values of R and (R) vs M or $\bar{Y}_{AV}$ (the mean value of the material property measured) for all materials in the ITP. Perform a least-squares regression for both relationships and record the coefficient of determination, designated Cd, for each parameter, R and (R).

Selectfor the mode of precision expression, the parameter R or (R) with the lowest value of Cd. This establishes which of the two modes of expression has the least relationship to the level of the measured property or, inversely, which parameter is the most independent of the measurement level. This lowest Cd, or most independent parameter, is to be used to prepare a final precision table in the format indicated by Table 6. The selected mode of expression applies to both repeatability and reproducibility. Follow the rules for expressing precision outlined in Clause 12 of this Technical Report.

12 Format for level 1 and level 2 precision-data table and precision clause in test method standards

12.1 Precision-data table

Precision is expressed in summary form in a Table 6 format. Each summary precision-data table should have a heading to indicate:

— whether a level 1 or level 2 precision procedure was used;

— the type of precision, type 1 or type 2, used (see 5.3);

— the property measured and its measurement units.

For each material tested, the following shall be recorded:

a) the material identification;

b) the mean level of the measured property;

c) the repeatability standard deviation, s_r;

d) the repeatability, r, in measurement units;

e) the relative repeatability, (r), in percent of the mean level;

f) the reproducibility standard deviation, s_R;

g) the reproducibility, R, in measurement units;

h) the relative reproducibility, (R), in percent of the mean level;

i) the number of laboratories in the final database used to determine the precision.

If there are no outliers, the value for item i) above is the number of laboratories in the original database. If outliers are found and option 1, deletion, is used, the number will be less than the number in the original database. If option 2, outlier replacement, is chosen, the number of laboratories that did not have outliers replaced should be indicated in this column with parentheses round the number. Explain this with a footnote to the table.

The calculation of pooled or average values is recommended only if the values for r and R are roughly equal for all materials. When there is a substantial difference in precision among several materials, a pooled or average precision has very little meaningful value or applicability. The precision-data table should also contain, as footnotes, an explanation of the table symbols used.

12.2 Precision clause

The results of the precision determination should be displayed in a clause in the test method standard entitled “Precision and bias”. The concept of bias is explained in Clause 3. The one or more paragraphs or subclauses should contain information on the following issues concerning the ITP and the precision determined.

A statement that the precision ITP was conducted in accordance with ISO/TR 9272 and the year the ITP was conducted. A statement that the reader should refer to ISO/TR 9272 for terminology and other details of the precision determination.

A caveat statement that the precision determined by the ITP may not be applied to acceptance or rejection testing of any group of materials or products without documentation that the results of the precision determination actually apply to the products or materials tested.

A statement giving:

a) the level of the precision, i.e. level 1 or level 2;

b) the type of precision, type 1 or type 2;

c) the number, p, of laboratories participating in the ITP;

d) the number of materials (or target materials) used, q, and a description of the materials;

e) the number of within-laboratory replicates, n;

f) the time span for the repeatability or within-laboratory replicates (hours, days);

g) the definition of a test result (average, median of a certain number x of determinations, or individual measurement);

h) the option chosen for outlier treatment (deletion or replacement);

i) any unusual features of the ITP.

A table of precision results, as described in 12.1 above, should be part of the clause. Ensure that the table (inserted into the test method standard in Table 6 format) gives the final number of laboratories (with original data) that remained after outlier deletion or replacement. Some comments on the outcome of the results should be given.

Generic statements on repeatability and reproducibility should be part of the precision clause, using the recommended text set forth below. A 95 % confidence level (p = 0,05) applies to these generic statements. "Table X" has been used in the statements to designate the final table as inserted into the test method standard.

— *Repeatability* — The repeatability, or local domain precision, of this test method has been established by the values given in Table X for each of the materials listed in the table. If calculated, pooled repeatability values are also listed in the table. Two single test results (obtained by the proper use of the test method specified in this International Standard) that differ by more than the tabulated values of r, in measurement units, and, if listed, (r), in percent, shall be considered suspect, i.e. to have come from different populations. Such a decision suggests that some appropriate investigative action be taken.

— *Reproducibility* — The reproducibility, or global domain precision, of this test method has been established by the values given in Table X for each of the materials listed in the table. If calculated, pooled reproducibility values are also listed in the table. Two single test results obtained in different laboratories (by the proper use of the test method specified in this International Standard) that differ by more than the tabulated values of R, in measurement units, and, if listed, (R), in percent, shall be considered suspect, i.e. to have come from different populations. Such a decision suggests that some appropriate investigative action be taken.

Bias is defined in terms of "bias deviation", a deviation of a measured value from a true or accepted reference value. Bias is not addressed in this Technical Report since, for essentially all the test methods that will be evaluated for precision, the determination of bias is not possible because no reference or true value exists or may be determined. For all such test methods, a statement should be included, as the last item in the precision clause, stating that bias has not been determined. Using the word bias as a synonym for bias deviation, the suggested statement text is as follows.

— *Bias* — Bias is the difference between a test value and a reference or true value. Reference values do not exist for this test method; therefore bias cannot be determined.

12.3 Report on the precision determination ITP

A full report on the precision determination should be given for any ITP. This is a comprehensive report of all ITP details, not the report that each participating laboratory prepares and returns as part of the ITP. This full report should contain information on the details of the organization and execution of the programme as follows:

a) the organization committee, where located, co-ordinator, dates of ITP;

b) the level of precision: level 1 or level 2;

c) the type of precision: type 1, type 2;

d) the number of laboratories, p (list their names without connection to the ITP lab number);

e) the number of materials or target materials, q, plus a description of these;

f) the definition of a test result, the number of replicates, n, and the time span for repeatability;

g) information on the technicians who conducted the testing: one or more, any special details;

h) details of the preparation of the materials and how homogeneity was documented;

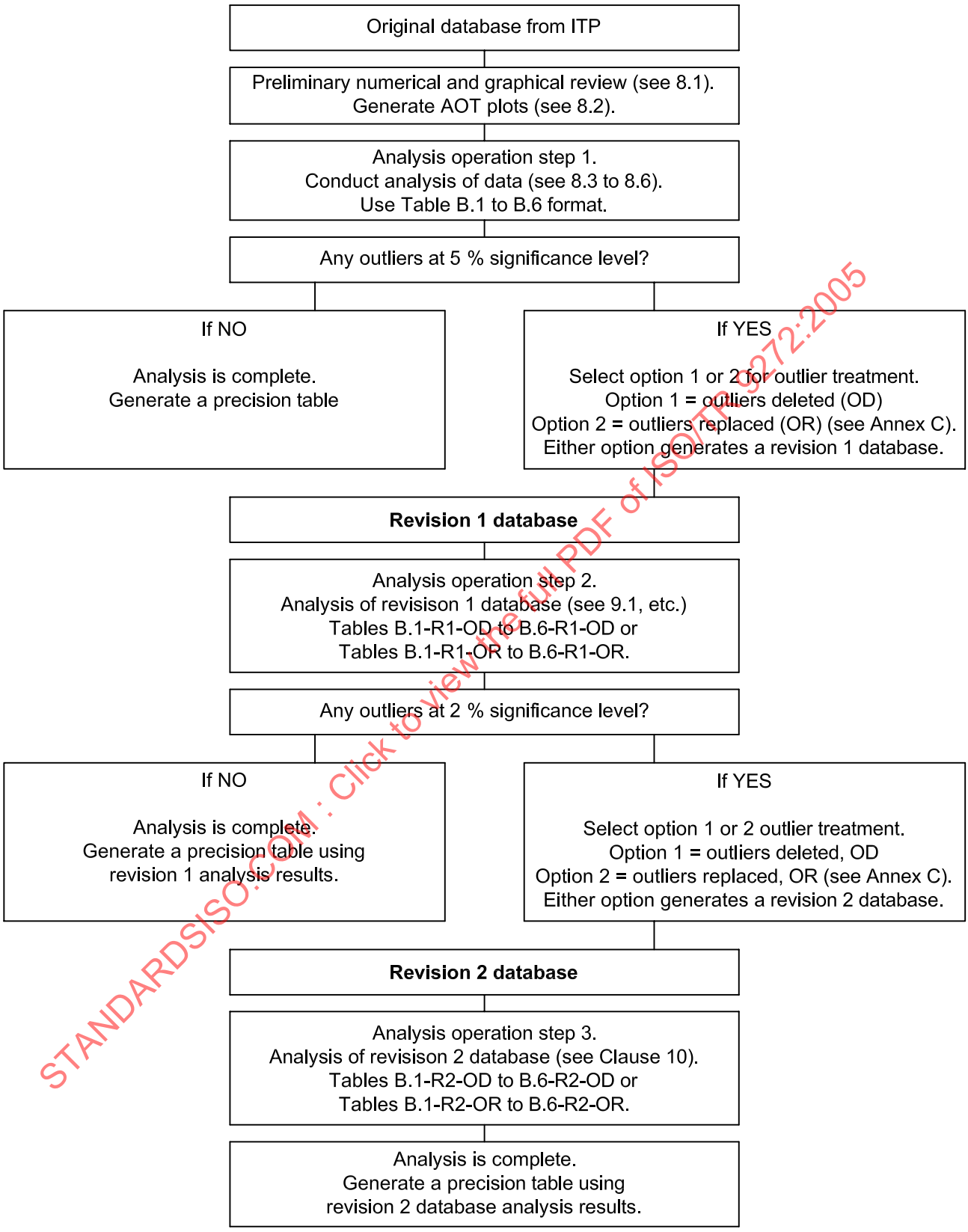
i) details on packaging and delivery of materials to the ITP participants;

j) copies of all ITP data reports from each participating lab;

k) the ITP analysis report, with all tables as designated in Annex E, a full description of all analysis steps, the options chosen for outlier rejection, plus all other required comments;

l) the table of precision results, plus any comments on the outcome;

m) a draft of the precision clause for inclusion in the test method standard.



NOTE See example of precision calculation in Annex D for tables with data.

Figure 1 — Decision tree diagram for ITP level 1 data analysis

Table 1 — Level 1 precision — Basic data[a]

Laboratory, $L(i)$	Material, $M(j)$					
	1	2	3	4	…	q
1						
2			Y_{ijk}			
3						
4						
…						
p						

Notation used:

There are a total of p laboratories: $L(i)$ = 1, 2, 3, ... p.

There are a total of q materials or levels: $M(j)$ = 1, 2, 3, ..., q.

There are a total of n replicates per cell: A cell = each combination of $L(i)$ and $M(j)$; normally n = 2.

Y_{ijk} = a single test result, where k = 1, 2, … $n(ij)$ and n normally = 2; see cell (2,3) in table for example.

Cell (ij): each cell contains n test result values.

[a] Table layout for uniform level ITP.

Table 2 — Level 1 precision — Cell averages[a]

Laboratory, $L(i)$	Material, $M(j)$					
	1	2	3	4	…	q
1						
2			Avg Y_{ijk}			
3						
4						
…						
p						

Notation used:

There are a total of p laboratories: $L(i)$ = 1, 2, 3, … p.

There are a total of q materials or levels: $M(j)$ = 1, 2, 3, …, q.

There are a total of n replicates per cell: A cell = each combination of $L(i)$ and $M(j)$; normally n = 2.

Avg Y_{ijk} = average of n test results.

[a] Table layout for uniform level ITP.

Table 3 — Level 1 precision — Cell standard deviations[a]

Laboratory, $L(i)$	Material, $M(j)$					
	1	2	3	4	…	q
1						
2			Std dev Y_{ijk}			
3						
4						
…						
p						

Notation used:

There are a total of p laboratories: $L(i)$ = 1, 2, 3, … p.

There are a total of q materials or levels: $M(j)$ = 1, 2, 3, …, q.

There are a total of n replicates per cell: A cell = each combination of $L(i)$ and $M(j)$; normally n = 2.

Std dev Y_{ijk} = standard deviation of cell (ij) for n test results.

[a] Table layout for uniform level ITP.

Table 4 — Initial data format for each material (Level 2 precision — Carbon black testing)

Date	Material, $M(j)$		Operator or technician
	Test result 1	Test result 2	
Day 1	xxx	xxx	xxxxx
Day 2	xxx	xxx	xxxxx
See notes to Table 5.			

Table 5 — Format for interlaboratory data (Level 2 precision — Carbon black testing)

Lab number	Material 1		Material 2		Material q	
	Cell avg	Cell std dev	Cell avg	Cell std dev	Cell avg	Cell std dev
1	xx	xx	xx	xx	xx	xx
2	xx	xx	xx	xx	xx	xx
…	xx	xx	xx	xx	xx	xx
p	xx	xx	xx	xx	xx	xx

Note 1 Materials are typically different grades or types of carbon black.

Note 2 The data in Table 4 (for each material) constitutes a “cell”, i.e. avg and std dev are calculated for four data values.

Table 6 — Example of level 1 and 2 precision table organization (Level 1 or 2 and type 1 or 2[a] — Precision for ISO XXXXX measured property = xxxxxx, in xx[b])

Material	Mean level	Within laboratory			Between laboratories			No. of labs[c]
		s_r	r	(r)	s_R	R	(R)	
A								
B								
C								
D								
Pooled or avg values								

Notation used:

s_r = within-laboratory standard deviation (in measurement units);

r = repeatability (in measurement units);

(r) = repeatability (in percent of mean level);

s_R = between-laboratory standard deviation (for total between-laboratory variation in measurement units);

R = reproducibility (in measurement units);

(R) = reproducibility (in percent of mean level).

See text of precision clause for discussion of precision results given in this table.

[a] Indicate the level of precision (1 or 2) and the type of precision (1 or 2) in the table heading.

[b] ISO XXXXX = reference number of test method standard; xxxxxx = property measured; xx = units of property.

[c] List number of labs in final database. Also list the option chosen: if option 2, indicate number of labs in parentheses ( ).

Annex A
(normative)

Calculating the h and k consistency statistics

A.1 General background

The test results of a typical interlaboratory test programme, when placed in a Table 2 and Table 3 format, may well contain cell values that appear to be outliers, they do not agree with the values obtained for other corresponding cells in either respective tables. It is necessary to review the data and make a decision on how to treat these outliers. This should identify any one, two or more potential outliers that have substantial deviations from the overall mean for a particular material in the database. Outlier treatment consists of rejection of all identified outliers and then using one of two options to address the particular outliers so identified. Option 1 is the deletion of the outliers to generate a reduced-size database. Option 2 is the replacement of the outliers by a procedure that maintains the character of the distribution of the non-outlier data.

Both the level 1 and level 2 precision clauses of this Technical Report use two particular parameters, called *consistency statistics*, to reject potential outliers, the h and k values as developed by J. Mandel (see 7.6 of ISO 5725-2:1994). The h statistic is a parameter used to review the between-laboratory cell averages for potential outliers and the k statistic is a parameter used to review the between-laboratory cell standard deviations (or ranges) for potential outliers. In distinction to most outlier rejection procedures that address only those extreme values that appear to be outliers, the h and k consistency statistic procedure calls for a calculation of an h and a k value for all laboratories (all cell values) for each material or q level in any ITP. After this calculation, step the subsequent outlier identification and rejection technique makes use of all the calculated values.

A.2 Defining and calculating the h statistic

A.2.1 The h-value

The between-laboratory "cell average" consistency statistic, h, is calculated using the cell averages (or means) for all laboratories and is defined as follows for each material or q level in the ITP:

$$h = d/s(Y_{AV}) \tag{A.1}$$

where

d = $[Y_{AV}(i) - \bar{Y}_{AV}]$;

$Y_{AV}(i)$ = individual cell average for laboratory (i);

$\bar{Y}_{AV}$ = average of *all* cells, for any material;

$s(Y_{AV})$ = standard deviation of cell averages for any material or q level across all laboratories.

The h-value is the ratio of the deviation, d, of each individual laboratory cell average from the overall cell average for all laboratories, divided by the standard deviation among the cell averages across all the laboratories. The h-value may be considered as a standardized variate (or z-function) with a mean of zero. Large h-values (+ or –) indicate substantial discrepancy from the overall zero average in multiples of the $s(Y_{AV})$ standard deviation.

A.2.2 Calculating critical h-values

After an h-value is calculated for each laboratory for each material, the values are reviewed to determine if any of the calculated h-values exceed a certain critical value. If a calculated h-value exceeds a critical h-value, designated h(crit), at some selected probability or significance level, the h-value in question is considered to represent an outlier and the value for the cell that generated the h-value, is identified for outlier treatment. The value of h(crit) depends on the number of laboratories in the ITP and for any probability or significance level, it may be calculated by:

$$h(\text{crit}) = (p - 1)t/[p(t^2 + p - 2)]^{1/2} \quad \text{(A.2)}$$

where

p = number of laboratories in the ITP;

t = Student's t at selected significance level, with df = $(p - 2)$, a 2-tailed value;

df = number of degrees of freedom.

A.3 Defining and calculating the k-statistic

A.3.1 The k-value

The "cell standard deviation" consistency statistic, k, is an indicator of how the within-laboratory individual cell standard deviation for any selected laboratory, compares to the overall (or pooled across all laboratories) "cell standard deviation". The usual approach to tests of significance for variability statistics is the use of the F-ratio, a ratio of two variances. However the k-value is expressed as a ratio of two standard deviations since it is easier to comprehend this ratio when reviewing data. The k-value is developed as follows.

In the usual F-ratio approach, the significance of any individual cell-variance compared to the pooled variance of all the cells (for any material) *excluding* the one cell being tested is given by:

$$F = s_i^2/[\Sigma s_{(p-i)}^2/(p - 1)] \quad \text{(A.3)}$$

where

s_i^2 = cell variance being tested for potential significance, laboratory (i);

$\Sigma s_{(p-i)}^2$ = sum of cell variances, excluding cell (i);

p = the number of laboratories in the ITP.

The k-value is defined by Equation (A.4) and is calculated for each material by:

$$k = s_i/s_r \quad \text{(A.4)}$$

where

s_i = cell standard deviation for laboratory i;

s_r = pooled cell standard deviation (across all laboratories) [this is the initially calculated repeatability standard deviation, see Equation (A.5) below].

A.3.2 Calculating critical k-values

For the purposes of calculating critical k-values, designated k(crit), the following development is presented. The repeatability variance is given by Equation (A.5):

$$s_r^2 = [\Sigma s_{(p\text{-}i)}^2 + s_i^2]/p \quad \text{(A.5)}$$

Combining Equations (A.3), (A.4) and (A.5) gives Equation (A.6):

$$k = \{[p/(1 + (p - 1))/F]\}^{1/2} \quad \text{(A.6)}$$

The number of degrees of freedom, df, for F in Equation (A.6) is $(n - 1)$ for the numerator and $(p - 1)(n - 1)$ for the denominator, where n = number of replicates per cell. Equation (A.6) may be used to calculate k(crit) for any values of p and n, at a selected significance level, by reference to the critical F value at the indicated df for the numerator and denominator.

A.4 Identification of outliers using the critical h and k values

When all the h and k values have been calculated using Equation (A.1) and Equation (A.4), respectively, and tabulated for any database, they are reviewed to determine if any of the calculated h and k values exceed the critical h and k values.

Table A.1 gives the 2 % and 5 % significance level (or p = 0,02, p = 0,05) critical values for both h and k, for various numbers of laboratories, p = 3 to 30, and cell replicates, n = 2, 3 or 4. This is used for the two-step procedure for reviewing the database for potential outliers as described in Clauses 8 and 9.

NOTE n = number of replicates per cell within each laboratory for each material or level (data for 5 % significance level taken from ISO 5725-2:1994).

Table A.1 — Critical h-values and k-values at 2 % and 5 % significance level

Number of labs, p	5 % critical h-value	5 % crit k-value for p and n			Number of labs, p	2 % critical h-value	2 % crit k-value for p and n		
		n = 2	n = 3	n = 4			n = 2	n = 3	n = 4
3	1,15	1,65	1,53	1,45	3	1,15	1,69	1,59	1,52
4	1,42	1,76	1,59	1,50	4	1,47	1,85	1,68	1,59
5	1,57	1,81	1,62	1,53	5	1,67	1,94	1,74	1,67
6	1,66	1,85	1,64	1,54	6	1,80	2,00	1,77	1,65
7	1,71	1,87	1,66	1,55	7	1,89	2,04	1,79	1,67
8	1,75	1,88	1,67	1,56	8	1,95	2,07	1,80	1,68
9	1,78	1,90	1,68	1,57	9	2,00	2,09	1,83	1,69
10	1,80	1,90	1,68	1,57	10	2,00	2,11	1,84	1,70
11	1,82	1,91	1,69	1,58	11	2,07	2,12	1,84	1,70
12	1,83	1,92	1,69	1,58	12	2,09	2,13	1,85	1,71
13	1,84	1,92	1,69	1,58	13	2,11	2,14	1,86	1,72
14	1,85	1,92	1,70	1,59	14	2,13	2,15	1,86	1,73
15	1,86	1,93	1,70	1,59	15	2,14	2,16	1,87	1,73
16	1,86	1,93	1,70	1,59	16	2,15	2,16	1,87	1,73
17	1,87	1,93	1,70	1,59	17	2,16	2,17	1,87	1,73
18	1,88	1,93	1,71	1,59	18	2,17	2,18	1,88	1,73
19	1,88	1,93	1,71	1,59	19	2,18	2,18	1,88	1,74
20	1,89	1,94	1,71	1,59	20	2,19	2,18	1,88	1,74
21	1,89	1,94	1,71	1,60	21	2,20	2,18	1,88	1,74
22	1,89	1,94	1,71	1,60	22	2,20	2,19	1,88	1,74
23	1,90	1,94	1,71	1,60	23	2,21	2,19	1,89	1,74
24	1,90	1,94	1,71	1,60	24	2,21	2,19	1,89	1,74
25	1,90	1,94	1,71	1,60	25	2,22	2,19	1,89	1,74
26	1,90	1,94	1,71	1,60	26	2,22	2,20	1,89	1,74
27	1,91	1,94	1,71	1,60	27	2,23	2,20	1,89	1,74
28	1,91	1,94	1,71	1,60	28	2,23	2,20	1,89	1,74
29	1,91	1,94	1,72	1,60	29	2,23	2,20	1,90	1,74
30	1,91	1,94	1,72	1,60	30	2,24	2,20	1,90	1,74

Annex B
(normative)

Spreadsheet calculation formulae for precision parameters — Recommended spreadsheet table layout and data calculation sequence

B.1 Calculation formulae

B.1.1 General

When a dedicated computer programme is not available to calculate precision, the repeatability and reproducibility may be calculated using typical spreadsheet procedures and algorithms. The final precision calculations involve a series of sums or totals. The calculation formulae are given in this clause. In Clause B.2, a recommended spreadsheet table layout is presented that facilitates the calculations. Clause B.3 gives some recommendations for setting up the table sequence and conducting the analysis. Figure 1 presents a decision tree diagram that gives guidance on the sequence of steps. Recall that p = number of laboratories in the ITP.

NOTE The calculations were set up for this annex using Lotus 123. It is assumed that any spreadsheet programme can be used; however, some of the particular algorithms may be slightly different than indicated in this annex.

B.1.2 Uniform level ITP design, n = 2

All laboratories in the ITP test all materials; each material has n = 2 replicates per cell and the summations are over all laboratories. A cell contains the n replicate values for each "laboratory/material" combination in a Table 1 format as given in the main body of the Technical Report. A replicate is a "test result", i.e. the mean or median value as specified by the test method.

$T_1 = \Sigma Y_{AV}$, where Y_{AV} is the cell average for laboratory i (B.1)

$T_2 = \Sigma (Y_{AV})^2$ (B.2)

$T_3 = \Sigma w^2$, where w = range of cell values, laboratory i (for n = 2 only) (B.3)

$T_4 = \Sigma s^2$, where s = cell standard deviation, laboratory i (B.4)

For the calculations outlined below use either T_3 or T_4. Equation (B.5) gives the repeatability standard deviation squared or variance, s_r^2:

$$s_r^2 = T_3/2p = T_4/p \quad \text{(B.5)}$$

Equation (B.6) gives the variance between laboratories s_L^2:

$$s_L^2 = \{[pT_2 - (T_1)^2]/p(p-1)\} - [s_r^2/2] \quad \text{(B.6)}$$

Since this between-laboratory variance does not contain the within-laboratory variance component, it is corrected for this by adding the within-laboratory variance. The variance that contains both the between-laboratory and the within-laboratory components is the reproducibility variance given by Equation (B.7):

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

$M_{AV} = \bar{Y}_{AV} = T_1/p$, material average for all laboratories (B.8)

The repeatability r and the reproducibility R are given by Equations (B.9) and (B.10):

$$r = 2{,}83(s_r^2)^{1/2} = \text{Repeatability} \quad \text{(B.9)}$$

$$R = 2{,}83(s_R^2)^{1/2} = \text{Reproducibility} \quad \text{(B.10)}$$

B.1.3 Uniform level ITP design, $n > 2$

For any ITP with n equal to more than two (2) but with a constant number of cell replications for each material/laboratory combination, the computation equations are identical to Equations (B.1) to (B.10) with the following exceptions: (1) the value of n is used in place of 2 in the last term of Equation (B.6) and (2) T_3 is not calculated, the value for s_r^2 being obtained by means of the T_4/p expression in Equation (B.5).

B.1.4 Non-uniform level design

For any ITP with an unequal number of replicates per cell:

$$T_5 = \Sigma[n_i(Y_{AV})_i] \quad \text{(B.11)}$$

where n_i = number of replicates in cell i and $(Y_{AV})_i$ = average for cell i

$$T_6 = \Sigma[n_i(Y_{AV})_i^2] \quad \text{(B.12)}$$

$$T_7 = \Sigma n_i \quad \text{(B.13)}$$

$$T_8 = \Sigma n_i^2 \quad \text{(B.14)}$$

$T_9 = \Sigma(n_i - 1)s_i^2$, where s_i^2 is variance for cell i

$$s_r^2 = T_9/(T_7 - p) \quad \text{(B.16)}$$

$$s_L^2 = \{\, | [T_6T_7 - T_5^2]/[T_7(p-1)] | - s_r^2 \}\{ [T_7(p-1)]/(T_7^2 - T_8) \} \quad \text{(B.17)}$$

$$s_R^2 = s_L^2 + s_r^2 \quad \text{(B.18)}$$

$$M_{AV} = \bar{Y}_{AV} = T_5/T_7 \quad \text{(B.19)}$$

B.2 Table layout for spreadsheet calculations

B.2.1 Table organization

This clause contains a listing of all the tables required with a brief description of the linking between the tables to permit all calculations to be automatically performed to give the values for r and R, once all tables have been set up and the basic table of data has been generated. The layout is for a uniform level design with $n = 2$. The description is directed mainly to analysis step 1. If outliers are found for step 1, then the calculation operations of step 2 and perhaps step 3 will be required. For a full understanding of these two additional steps, it is necessary to completely review the precision determination example in Annex D, which gives instructions for these additional calculations.

For this annex, the tables will be identified as B.1, B.2, etc. These correspond to tables in Annex D designated D.1, D.2, etc. Starting with Table B.2, the tables differ from the format of Tables 2 and 3 in the main body of the Technical Report in the use of a double or side-by-side data display format. This double table set-up permits rapid viewing of the data and calculated parameters as data is entered and processed.

There are potentially three analysis operation steps for any ITP. The number of steps actually required depends on the quality or uniformity of data in the database. If outliers are found, then a second and perhaps a third analysis step will be required. Each of these analysis operations should be conducted on a separate

"sheet" or tabbed page of the computer spreadsheet programme. This facilitates the analysis and avoids confusion. If outliers are found for any analysis operation, there are two options to continue with the analysis:

a) *Outlier option 1: Removal by cell deletion* — The simplest option for outliers is the deletion of the outlier from the database as expressed in a Table B.1 format. See B.3.2 below for more details on this.

b) *Outlier option 2: Cell replacement values for outliers* — If this option is chosen, cell replacement values are calculated by the procedures described in Annex C. This option involves more work but it may be the only option for a limited ITP database with a small number of laboratories.

The three potential analysis steps are described in Clauses 8, 9 and 10. If there are no outliers, only analysis step 1 is used. If outliers are present, analysis steps 2 and 3 may be required depending on the extent of outliers in the database. The table description outlined below is for analysis step 1, the first set of calculations for any ITP (see Clause 8), prior to the possible rejection of any incompatible values as outliers.

The word "cell" is used in two different contexts: it is the intersection of a row with a column in a computer spreadsheet; it is also, for any ITP, the combination of a laboratory and a material as in Table 1 in the main body of the Technical Report. The word cell will be italicized when it refers to a computer spreadsheet. In many cases, there is a dual usage or meaning (a Table 1 cell is also a spreadsheet cell).

Although, as described below, Table B.1 may contain blank table cells, all table cells that have data must contain the number of replicate values characteristic of the design of the ITP. For most level 1 precision ITPs, n = 2 and each cell must contain both values. The original database generated in some ITPs may be one where one or more laboratories report only one value for a particular material, i.e. they did not fully participate and only supplied partial data. The partial data for such a laboratory cannot be used since the spreadsheet programme as set up in this annex requires that all Table B.1 cells (for analysis step 1, 2 or 3) have the required number of replicates.

Table number and name	Table description
B.1 — Basic data from ITP	This is the basic Table 1 format (as discussed in main body of Technical Report); rows = laboratories; columns in replicate 1, 2 format = materials. Two spreadsheet columns are required for each material. Each (double column) ITP cell contains two test results. In generating all tables beyond Table B.1, preserve the same row/column identification for laboratories and materials.
B.2 — Cell averages, averages squared	This is a dual table, cell averages in left side and cell averages squared in the right side, each side preserving the laboratory/material row vs column format of Table B.1. Totals are calculated for each material column; Cell average totals = T_1, cell average squared totals = T_2. Also calculate, for the left section, the grand cell average (all laboratories) and the variance and standard deviation of the cell averages (across all laboratories). NOTE Do not truncate significant figures for any total in any of these tables. Retain four significant digits for all calculations.
B.3 — Cell avg deviations, d- and h-values	A dual table: cell deviations d, d = cell i – (all-cell avg); in the left section and cell h-values in the right section. Review the cell h-values and indicate all that are significant at the 5 % level by making value bold and italic. See Annex A for calculation of h-values.
B.4R — Cell ranges and ranges squared	A dual table: cell ranges on left and cell ranges squared on the right. For each left-hand-side cell, the cell range may be obtained from Table B.1 using an appropriate @IF function to convert those negative difference values to positive values for the cells in Table B.4R. It is useful to obtain the average range for each material. Calculate the cell squared totals T_3 for each material.

B.4S — Cell standard deviations and variances	A dual table, with cell standard deviations on the left and cell variances on the right. It is convenient to calculate the pooled variance for each column of standard deviations; place these at the bottom of each left-side column. Calculate the total for the cell variances; place these values at bottom of each column of variances on the right side. Total of cell variances for each material = T_4.
B.5 — Cell k-values	A single table for cell k-values. See Annex A for calculation of k-values. For each k-value that equals or exceeds the 5 % significance level value, indicate by making the value bold and italic.
B.6 — Calculations for precision	A table giving the sequence of calculations for precision. The calculations are performed for each material, thus a column is required for each material. Insert values for T_1, T_2 and either T_3 or T_4 by means of spreadsheet linking to the appropriate preceding tables. Calculation 1 is a calculation of s_r^2, using either T_3 or T_4. Calculation 2 determines s_L^2 using T_1 and T_2. Calculation 3 is a calculation of s_R^2, using s_L^2 and s_r^2. Calculation 4 determines r and calculation 5 determines R. At the bottom of Table B.6, material means (averages) are given as well as the standard deviations s_r and s_R. Also listed is a sub-table for step 1 and, if used, step 2 outlier review at the 5 % and 2 % significance levels. This sub-table indicates the outlying laboratories for both h and k.

NOTE The values for n and p in Table B.6 can either be active or be a fill-in format. The value of n will be 2, but p will vary depending on the number of cells for laboratories deleted for either h or k values. For active p values, a count function should be performed for the cell values in Table B.5-R1-OD or B.5-R2-OD (see B.3.1) for each material. This counts the number of laboratories after deletions of both h and k. The count result enters the appropriate cell of Table B.6. For a fill-in operation, the values in Table B.6 must be inserted manually.

B.2.2 Setting up the spreadsheet

Begin on sheet 1 of a spreadsheet programme. This will be used for analysis step 1. The first set of calculations is for the original database. For any subsequent analysis operations with a complete set of recalculations after outliers are removed from the database or outliers replaced, one or more additional computer programme sheets will be used. Calculations are facilitated if each table occupies a single screen area, using the "page down" command to go to the next table. Refer to the Annex D example for more details on steps 2 and 3.

a) *Link Table B.2 to Table B.1* — For lab 1 and material 1, use the *average* @function to calculate the average for cell 1 in Table B.2, using the corresponding two adjacent (spreadsheet) *cells* in row 1 of Table B.1 (for lab 1 and material 1) as the argument spreadsheet range. Repeat for all table cells. After this is completed, calculate the cell average squared values for all cells on the right side of Table B.2 by the appropriate spreadsheet squared function algorithm using the left-hand-side cell averages.

b) *Link Table B.3 to Table B.2* — For material 1, using the appropriate spreadsheet algorithm, subtract from each laboratory cell average on the left side of Table B.2 the overall cell average. This gives d. Divide each calculated d by the standard deviation of all cell averages to give the calculated h-value. Repeat for all materials. The calculation output for h-values is entered into the corresponding (row/column) *cell* in the right-side section of Table B.3.

c) *Link Table B.4 to Table B.1* — For lab 1 and material 1, calculate the standard deviation for cell 1 in Table B.4 by means of the @function for standard deviation, using the corresponding two adjacent *cells* in row 1 of Table B.1 (lab 1 and material 1) as the argument spreadsheet range. Repeat for all cells. Ensure that the divisor for the standard deviation calculation is $(n - 1)$, not n, where n = number of values for the standard deviation calculation for each material. In spreadsheet terminology, this is often designated a "sample" calculation. Using the appropriate algorithm, square each cell standard deviation value; the result is entered into the corresponding *cell* on the variance or right side of Table B.4.

d) *Link Table B.5 to Table B.4S* — For material 1, divide each individual (within) cell standard deviation by the pooled value for (within) cell standard deviations (this is the square root of the pooled or mean variance) to obtain k-values. Repeat for all materials. The k-values are entered into the corresponding *cells* in Table B.5.

e) *Link Table B.6 to Tables B.2, B.4S and/or B.4R* — For material 1, use the appropriate spreadsheet function or algorithm to bring the totals T_1, T_2, T_3 and/or T_4 into Table B.6. Repeat this for all materials. The source for each total should be the total at the bottom of each of the appropriate columns in Tables B.2, B.4S or B.4R. For calculation 1 in Table B.6, use the formula given in the table to calculate each of the parameters for all materials in the ITP. The formula should use the active values of n and p as well as the values for that material brought in from Tables B.2, B.4S or B.4R. When calculation 5 of Table B.6 is complete, the entry of values for T_1, T_2, T_3 and/or T_4 along with values for p and n (by means of their linkages to preceding tables) will produce an immediate result for all intermediate and final precision calculations in the table.

B.3 Sequence of database calculations for precision

B.3.1 Outliers in analysis step 1 (sheet 1)

As noted above, the step 1 analysis operation or set of calculations should be performed on sheet 1 of the computer spreadsheet programme. If any incompatible values are declared as outliers at the 5 % significance level, the database shall be revised in accordance with 8.4 to either delete outliers for any laboratory or insert replacements into the database for those cells that contain outliers. If any outliers are found, it is necessary to conduct analysis step 2 (sheet 2) on the revision 1 (R1) database. The calculations for analysis of the revision 1 database are facilitated by copying all of the executed Tables B.1 to B.6 on sheet 1 onto corresponding locations on sheet 2 of the spreadsheet, with all programmed calculations active, i.e. not as values. These tables on sheet 2 are now designated as (1) Table B.1-R1-OR to Table B.6-R1-OR for replaced outliers or (2) Table B.1-R1-OD to Table B.6-R1-OD for deleted outliers.

B.3.2 Outliers in analysis step 2 (sheet 2): Option 1 — Outlier deletion

All deletion operations can be facilitated by marking, on a printed-out Table B.1, all table cells that have significant h and k values. To delete data, simply delete from Table B.1 all the cells that have a 5 % significance level h or k value. Cell refers here to the ITP design, not to the spreadsheet *cells*, i.e. delete both values in each ITP design cell which occupies two spreadsheet *cells*. When this is done, the typical spreadsheet programme will give an ERR indication at several calculation cell locations in Table B.2-R1-OD to Table B.6-R1-OD. This is due to the deletion of one or more argument values in Table B.1-R1-OD and some subsequent tables as well.

ERR notations will appear in two general locations:

a) In columns as data entries that come from tables above them in the sequence of tables, i.e. values used to calculate parameters for a particular column such as averages, standard deviations, etc.

b) At the bottom of columns where averages, standard deviations, etc., were previously located. To correct the tables, start with the first table that contains a spreadsheet *cell* that has an ERR notation, and delete the ERR cell that is a *data entry*, not an ERR *cell* at the base of a column. Correcting the data entry value or *cell* will automatically correct the ERR (calculated value) at the base of the column.

The use of a spreadsheet “delete” operation for any ERR *cell* will make the *cell* in question blank. Continue this for all tables until all ERR indications are removed and replaced by blank values, not zeros. This will produce correct calculations for all parameters. Also remove from all tables any *zero cell values* that are generated by the deletions from any of the preceding tables. If they are not removed, the bottom of the table column calculations will be in error. For option 1, outlier deletion, the revised precision parameters will automatically be calculated, and will appear in Table B.6-R1-OD of sheet 2 after all ERR entries are removed.

B.3.3 Outliers in analysis step 2 (sheet 2): Option 2 — Outlier replacement

When this option is chosen, data replacement values or DRs (see Annex C for definitions on replacement values) are inserted into the cells that contain outliers. Insert into the experimental design cells of Table B.1 (individual) *cell data* replacement (test result) values, DR1 and DR2, as determined in Annex C. These will be in cells that have a significant h or k value. Correct any possible ERR occurrences, if they appear, as described in B.3.2.1 and B.3.2.2. For option 2, insertion of data replacement values or DRs, the revised precision parameters will automatically be calculated and appear in Table B.6-R1-OR of sheet 2.

B.3.4 Outliers in analysis step 3 (sheet 3)

The precision values for (sheet 2) revision 1 analysis are accepted as final if there are no outliers at the 2 % significance level.

a) If any outliers are found at the 2 % significance level, either follow the procedure cited above (for 5 % significance) to do an option 1 deletion of all outliers to generate a revision 2 OD database or select option 2 and calculate replacement values. When these are inserted into the revision 1 OR database, a revision 2 OR database is generated.

b) If outliers are found, copy the executed Table B.1-R1-OR to Table B.6-R1-OR or Table B.1-R1-OD to Table B.6-R1-OD of spreadsheet sheet 2 to spreadsheet sheet 3 with active values as above. These revision 2 tables, when completed as indicated below, will be designated Table B.1-R2-OR to Table B.6-R2-OR or the corresponding Table B.1-R2-OD to Table B.6-R2-OD. The purpose of a sheet 3 analysis is to delete or replace the 2 % significance outliers and thereby generate final revision 2 precision values.

c) Once outlier values have been deleted from any *cell* or data replacement values have been calculated (using Annex C) and inserted into the appropriate *cells* of Table B.1-R2-OR or Table B.1-R2-OD in sheet 3; the new precision values will appear in sheet 3 Table B.6-R2-OR or Table B.6-R2-OD after any ERR indications are removed. These sheet 3 Table B.6-R2-OR or Table B.6-R2-OD values are the final precision parameters, r and R, for the ITP.

B.3.5 Precision result rounding

The final precision results as given in Table B.6, Table B.6-R1 or Table B.6-R2 (with either outlier option) are transferred into a Table 6 format (see 12.1) for insertion into the test method standard. When this is done, the final precision parameters should be rounded to the number of significant digits or figures that are technically attainable in usual practice with the test method, with perhaps one more significant figure than normally employed. Excessive figures beyond this shall not be retained.

Annex C
(normative)

Procedure for calculating replacement values for deleted outliers

C.1 Introduction

If outliers are found in analysis step 1 at the 5 % significance level, there are two options. Option 1 is to delete the outliers and thereby generate a revised, or R1, database. Option 2 is to replace the outliers in a way that essentially preserves the distribution of the non-outlier data as described in more detail in Clause C.2. This annex provides the algorithms to address the replacement process when outliers are found at either the 5 % or the 2 % significance level.

Outlier option 2 (replacement) is usually the choice when outliers are found with a small database with a limited number of laboratories (ca 6 or less). Replacing outlier values, rather than deleting them, preserves the size of the database. The procedure for calculating replacement values, however, must be one that is "consistent with the observed data distribution" in the database. The replacement procedure described in this annex fulfils this objective. It consists of a determination or calculation of two types of replacement.

C.2 The replacement procedure

The replacement procedure (for either step 1 or step 2) is one that replaces outliers with realistic values. The initial operation determines replacement values for each outlier "cell average" and each outlier "cell standard deviation". The first type of replacement is designated a *parameter replacement* or PR. There are two possible types of PR as described below that might be inserted into the database. Although only one is selected, both are described in order to demonstrate the merit of the selected second type of replacement.

C.2.1 Distribution mean parameter replacement

The first possible approach for a PR is to insert into the database a value equal to the distribution or actual database *mean* of all cell values for any material. There are two types of distribution *mean*:

a) of cell averages;

b) of cell standard deviations or cell ranges.

The word "*mean*" applies to both. If only one PR is being considered and there are ten or more laboratories, this will not substantially change the nature of the distribution.

However, if two or more outliers are being replaced and the number of laboratories is much less than ten, this may narrow the distribution and thus give a falsely optimistic value of:

a) the standard deviation for the final precision results (if no further outliers are found);

b) the denominator standard deviation for the h and/or k statistics that will be used for outlier review at the 2 % significance level.

For this reason, this type of replacement is not chosen.

C.2.2 Ascending order trend (AOT) parameter replacement

The alternative approach for a PR is to use a value that substantially preserves the observed distribution as illustrated by the ascending order trend plots discussed in 8.2. This is designated an *ascending order trend* or AOT replacement or PR for a cell mean. Each AOT replacement or PR is in essence a predicted value, one that would be expected for the laboratory in question in the absence of the unexpected perturbation that generated the outlier illustrated by the "off-the-line" behaviour in the AOT plot. This AOT replacement does not narrow the observed distribution in the same sense as a distribution *mean* value replacement.

C.3 Outlier replacement categories

There are two different categories for outlier replacements: *parameter* replacements (PRs) as discussed above and *data replacement values* (DRs). After PRs have been determined for all outlier cell averages and cell standard deviations (or ranges), the next step is the calculation of DRs for each cell of Table B.1 format that contained a *parameter* outlier.

DRs are required to insert into a Table B.1 data format (to generate a Table B.1-R1-OR) to permit a recalculation of the revised precision values based on the new R1 database (see Annex B and the Table B.1 to B6 series). Once the initial basic data Table B1 is revised to generate a Table B.1-R1-OR, all the succeeding tables, B.2-R1-OR to B.6-R1-OR, are recalculated by the automatic calculation process described in Annex B. The procedures described (for this Annex C) are for uniform level designs with two cell values or $n = 2$. The procedures may be slightly amended for $n = 3$ situations. The precision example in Annex D on Mooney viscosity testing illustrates the entire AOT replacement process and the operations described in this annex as well as Annexes A and B.

C.4 PRs for outliers at 5 % significance level

Outlier values at the 5 % significance level shall be replaced using the AOT replacement procedure described in a) to c) below. These procedures apply in principle to any of three databases: the original database, the R1 database or the R2 database. The R1 and R2 databases will potentially contain PRs determined by a previous outlier replacement process.

a) *PRs: "Cell average" outliers* — For each material, visually fit a (least-squares type) straight line through the central data point region of the cell average AOT plot and extend the line to both extreme ends of the plot. Alternatively, a linear regression may be used to fit the straight line; however, do not include in the data set any questionable outlier end points. For the outlier values (low or high end of plot), determine the difference between the outlier value (plotted point) and that point on the extended line at the x-axis location of the laboratory in question. Add this difference to or subtract it from the outlier value to produce a new value that is "on the fitted line" at that x-axis location. For each outlier, this "on the line" value is the *cell average* PR for that laboratory.

b) *PRs: "Cell range" outliers* — For each material, visually fit a straight line through the central value point region of the cell range AOT plot and extend the line to the high value end of the plot. Repeat the procedure given in a) above to determine a new value on the fitted line. For each outlier, this "on the line" value is the *cell range* PR for that laboratory.

c) *PRs: "Cell standard deviation" outliers* — If cell standard deviations were calculated initially rather than cell ranges, determine a *standard deviation* PR using the same procedure as described for cell range outliers in b) above. For ITP designs that have $n = 2$, the replacement cell standard deviation (std dev) can be converted to a cell range, w, by using $w = (\text{std dev}) \times (2)^{1/2}$. In the equations listed below, a value for the range is required for calculating DRs.

NOTE The equations for calculating DRs using PRs for ranges as given below can be altered for use with standard deviations rather than ranges. For ITPs where $n = 2$, substitute the value of the range w, i.e. (std dev) × 1,414, into the equations.

C.5 DRs for outliers at 5 % significance level

After PRs have been determined for all outlier cell averages and cell standard deviations (or ranges) at the 5 % significance level, the next step is the calculation of DRs for insertion into a Table B.1 format. For the DR process, procedures are used that maintain the values not declared as outliers, at their observed values in the database. As an example, when only a replacement cell average is required (i.e. the cell range or standard deviation is not an outlier), the actual or existing cell range shall not be changed by the replacement. Also when only a replacement cell range or standard deviation is required, the existing cell average shall be maintained. There are four possible combinations of PRs that require DRs. The procedures for these are described in steps a) to d) below.

a) *Cell average outlier with non-outlier cell range* — For the two DRs for a cell average outlier, add one-half and subtract one-half of the original or *existing cell range*, ECR, to and from the PR (cell avg), using Equations (C.1) and (C.2). This gives two cell values, DR1 and DR2, that yield the replacement cell average. Insert the replacement values into the Table B.1 format database.

$$DR1 = PR(\text{cell avg}) + ECR/2 \quad (C.1)$$

$$DR2 = PR(\text{cell avg}) - ECR/2 \quad (C.2)$$

To avoid the confusion of excessive notation, all DRs (each of four categories) are identified as DR1 and DR2.

b) *Cell average outlier with cell range outlier* — For the two DRs for this situation, add one-half and subtract one-half of the AOT-plot-determined PR (cell range) to and from the PR (cell avg) using Equations (C.3) and (C.4). This gives the two new cell *data* values DR1 and DR2 that yield the replacement cell average and the replacement cell range. Insert the DRs into the Table B.1 format database.

$$DR1 = PR(\text{cell avg}) + PR(\text{cell range})/2 \quad (C.3)$$

$$DR1 = PR(\text{cell avg}) - PR(\text{cell range})/2 \quad (C.4)$$

c) *Cell range outlier with non-outlier cell average* — For the two DRs required for this situation, add one-half and subtract one-half of the AOT-determined PR (cell range) to and from the original or *existing cell average*, ECA, using Equations (C.5) and (C.6). This gives the two new cell *data* values DR1 and DR2 that yield the original cell average and the replacement cell range. Insert these into the Table B.4R format database.

$$DR1 = ECA + PR(\text{cell range})/2 \quad (C.5)$$

$$DR2 = ECA - PR(\text{cell range})/2 \quad (C.6)$$

d) *Cell range outlier with cell average outlier* — Follow the same procedure as in b) above. This gives two cell *data* values with the replacement cell average and the replacement cell range. Insert these into the Table B.1 format database.

C.6 PRs for outliers at 2 % significance level

For an analysis step 2 review of the revised or R1 database, follow the instructions in Clauses C.5 and C.6 that apply to a significance level of 2 %.

a) *PRs: "Cell average" outliers* — For each material, replot the cell average data to give a new AOT plot, using the revised data of Table B.1-R1-OR. The data in the Table B.1-R1-OR format will have new replacement values for all 5 % significance outliers. Follow the procedure described in Clause C.5 to determine the PR "cell average" for outliers at the 2 % significance level.

b) *PRs: "Cell range" outliers* — For each material, replot the cell range data in an AOT plot, using the revised data of Table B.1-R1-OR. Follow the procedure described in Clause C.5 to determine the PR "cell range" for outliers at the 2 % significance level.

c) *PRs: "Cell standard deviation" outliers* — If cell standard deviations were calculated initially rather than cell ranges, calculate a replacement standard deviation using the cell range procedure described in Clause C.5. As noted above, for ITP designs with n = 2, the replacement cell standard deviation (std dev) can be converted to a cell range, w, by using w = (std dev) × $(2)^{1/2}$.

C.7 DRs for outliers at 2 % significance level

After PRs have been determined for all outlier cell averages and cell standard deviations (or ranges) at the 2 % significance level, the next operation is the calculation of DRs for Table B.1 format. These are required to generate a Table B.1-R2-OR format to permit a recalculation of the revised precision values (repeatability, reproducibility) based on the new R2 database (see Annex B). Just as for the 5 % significance level calculations, there are four possible combinations of *parameter* outliers that require *data* replacements for an R2 database. The outliers are at the 2 % significance level and the database being considered for revision is the R1 database. After 2 % significance level outliers have been replaced (both PRs and DRs) in an R1 database, it becomes an R2 database and is used to calculate the final or terminal values of repeatability and reproducibility. Refer to the flow-sheet diagram in Figure 1.

For the four outlier combination categories discussed in Clause C.5, repeat the calculations for DRs based on PRs determined using AOT plots of the R1 database. Use the equations given in these sections.

Annex D
(normative)

An example of general precision determination — Mooney viscosity testing

D.1 Introduction

This annex presents a detailed example of a level 1 "three-step analysis" precision determination with emphasis on how outliers are detected and how the original database is revised to obtain robust precision estimates that are free of outlier effects. All precision calculations are given, starting with a basic Table 1 (or equivalent Table B.1) format, using the calculation formulae and other operations in the series of tables described in Annex B. All tables in this Annex D will have identifications analogous to their Annex B identifications but using the D designation. Thus Table D.1 in this annex is equivalent to Table B.1 in Annex B, Table D.2 is equivalent to Table B.2, etc.

Two outlier treatment options may be chosen. Option 1 is the deletion of all outliers and the calculation of precision results on the revised and reduced database. Option 2 is the replacement of outliers with AOT replacements and the calculation of precision results on the revised database. Both of these options are given in this example. Although not illustrated in this Technical Report, calculations have been conducted for this database using the alternative analysis algorithms A and S given in ISO 5725-5. A comparison of the precision results for options 1 and 2 and the ISO 5725-5 analysis is presented and the outcome is discussed. An additional feature is illustrated: the use of technical judgement by the statistical analyst to override the outcome of a particular objective outlier rejection procedure. The reasons for this are cited.

The ITP for Mooney viscosity testing was conducted in the mid-1980s using the edition of ISO 289 that existed at that time. Four materials (rubbers) were used and nine laboratories participated in the ITP. The rubbers, identified as materials 1 to 4, and some of the details of the testing are described as follows:

Material number	Material description	Test conditions
1	SBR 1712 (37,5 oil ext.)	ML1+4@100 °C
2	IIR (butyl) NIST SRM 388	ML1+8@100 °C
3	NR (natural rubber)	ML1+4@100 °C
4	SBR 1712 BMB (37,5, 65 N339)	ML1+4@100 °C

NIST = National Institute of Standards and Technology, the new name for the National Bureau of Standards in USA.

SRM = Standard reference material as developed by NIST.

BMB = Black masterbatch: 37,5 oil + 65 of carbon black N339.

Samples of each of the four materials were sent out to the nine participating laboratories and viscosity tests were conducted on two separate days one week apart. A test result was one determination (measurement) of Mooney viscosity at the indicated time and temperature. Thus for this ITP, p = 9, q = 4 and n = 2. A type 1 precision was determined with one additional operation just prior to testing: materials 1, 3 and 4, were mill-massed. Material 2, the IIR SRM, was not mill-massed since this was not specified for this reference material.

D.2 Organization of the Mooney example precision determination

The ordinary practice to determine precision for any given ITP is to use the sequence of steps as outlined in Figure 1 and discussed in the overview (see Clause 7). The detailed instructions are in Clauses 8, 9 and 10. If outliers are found for step 1, one of the two outlier options is selected and the analysis proceeds to step 2 and on to step 3, if needed, based on this decision (see Figure 1). However, to better illustrate precision determination in this example, calculations are given for both outlier options. Although outlier replacement is option 2, the calculations for this option will be demonstrated first as part 1. After that, the simpler option 1 approach of outlier deletion will be demonstrated as Part 2. The preliminary data and graphical review, given below, is not repeated for the Part 2 outlier deletion option.

D.3 Part 1: Level 1 analysis — Option 2: Outlier replacement

D.3.1 Analysis step 1 — Preliminary review

Table D.1, as set up in sheet 1 of the computer spreadsheet programme (see Annex B), is a tabulation of the original data in a format as specified in 8.3. Although it is not necessary for the analysis steps to follow, it is informative to obtain averages and standard deviations of all columns in the table and the results for these calculations are illustrated.

The next operation is to generate tables in the format of Tables 2 and 3 as outlined in 8.4 a) and 8.4 b). As previously discussed, the basic Table 2 and 3 data tabulation is combined with other tabulations and calculations in a dual table format. This dual table format is required for the full analysis and is fully described in Annex B. Thus the Table 1 format as called for in 8.4 a) is given on the left side of Table D.2 and the Table 3 data tabulation format as called for in 8.4 b) is given on the left side of Table D.4S for within-cell standard deviations, or in Table D.4R for within-cell ranges.

The graphical examination of the ITP data is conducted using Figures D.1 to D.4. Figure D.1 illustrates plots of "cell average" Mooney viscosity vs laboratory number in ascending viscosity order for materials 1 and 2 and Figure D.2 illustrates similar plots for materials 3 and 4. These plots serve a dual purpose: an initial review of the original data and a second operation to calculate the outlier option 2 AOT replacement values for outliers as described in C.2.2 in Annex C.

Figure D.1 indicates that there may be two potential outliers for material 1 — one low outlier for lab 9 and perhaps a high outlier for lab 6. These deviate from the central-region essentially linear trend line. This trend line will be used in the AOT replacement operation to be conducted later. For material 2, one high potential outlier for lab 1 is indicated. In Figure D.2, material 3 has one low potential outlier for lab 9 and material 4 has two potential outliers — low for lab 9 with a less likely high value for lab 8.

Similar plots for cell ranges in Figures D.3 and D.4 are slightly different from the cell average plots. There are no low end outliers. All low values indicate good agreement and as a result these plots have more of an initial low end curvilinear nature prior to a central linear region. Material 1 has a two potential high end cell range outliers for lab 4 and lab 1. Material 2 has no potential outliers. Materials 3 and 4 in Figure D.4 both have potential outliers for lab 4 and perhaps one for lab 9. The plots of Figures D.1 to D.4 give an overall impression of the degree of data uniformity for each of the four materials. The other features of the figures will be discussed later.

D.3.2 Precision calculations and outlier review for original database

The step 1 analysis begins by calculating the precision values of r and R for the original database. The initial calculation of r and R using the procedures set forth in Annex C establishes a starting point or foundation for comparisons of the reduction in these two parameters as outliers are deleted. Next is an examination of the database to detect any potential outliers at the 5 % significance level. Both of these operations will be conducted in parallel and described as each table in the sequence Table D.1 to Table D.6 is reviewed.

Table D.2, set up in the dual format for all four materials, has cell averages on the left and cell averages squared on the right. Two totals, T_1 for "cell averages" and T_2 for "cell averages squared" (as required for final

precision analysis in Table D.6), are obtained for each column or material in the table. Also indicated are results for the overall cell average and the variance and standard deviation for individual cell averages for all nine laboratories.

Table D.3 contains the "cell average" deviations, d, on the left and the cell h-values on the right, where for each material:

$$d = Y_{AV}(i) - \bar{Y}_{AV} \quad \text{(D.1)}$$

$$h = d/s(Y_{AV}) \quad \text{(D.2)}$$

where

$Y_{AV}(i)$ = cell i average;

$\bar{Y}_{AV}$ = average of all cell averages;

$s(Y_{AV})$ = standard deviation of cell averages (see Annex A).

The values for $\bar{Y}_{AV}$ and $s(Y_{AV})$, descriptively indicated, are found at the bottom of the left section of Table D.3. Below the right side of the table, an inset sub-table gives the h(crit) at the 5 % significance level for the indicated number of laboratories, i.e. p = 9. Critical values for both h and k are given in Table A.1 of Annex A. The calculated column h-values (for each material) that equal or exceed the critical value 1,78 have a bold italic indication. There are four cells with significant h-values: lab 1/material 2, and lab 9/materials 1, 3 and 4.

Tables D.4R and D.4S indicate the variation in the day-1 vs day-2 test results. Actually, only one of these two tables is absolutely needed but both have been generated for this example. Table D.4R contains the "within cell" ranges on the left and the cell ranges squared on the right. For each material, the "cell range" squared total, T_3, is given. Cell ranges for an ITP programme with n = 2 may be converted into standard deviations by std dev = $w/(2)^{1/2}$, where w is the range. Table D.4S has "within-cell" standard deviations on the left and variances (standard deviations squared) on the right. On the right side, the total of all variances, T_4, as well as the pooled or average variance is given for each material.

The analysis of cell standard deviations for outliers is conducted by means of Table D.5. This tabulation of the k-values for all cells for each material is generated using:

$$k = s_i/s_r \quad \text{(D.3)}$$

where

s_i = cell standard deviation for laboratory i;

s_r = pooled cell standard deviation (across all labs) (see Annex A).

The pooled standard deviations (square root of pooled or average variance) are given at the bottom of both Table D.4S and Table D.5. Table D5 has an inset sub-table that gives k(crit) at the 5 % significance level for p = 9 and n = 2. There are three calculated k-values equal to or above the critical value of 1,90: materials 1, 3 and 4 for lab 4. These cells have a bold italic indication.

This completes analysis step 1.

Before proceeding to step 2, it is informative to consult Table D.6, the precision results for the original database. The r-values vary from 0,74 to 3,43 and the R-values from 1,97 to 15,15. If no outliers had been detected in the step 1 analysis, this table would constitute the end of the analysis and the values as they appear in Table D.6 would be used to prepare a final table of precision results for entry into the test method standard. In addition to the five internal calculations of Table D.6 to give the final values for r and R, the table also gives the mean value for each material as well as the repeatability standard deviation s_r and the reproducibility standard deviation s_R. The results of the step 1 outlier analysis for the h and k statistics are given in a sub-table at the bottom of Table D.6. The step 1 outlier analysis has indicated a number of outliers at the 5 % significance level. The presence of these outliers calls for a step 2 analysis operation on a revised ITP database.

D.3.3 Analysis step 2 — Outlier treatment

The step 2 analysis process is twofold:

a) it generates a revised database on which the second round of calculations is conducted to obtain revised values for r and R, using the procedures set forth in Annex B;

b) the revised database is examined to detect any potential outliers at the 2 % significance level.

D.3.3.1 Table nomenclature

The step 2 analysis begins with the calculations for option 2 replacements for the 5 % significance outliers as detected in step 1. In preparation, a second set of spreadsheet tables is generated. To make comparisons and table identification in step 1 and step 2 easier, the table designations for step 2 retain the D.1 to D.6 identification with two added symbols. First, R1 is added, i.e. Table D.1 in step 1 becomes Table D.1-R1 for step 2. The second addition, for option 2 tables, is the symbol OR, where OR designates "outliers replaced". Thus Table D.1 for step 1 becomes Table D.1-R1-OR for the step 2, option 2, operation. Recall that step 1 is conducted on the original database.

D.3.3.2 Step 2 analysis — Replacement of 5 % significance outliers

To implement outlier option 2, AOT replacement values must be obtained for the outliers in the step 1 analysis. Refer to Annex C for the AOT procedure. Basically, two calculations need to be performed. The first to obtain AOT cell *mean* replacements, where *mean* applies both to the cell averages and to cell standard deviations or ranges. These replacements are defined as *parameter replacements* or PRs (see Annex C). Once this has been done, the second procedure is the calculation of cell *data replacement* values or DRs that are necessary to begin the calculation of the new set of precision values for the R1 database.

a) *PRs (cell mean replacements)* — This operation for "cell averages" is conducted using the procedure of Annex C in conjunction with Figures D.1 to D.4. In Figure D.1, the value for lab 9 was declared an outlier in the step 1 analysis. The PR of 51,4 for lab 9 material 1, indicated by a cross, was obtained by the Annex C procedure. The PR of 71,7 for material 2 was obtained for lab 1 using the same procedure. In Figure D.2, the PRs (71,0, 94,5) for lab 9 for both materials were calculated in the same manner. In Figure D.3, the range PR for lab 4 was calculated as 0,85. In Figure D.4, the range PRs 2,20 and 1,20 were obtained for lab 4 for materials 3 and 4, respectively, using the same procedure. The PRs for cell averages are tabulated as item 1 in Part A of Table D.7 and the PRs for cell ranges are tabulated as item 2 in Part A of Table D.7.

   The next operation is to convert these PRs into DRs (cell *data* replacements). The DRs are required for entry into a Table D.1 format to generate a new Table D.1-R1-OR.

b) *DRs (cell average data replacements)* — As outlined in Annex C, there are two types of DR. For this example, all DRs are of the first type: "cell average outlier with non-outlier cell range". Thus the cells scheduled for replacement do not have accompanying cell range (or standard deviation) outliers. The DRs for this first type can be calculated for any selected cell using:

   1) the PRs obtained above;

   2) the existing cell range (ECR) for that cell, using Equations (C.1) and (C.2) in Annex C.

   The *data* entries in item 3 in Part B of Table D.7 were obtained using these two equations with (1) the PRs in Part A and (2) the cell ranges (ECRs) that exist for the four cells in question (these are listed in parentheses next to the replacement averages in Part A). The calculated (duplicate) DRs are shown in item 3 in Part B of Table D.7.

c) *DRs ("cell range" data replacements)* — The PRs listed in item 2 in Part A of Table D.7 must be converted to DRs. All three of these are of the third type, i.e. "cell range outlier with non-outlier cell average". The conversion from PRs to (duplicate) DRs is achieved using:

1) the PR obtained above;

2) the existing cell average (ECA) for that cell and Equations (C.5) and (C.6) in Annex C.

The results of these calculations are shown in item 4 in Part B of Table D.7.

D.3.3.3 Step 2 analysis — Precision for revised database with outlier replacements

Once the outlier replacements have been calculated and tabulated in Table D.7, the revised database can be re-analysed. This begins with Table D.1-R1-OR. The DRs of Table D.7 are substituted for the individual cell outlier values in Table D.1-R1-OR, indicated with italics. When the replacement values for all cells have been entered into Table D.1-R1-OR, the revision 1 (R1) precision results appear in Table D.6-R1-OR.

Table D.6-R1-OR indicates that the repeatability r has been reduced, with an interval of 0,76 to 2,92; and R spans the range 1,76 to 11,27. On an overall (pooled) basis, the repeatability r has been improved by a reduction factor of 0,88 (i.e. 12 % less for r) and the reproducibility R has been improved by a reduction factor of 0,76 (24 % less for R) using the R1 database generated by the outlier replacement procedure.

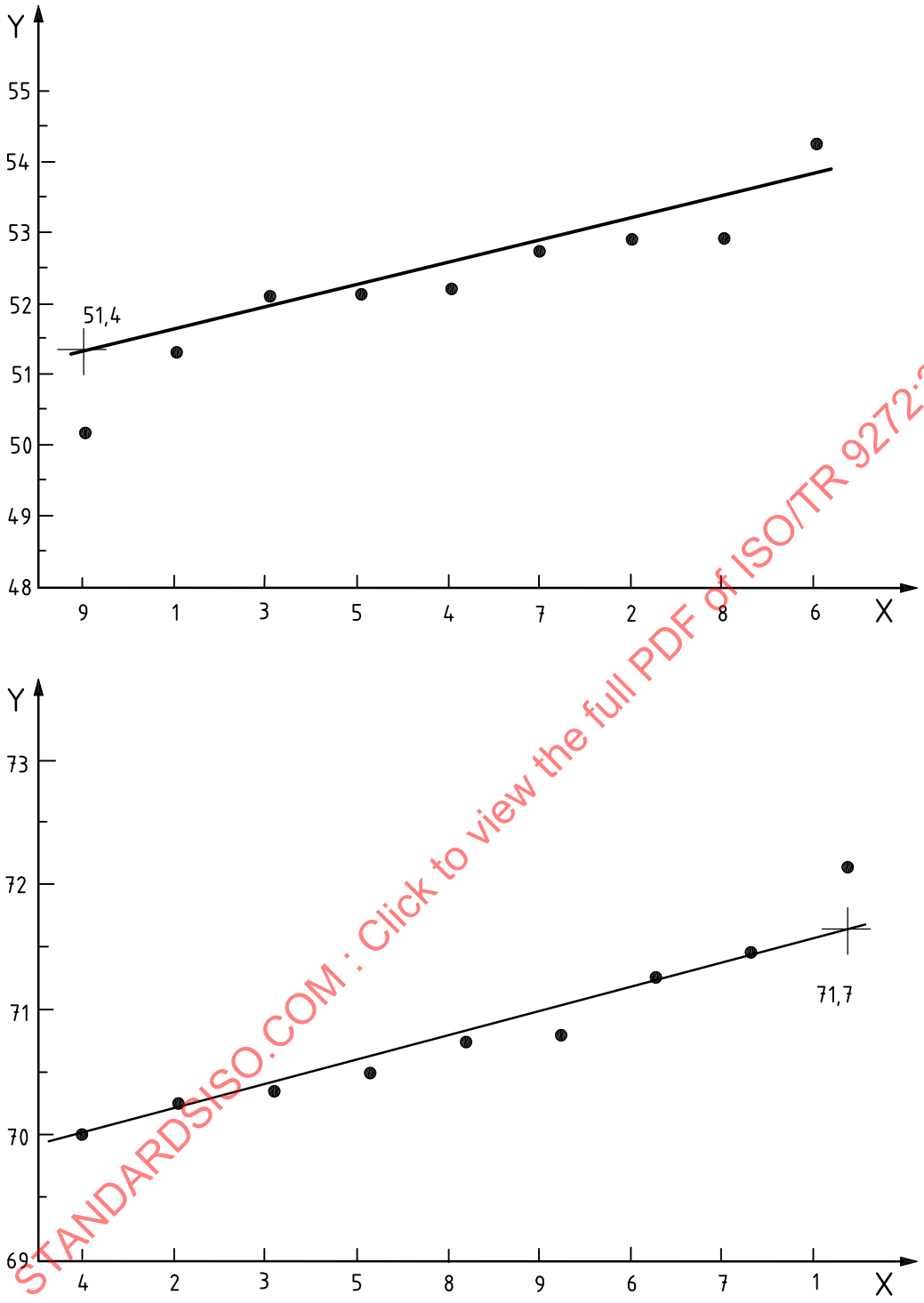
D.3.3.4 Step 2 analysis — Detection and replacement of 2 % significance outliers

When the replacement values for the 5 % outliers are entered into the Table D.1 format (i.e. in Table D.1-R1-OR), the calculation operations for all subsequent tables follow automatically. Critical values for h and k at the 2 % significance level are obtained from Table A.1 in Annex A. Table D.3-R1-OR shows a cell average outlier for material 4 in lab 8. The calculated h-value of 2,07 exceeds the critical h-value of 2,00. Table D.5-R1-OR indicates that the cell range (and standard deviation) for material 1 in lab 1 is an outlier with a calculated k-value of 2,15, exceeding the 2 % critical value 2,09.

The final action for a step 2 analysis is the replacement of the data values found to be outliers at the 2 % significance level. Figure D.5 illustrates AOT plots for material 1 with the range value of 0,80 indicated as the replacement of outlier value 1,10 for lab 1. Also indicated is the plot for material 4 with the PR or cell average replacement value of 99,2 for the outlier 101,5 for lab 8. The two outlier PRs need to be converted into DRs. The cell range *mean* of 0,80 and the cell average *mean* of 99,2 are both converted to DRs using the Annex C equations. These replacement values are shown in Table D.7 in bold italic font.

D.3.3.5 Analysis step 3 — Final operation for Part 1

When the DRs for the two 2 % significance outlier values in the step 2 analysis are inserted into Table D.1-R1-OR in place of the outlier values, a new table, Table D.1-R2-OR, is generated. This new Table D.1-R2-OR is a revision 2 database. Refer to the sequence Table D.1-R2-OR to Table D.6-R2-OR; the last table gives the final revision 2/option 2 repeatability and reproducibility. Comments on the improved precision, i.e. the reduction in r and R, will be postponed until the option 1 analysis is conducted in Part 2.

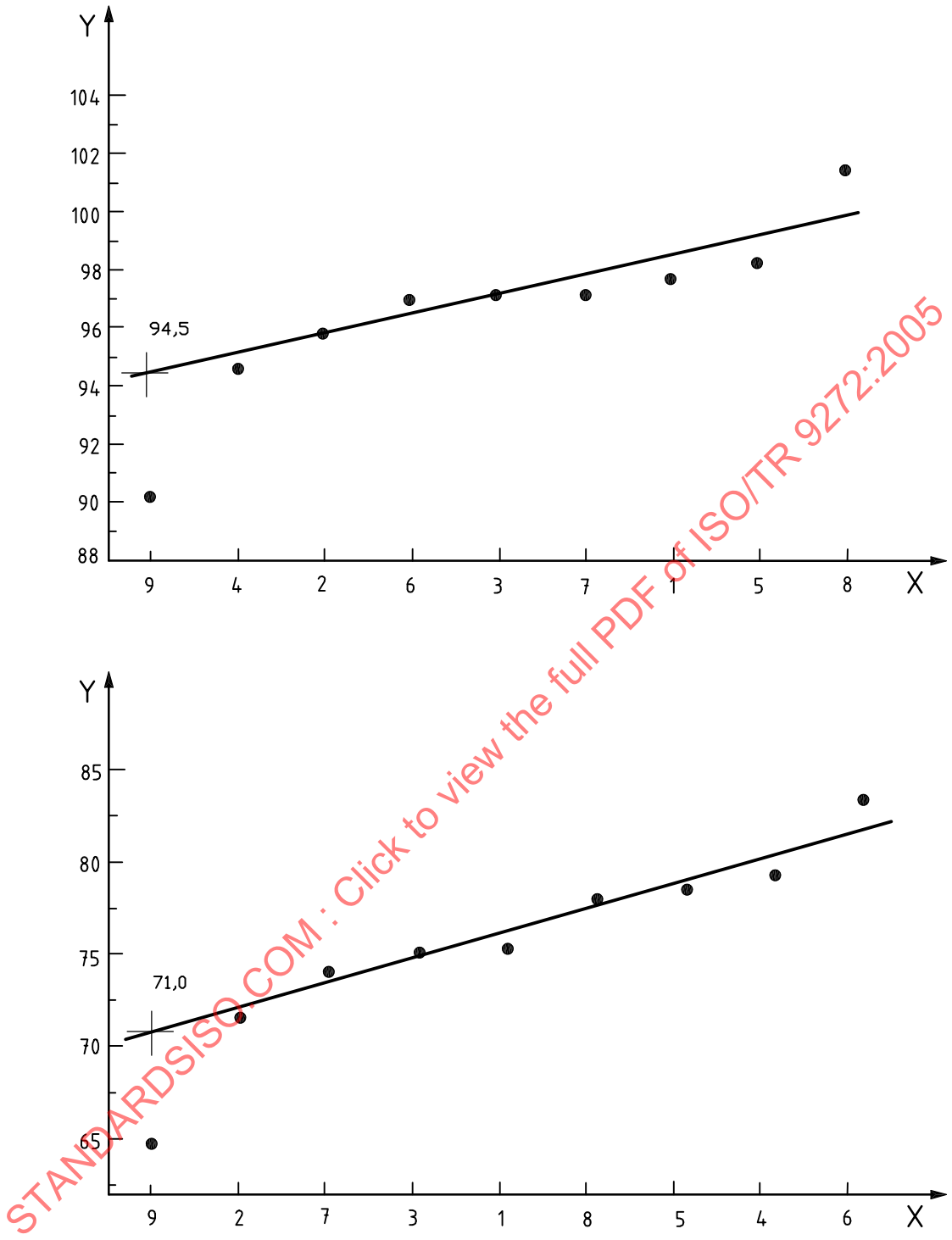


Key

X lab number (in ascending order of result)

Y Mooney viscosity ML

Figure D.1 — AOT plots of original cell averages for materials 1 (upper plot) and 2 (lower plot)
(with linear trend lines and PRs indicated)

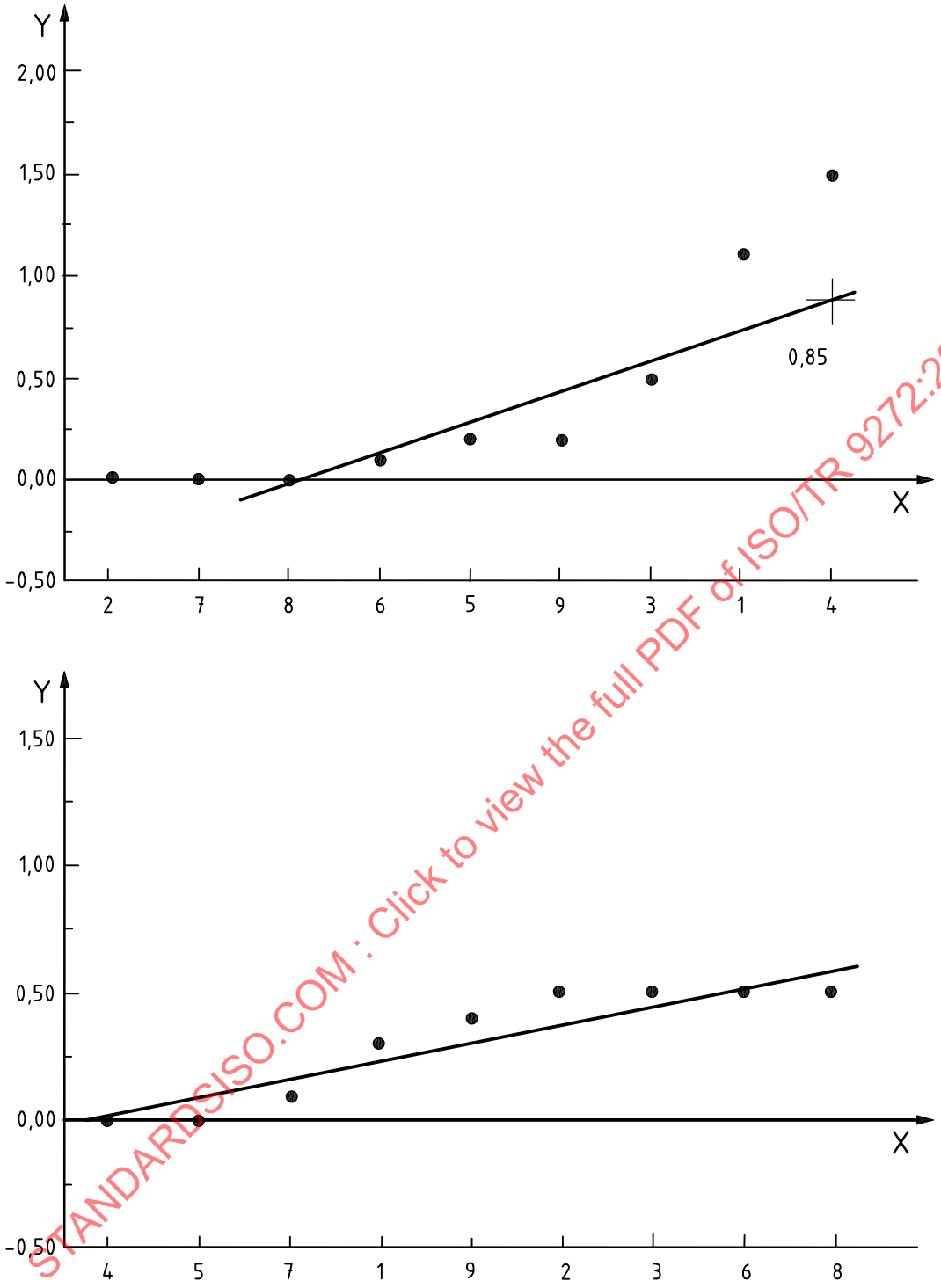


Key

X lab number (in ascending order of result)

Y Mooney viscosity ML

Figure D.2 — AOT plots of original cell averages for materials 3 (upper plot) and 4 (lower plot)
(with linear trend lines and PRs indicated)

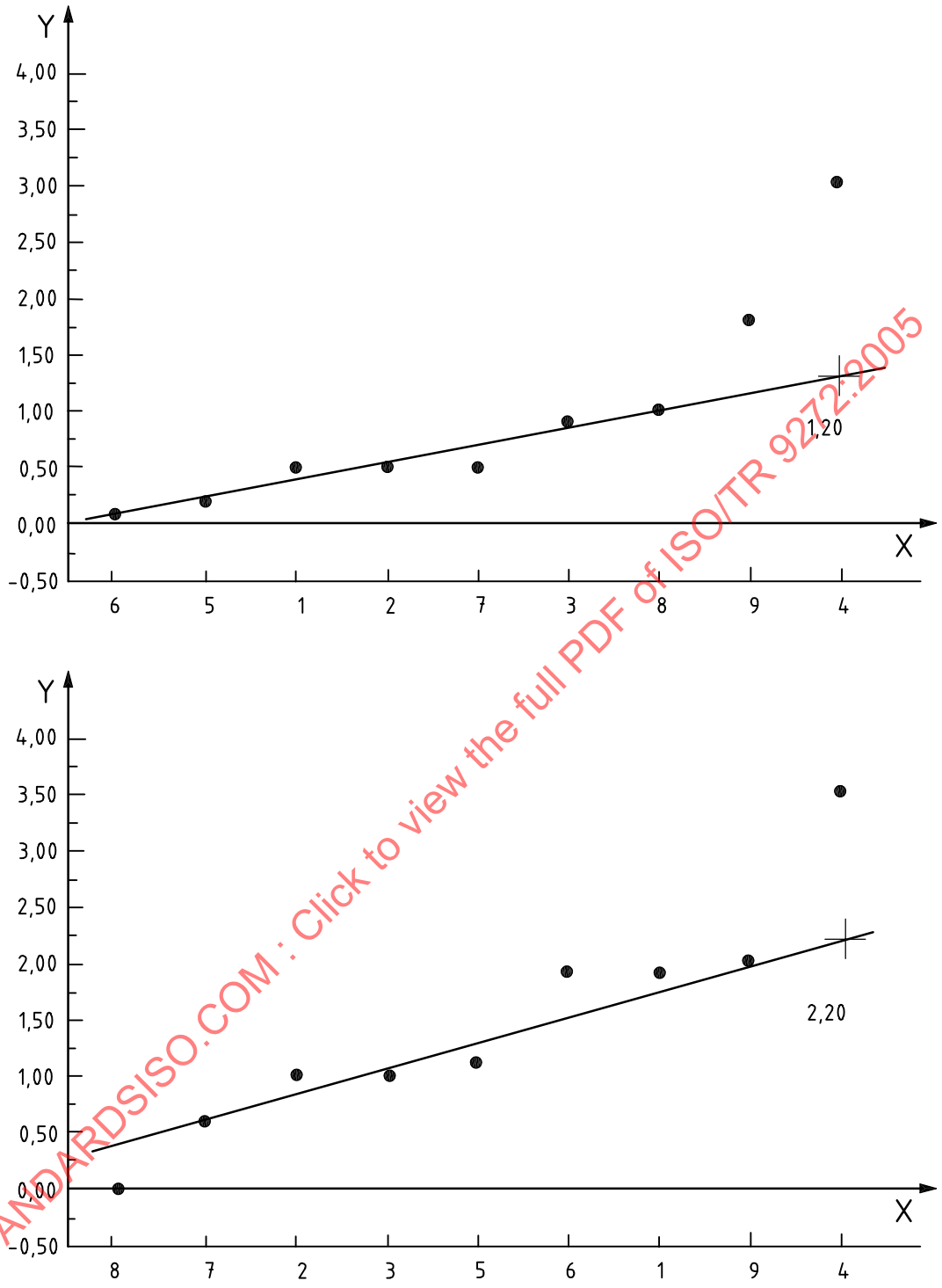


Key

X lab number (in ascending order of result)

Y cell range (absolute value)

Figure D.3 — AOT plots of original cell ranges for materials 1 (upper plot) and 2 (lower plot)
(with linear trend lines and PRs indicated)

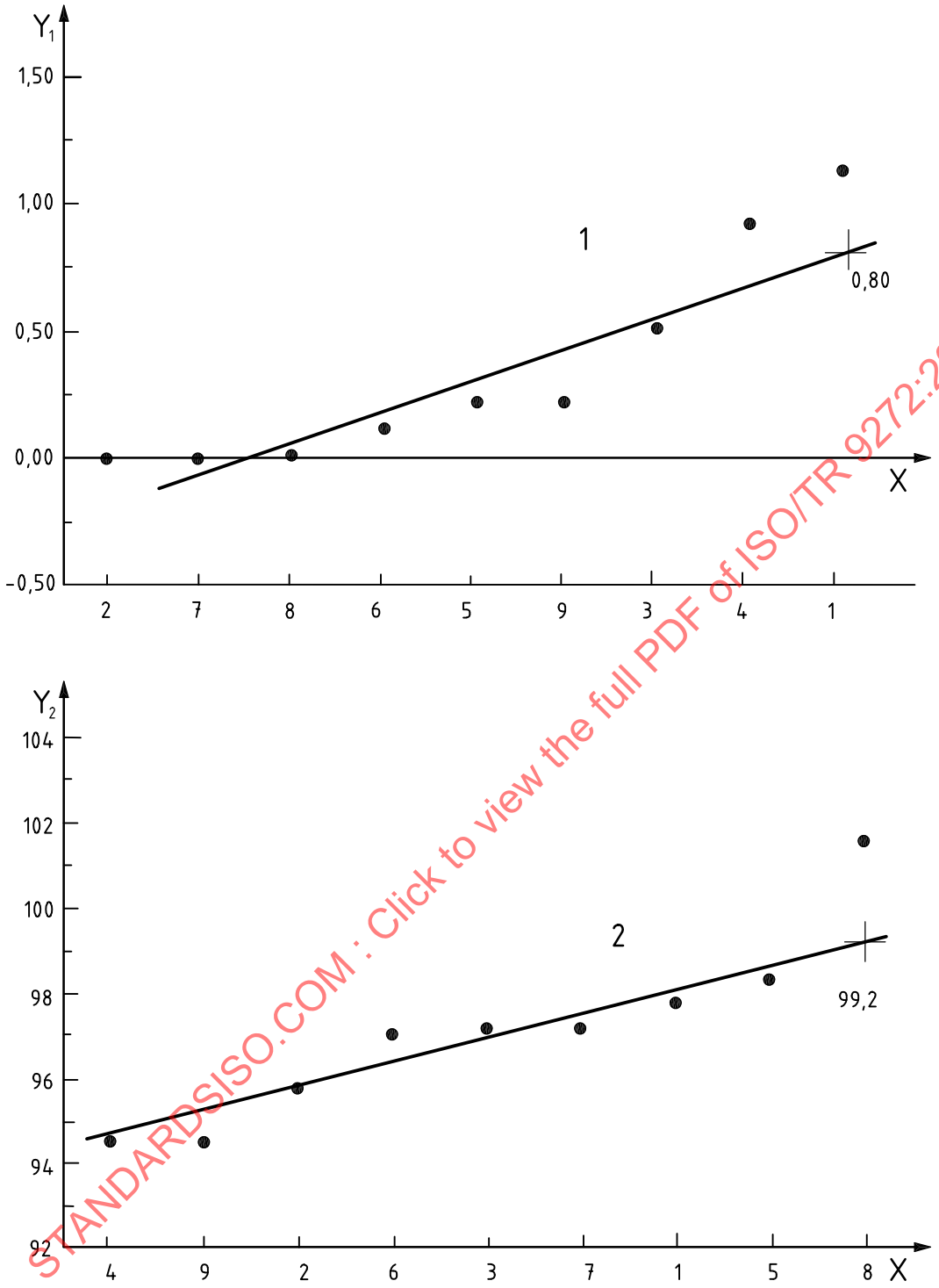


Key

X lab number (in ascending order of result)

Y cell range (absolute value)

Figure D.4 — AOT plots of original cell ranges for materials 3 (upper plot) and 4 (lower plot)
(with linear trend lines and PRs indicated)



Key

X lab number (in ascending order of result)

Y_1 cell range (absolute value)

Y_2 Mooney viscosity ML

1 material 1: 5 % significance k outliers replaced

2 material 4: 5 % significance h outliers replaced

Figure D.5 — AOT plots for revision 1 database for materials 1 (upper plot) and 4 (lower plot)
(with linear trend lines and PRs indicated)

D.4 Part 2: Level 1 precision analysis — Option 1: Outlier deletion

D.4.1 Analysis step 1 — Preliminary review

A substantial portion of the work for Part 2/option 1 has already been done in Part 1. Figures D.1 to D.5, Table D.7 and the two sub-tables at the bottom of Table D.6-R2-OR all indicate the values that have been declared h and k outliers in the Part 1 analysis. If option 1, outlier deletion, had been an initial analysis decision or a decision after step 1, the preliminary review of data and the precision calculations and outlier review of the original database as described above would be the first operation for a Part 2 analysis. These constitute Part 2/step 1 and do not need to be repeated here. For this Part 2/level 1 precision analysis option 1 (outlier deletion) discussion, the final table identification symbol for step 2 analysis is OD, which signifies "outlier deletion".

D.4.2 Analysis step 2

D.4.2.1 Deletion of 5 % significance outliers

Since all outliers have been detected in Part 1, the deletion process is all that is required for this Part 2 analysis. However, in the ordinary analysis of an ITP, if option 1 is chosen as an initial decision, the outlier detection steps for both the 5 % and 2 % significance oultiers would be required prior to the action now described.

Table D.1-R1-OD shows the results of the deletion process on the original database, in Table D.1, to generate the revision 1 database. The tabulated values that have been declared significant at the 5 % level for h and k outliers have been deleted. Tables D.2-R1-OD to D.6-R1-OD are also shown with the blank cells at the locations indicated by the deleted 5 % outliers. In the spreadsheet analysis, all of the blank cells in this series of tables will initially have an ERR indication. As explained in Annex B, each ERR value shall be deleted to produce a blank cell. The final precision results are given in Table D.6-R1-OD. Comparing the results of the outlier replacement option 2 with the outlier deletion option 1, Table D.6-R1-OR vs Table D.6-R1-OD, indicates that option 1 in general gives smaller values for both r and R. A more detailed discussion of the two options will be conducted later in Clause D.9.

D.4.2.2 Deletion of 2 % significance outliers

The next operation is the deletion of cell values that have been declared as outliers at the 2 % significance level. Note at the bottom of Table D.6-R1-OD that two values are indicated; the cell average for material 4 for lab 8 and cell range (or standard deviation) for material 1 for lab 1. The case of material 1/lab 1 requires some consideration by the analyst. Refer to Table D.4R-R1-OD. If the lab 1 range of 1,10 is deleted we are left with six range values much smaller than 1,10, three of which are zero.

Although it is possible to get perfect agreement for two Mooney viscosity measurements one week apart in three of the laboratories, this occurrence must be viewed with some caution. Most technicians know when a special test or ITP is being conducted and they know that good agreement is the goal. A temptation exists to make the results look good. The analyst's judgement in this instance is that the pooled standard deviation (pooled range) would be unrealistically low if the lab 1 value of 1,10 were to be deleted. Therefore, a decision is made to override the objective analysis outcome and not delete the 1,10.

In the Part 1 analysis, the lab 1 range of 1,10 for material 1 was removed, but it was replaced by a value of 0,80. This is different from an outright deletion that removes a laboratory from the list of participants for any material. The deletion of only the material 4/lab 8 value from the revision 1 database yields Table D.1-R2-OD. This table represents the *Revision 2* database.

D.4.3 Part 2: Analysis step 3

The final precision results for Part 2/option 1 are given in Table D.6-R2-OD. Comparing the results of outlier replacement option 2 with outlier deletion option 1, Table D.6-R2-OR vs Table D.6-R2-OD, indicates that option 1 in general gives smaller values for both r and R.

The decision to retain the material 1/lab 1 range of 1,10 brings up a possibility for consideration: the combined use of option 1 and option 2 for outlier treatment. In the case of the Part 2/step 2 analysis, it is possible for the analyst to use the option 2 AOT replacement of 0,80 for this lab's range value, rather than allowing the original value of 1,10 to remain in the *Revision 2* database. This is an alternative option that may be used. It is a judgement call by the analyst.

D.4.4 Discussion of precision results

D.4.4.1 Option 1 vs option 2 vs ISO 5725-5 procedure

Table D.8 summarizes the results of this Mooney viscosity example. The repeatability and reproducibility for each material, as well as a pooled or overall material value, are indicated for:

a) the original database;

b) the use of the ISO 5725-5 robust analysis procedure (the calculations are not given here);

c) AOT outlier replacement (OR) option 2;

d) outlier deletion (OD) option 1.

Each of procedures b), c) and d) constitute one type of “robust” analysis. The goal of a robust analysis is the elimination or drastic reduction of the influence of outliers. Table D.9 indicates the degree of reduction for each of the three procedures in terms of a reduction factor. A reduction factor of 0,60 indicates that the precision parameter obtained for the robust procedure was 60 % of the value for the original or non-revised database or a 40 % change.

D.4.4.2 ISO 5725-5 vs three-step analysis procedure

Comparing ISO 5725-5 to the two options (AOT replacement, OR, and outliers deleted, OD) for the three-step analysis in this Mooney example precision determination indicates the following:

a) For repeatability, the alternative ISO 5725-5 procedure gives some improvement over the other two (option 1, option 2) procedures for material 1: factors of 0,60 vs 0,68, 0,71. There is no difference for material 2 (butyl rubber); all three robust procedures are essentially the same. There are substantial improvements for both options vs the ISO 5725-5 procedure for material 4 and especially for material 3. The pooled values indicate the overall performance in favour of options 1 and 2 compared to the ISO 5725-5 procedure.

b) For reproducibility, both option 1 and option 2 give improvement over the ISO 5725-5 procedure for materials 1, 4 and again especially for material 3. The pooled values for both repeatability and reproducibility indicate that either option 1 or option 2 is better in reducing the influence of outliers than the ISO 5725-5 procedure.

D.4.4.3 Option 1 (deletion) vs option 2 (replacement)

Comparing these two options for the three-step analysis indicates the following:

a) For repeatability, the two options are essentially equal for materials 1 and 2. However, for material 4 and especially material 3, the option 1 outlier deletion procedure gives increased reductions or substantially improved repeatability. The pooled value gives an overall 13 % advantage for option 1 (deletion).

b) For reproducibility, the two options are essentially equal for material 1 and material 4, but option 1 (deletion) gives improvement for material 2 and substantial improvement for material 3. The pooled value gives an overall 6 % improvement for option 1.

D.4.4.4 Comparison of precision for the four materials

The relative precision performance among the four materials for the option 1 (deletion) procedure is indicated in Table D.8. These results have been inserted into the Table 6 precision-summary format as described in Clause 12. The precision in this format for the Annex D example is given in Table D.10 which lists all the precision parameters and also the final number of laboratories in the ITP database after deletion of all outliers.

Materials 1, 2 and 3 give repeatability values, r, that are roughly equal: 0,92, 0,76 and 1,03, respectively. These three r-values differ substantially, as a group, from those obtained for the original database: 1,29, 0,74 and 2,54, respectively, for materials 1, 2 and 3. The outlier removal operation has reduced the r-parameter and gives an indication that all three are very nearly equal. In a sense, this is not too surprising since materials 1, 2 and 3 are all non-pigmented or clear rubbers: SBR, butyl (a NIST reference rubber) and natural rubber, respectively. These three might be expected to respond very similarly to this test within the confines of a single laboratory.

Material 4 is an SBR black masterbatch or SBR-BMB with 65 phr (parts per hundred parts of rubber, by mass) of N339 carbon black. Note that the repeatability for material 4 is substantially poorer (higher r) compared to the other three by a factor of 2,7 on an overall basis. Reasons for this lack of precision are discussed below.

The option 1 (deletion) reproducibility, R, for materials 1 and 3 is essentially equal (2,71 and 2,50) while material 2 has the lowest R at 1,49. Again material 4 is very high (R = 10,84), roughly by a factor of 5 compared to the other three materials on a overall basis. This is about twice the repeatability comparative precision factor of 2,7. For materials 1 to 4, the option 1 reproducibility is substantially improved (lower R) compared to the original database R-values of 3,37, 1,97, 8,84 and 15,15, respectively. Note the considerable differences for the original database R-values between materials 1, 2 and 3 compared to the much more nearly equal values (for materials 1, 2, 3) noted above.

The roughly equal reproducibility, R, for materials 1 and 3 (SBR and NR) is again a reasonably expected outcome: similar test response in a between-laboratory sense for these two unpigmented rubbers. Material 2 (butyl reference rubber) is produced to have high uniformity (good homogeneity bale to bale). It is used as a reference rubber to check the operation of Mooney viscometers. This uniformity undoubtedly accounts for part of its good reproducibility performance. Also this rubber was not subjected to the mill-massing operation.

D.4.4.5 SBR-BMB precision

The very poor performance with material 4, the SBR-BMB, was the subject of further investigation when this ITP was conducted. Subsequent laboratory work showed that the problem could be attributed to the procedure used to mill-mass the rubber prior to conducting the Mooney test. In the mill-massing procedure, the mill temperature, the mill nip (opening) and the time on the mill were not sufficiently well controlled and all were found to play a very important role in the amount of rubber breakdown. Variation in this prior mill-massing operation was the source of the poor precision; variable breakdown leads to variable viscosity.

The breakdown for the SBR-BMB was a combination of (1) rupture of rubber/carbon black intermolecular bonding and (2) ordinary chain rupture. The clear mill-massed rubbers, SBR 1712 and NR, also suffered some chain rupture, but the existence of the additional greater-magnitude breakdown mechanism for the SBR-BMB made it much more susceptible to mill-massing variations and produced the poor precision. ISO 289 was subsequently revised to eliminate the mill-massing operation for BMB rubbers.

Due to the poor precision (high r and R) for the SBR-BMB, this material was not included in the pooled-value calculations in Table D.10. Pooling is recommended only when the precision values are reasonably close for all materials in any ITP.

D.4.4.6 Final observations — Mooney example

The three-step analysis outlier removal operation using the h and k consistency statistics, step 1 at the 5 % significance level and step 2 at the 2 % significance level in the revised database, has given improved repeatability and reproducibility, compared to the original database. Option 1 yields nearly equal r-parameters for all three unpigmented rubbers and nearly equal R-parameters also. A good analysis outcome can be obtained for either option 1 or option 2, but option 1 involves less computation and it yields better precision, i.e. lower overall values for r and R. Option 1 is the preferred choice when there are *nine or more* laboratories in any ITP.

The three-step option 1 analysis has in essence isolated a “core group” of laboratories that have good control of Mooney viscosity testing. Table D.1-R2-OD indicates that laboratories 4 and 8 each had three outliers deleted. These two laboratories have poor control over testing and are in need of improvement. Laboratory 1 is also in need of some remedial efforts: it had two outliers, one of which was not deleted in option 1 as indicated above. Laboratory 8 had one outlier and it may need to give some attention to its test procedures. The “core group” of five laboratories (2, 3, 5, 6 and 7) had good control over their test domain. For materials 1, 2 and 3, the relative repeatability, (r), was 1,8 %, 1,1 % and 1,0 % and the relative reproducibility, (R), was 5,4 %, 2,2 % and 2,5 %, respectively. The precision attained by this “core group” should be the benchmark for Mooney viscosity testing in the rubber manufacturing industry.

Table D.1 — Mooney viscosity — Original basic data from the ITP

Lab No.	Material 1		Material 2		Material 3		Material 4	
	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2
1	50,8	51,9	72,0	72,3	98,0	97,5	74,3	76,2
2	53,0	53,0	70,0	70,5	95,5	96,0	71,0	72,0
3	52,4	51,9	70,1	70,6	96,7	97,6	74,6	75,6
4	53,0	51,5	70,0	70,0	96,0	93,0	81,0	77,5
5	52,3	52,1	70,5	70,5	98,2	98,4	78,0	79,1
6	54,4	54,3	71,5	71,0	97,0	97,1	82,4	84,3
7	52,8	52,8	71,5	71,4	96,9	97,4	73,8	74,4
8	53,0	53,0	71,0	70,5	102,0	101,0	78,0	78,0
9	50,1	50,3	71,0	70,6	91,0	89,2	65,6	63,6
Day avg	52,42	52,31	70,84	70,82	96,81	96,36	75,41	75,63
2-Day avg		52,37		70,83		96,36		75,52
Betw-lab S dev	1,28	1,13	0,74	0,67	2,88	3,41	5,17	5,66
Pooled betw-lab S dev		1,21		0,71		3,16		5,42

Table D.2 — Cell averages and cell averages squared — Original data

Cell averages					Cell averages squared				
Lab No.	**Material 1**	**Material 2**	**Material 3**	**Material 4**	**Lab No.**	**Material 1**	**Material 2**	**Material 3**	**Material 4**
1	51,35	72,15	97,75	75,25	1	2 636,82	5 205,62	9 555,06	5 662,56
2	53,00	70,25	95,75	71,50	2	2 809,00	4 935,06	9 168,06	5 112,25
3	52,15	70,35	97,15	75,10	3	2 719,62	4 949,12	9 438,12	5 640,01
4	52,25	70,00	94,50	79,25	4	2 730,06	4 900,00	8 930,25	6 280,56
5	52,20	70,50	98,30	78,55	5	2 724,84	4 970,25	9 662,89	6 170,10
6	54,35	71,25	97,05	83,35	6	2 953,92	5 076,56	9 418,70	6 947,22
7	52,80	71,45	97,15	74,10	7	2 787,84	5 105,10	9 438,12	5 490,81
8	53,00	70,75	101,50	78,00	8	2 809,00	5 005,56	10 302,25	6 084,00
9	50,20	70,80	90,10	64,60	9	2 520,04	5 012,64	8 118,01	4 173,16
T_1 =	471,300	637,500	869,250	679,700	T_2 =	24 691,150	45 159,925	84 031,473	51 560,680
Cell avg	52,37	70,83	96,58	75,52					
Var cell avg	1,342 5	0,459 4	9,551 3	28,528 2					
S dev cell avg	1,159	0,678	3,091	5,341					
NOTE Variance cell avg = s^2(Y_{av})									

Table D.3 — Cell avg "dev", d- and h-values — Original data

Cell deviations, d					Cell h-values				
Lab No.	**Material 1**	**Material 2**	**Material 3**	**Material 4**	**Lab No.**	**Material 1**	**Material 2**	**Material 3**	**Material 4**
1	–1,02	1,32	1,17	–0,27	1	–0,88	***1,94***	0,38	–0,05
2	0,63	–0,58	–0,83	–4,02	2	0,55	–0,86	–0,27	–0,75
3	–0,22	–0,48	0,57	–0,42	3	–0,19	–0,71	0,18	–0,08
4	–0,12	–0,83	–2,08	3,73	4	–0,10	–1,23	–0,67	0,70
5	–0,17	–0,33	1,72	3,03	5	–0,14	–0,49	0,56	0,57
6	1,98	0,42	0,47	7,83	6	1,71	0,61	0,15	1,47
7	0,43	0,62	0,57	–1,42	7	0,37	0,91	0,18	–0,27
8	0,63	–0,08	4,92	2,48	8	0,55	–0,12	1,59	0,46
9	–2,17	–0,03	–6,48	–10,92	9	***–1,87***	–0,05	***–2,10***	***–2,04***
All-lab cell avg	52,37	70,83	96,58	75,52	***h*(crit) 5 % significance level at indicated p**				
S dev cell avgs	1,159	0,678	3,091	5,341	p =	9	9	9	9
Bold and italic = significant values					***h*(crit)**	1,78	1,78	1,78	1,78
					Lab No. > *h*(crit)	9	1	9	9
h = $dls(Y_{av})$, where d = avg cell i – (avg all cells); $s(Y_{av})$ = S dev of cell avgs									

Table D.4R — Cell ranges and ranges squared — Original data

Cell ranges					Cell ranges squared				
Lab No.	Material 1	Material 2	Material 3	Material 4	Lab No.	Material 1	Material 2	Material 3	Material 4
1	1,100	0,300	0,500	1,900	1	1,210	0,090	0,250	3,610
2	0,000	0,500	0,500	1,000	2	0,000	0,250	0,250	1,000
3	0,500	0,500	0,900	1,000	3	0,250	0,250	0,810	1,000
4	1,500	0,000	3,000	3,500	4	2,250	0,000	9,000	12,250
5	0,200	0,000	0,200	1,100	5	0,040	0,000	0,040	1,210
6	0,100	0,500	0,100	1,900	6	0,010	0,250	0,010	3,610
7	0,000	0,100	0,500	0,600	7	0,000	0,010	0,250	0,360
8	0,000	0,500	1,000	0,000	8	0,000	0,250	1,000	0,000
9	0,200	0,400	1,800	2,000	9	0,040	0,160	3,240	4,000
Avg range	0,400	0,311	0,944	1,444	T_3 =	3,800 0	1,260 0	14,850 0	27,040 0

T_3 = Sum “cell ranges squared”

Calculation algorithm for any ITP cell range, with duplicates in cells cxx and dxx:

@IF[(cxx-dxx)<0, (cxx-dxx)*–1, (cxx-dxx)]

Table D.4S — Cell standard deviations and variances

Cell std deviations					Cell variances				
Lab No.	Material 1	Material 2	Material 3	Material 4	Lab No.	Material 1	Material 2	Material 3	Material 4
1	0,778	0,212	0,354	1,344	1	0,605 0	0,045 0	0,125 0	1,805 0
2	0,000	0,354	0,354	0,707	2	0,000 0	0,125 0	0,125 0	0,500 0
3	0,354	0,354	0,636	0,707	3	0,125 0	0,125 0	0,405 0	0,500 0
4	1,061	0,000	2,121	2,475	4	1,125 0	0,000 0	4,500 0	6,125 0
5	0,141	0,000	0,141	0,778	5	0,020 0	0,000 0	0,020 0	0,605 0
6	0,071	0,354	0,071	1,344	6	0,005 0	0,125 0	0,005 0	1,805 0
7	0,000	0,071	0,354	0,424	7	0,000 0	0,005 0	0,125 0	0,180 0
8	0,000	0,354	0,707	0,000	8	0,000 0	0,125 0	0,500 0	0,000 0
9	0,141	0,283	1,273	1,414	9	0,020 0	0,080 0	1,620 0	2,000 0
Pooled S dev	0,459	0,265	0,908	1,226	T_4 =	1,900 00	0,630 00	7,425 00	13,520 00
					Pooled variance	0,211 1	0,070 0	0,825 0	1,502 2

Table D.5 — Cell k-values — Original data

Lab No.	Material 1	Material 2	Material 3	Material 4
1	1,69	0,80	0,39	1,10
2	0,00	1,34	0,39	0,58
3	0,77	1,34	0,70	0,58
4	***2,31***	0,00	***2,34***	***2,02***
5	0,31	0,00	0,16	0,63
6	0,15	1,34	0,08	1,10
7	0,00	0,27	0,39	0,35
8	0,00	1,34	0,78	0,00
9	0,31	1,07	1,40	1,15
Pooled S dev	0,459	0,265	0,908	1,226
***k*(crit) 5 % signif level at *n* = 2, indicated *p*:**				
***p* =**	9	9	9	9
***k*(crit) =**	1,90	1,90	1,90	1,90
Lab No. > *k*(crit)	4	none	4	4
Bold and italic = Significant values $k = s(i)/s_r$, where $s(i)$ = indiv cell std dev; s_r = pooled all-lab std dev				

Table D.6 — Mooney viscosity: Calculation for precision — Original data

ITP for n =		2	2	2	2
p =		9	9	9	9
		Material 1	Material 2	Material 3	Material 4
T_1 =		471,300	637,500	869,250	679,700
T_2 =		24 691,150	45 159,925	84 031,473	51 560,680
T_4 =		1,900 00	0,630 00	7,425 00	13,520 00
Calcn 1	(s_r)^2 = T_4/p =	0,211 1	0,070 0	0,825 0	1,502 2
(s_L)^2 = {[pT_2 – (T_1)^2]/$p(p-1)$} – [(s_r)^2/2]					
Calcn 2	(s_L)^2 =	1,236 9	0,424 4	9,138 8	27,777 1
(s_R)^2 = (s_L)^2 + (s_r)^2					
Calcn 3	(s_R)^2 =	1,448 1	0,494 4	9,963 8	29,279 3
r = 2,8 [(s_r)^2]^0,5 = Repeatability					
Calcn 4	r =	1,287	0,741	2,543	3,432
R = 2,8 [(s_R)^2]^0,5 = Reproducibility					
Calcn 5	R =	3,37	1,97	8,84	15,15
		Material 1	Material 2	Material 3	Material 4
Material averages		52,37	70,83	96,58	75,52
Standard deviation, s_r =		0,459	0,265	0,908	1,226
Standard deviation, s_R =		1,203	0,703	3,157	5,411
	Relative (r)	2,46	1,05	2,63	4,54
	Relative (R)	6,43	2,78	9,15	20,06

Step 1: Outliers at 5 % significance level for materials 1 to 4					
		Material 1	Material 2	Material 3	Material 4
For h:	Lab No.	9	1	9	9
For k:	Lab No.	4	none	4	4

Table D.7 — Replacement values for outliers

Part A — AOT parameter replacement values (PRs)				
1. AOT PRs for cell average outliers				
Lab No.	**Material 1**	**Material 2**	**Material 3**	**Material 4**
1		71,7 (0,30)		
8			***99,2 (1,00)***	
9	51,4 (0,20)		94,5 (1,80)	71,0 (2,00)
NOTE Cell mean replacement (cell averages) listed with individual cell range in parentheses.				
2. AOT PRs for cell range outliers				
Lab No.	**Material 1**	**Material 2**	**Material 3**	**Material 4**
1	***0,80 (51,35)***			
4	0,85 (52,25)		1,20 (94,50)	2,20 (79,25)
NOTE Cell PRs (cell ranges) listed with indiv cell avg in ( ).				
Part B — AOT (cell) data replacement values (DRs)				
3. AOT DRs for cell average outliers				
Lab No.	**Material 1**	**Material 2**	**Material 3**	**Material 4**
1		71,6, 72,0		
8			***98,7, 99,7***	
9	51,3, 51,5		93,6, 95,4	70,0, 72,0
4. AOT DRs for cell range outliers				
Lab No.	**Material 1**	**Material 2**	**Material 3**	**Material 4**
1	51,8, 51,0			
4	51,8, 52,7		93,9, 95,1	74,2, 76,4
NOTE Bold and italic = values significant at 2 % level.				

Table D.8 — Comparison of outlier handling procedures

Part 1 **Outlier procedure**	**Repeatability,** r **Material 1**	**Material 2**	**Material 3**	**Material 4**	**Pooled precision,** r
Original database (no outliers deleted)	1,29	0,74	2,54	3,43	2,26
Alternative ISO 5725-5 robust analysis	0,78	0,74[a]	2,18	3,22	2,02
AOT outlier replacement, option 2[b]	0,88	0,76	1,55	2,92	1,75
Outliers deleted, option 1[b]	0,92	0,76	1,03	2,46	1,46
Part 2 **Outlier procedure**	**Reproducibility,** R **Material 1**	**Material 2**	**Material 3**	**Material 4**	**Pooled precision,** R
Original database (no outliers deleted)	3,37	1,97	8,84	15,15	8,98
Alternative ISO 5725-5 robust analysis	3,09	1,97[a]	6,76	14,62	8,26
AOT outlier replacement, option 2[b]	2,64	1,76	4,66	11,27	6,30
Outliers deleted, option 1[b]	2,71	1,49	2,50	10,84	5,77

[a] Analysis not conducted for material 2.

[b] Final precision results.

Pooled (or mean) precision across four materials calculated on basis of variance or std dev squared.

NOTE See Table D.7 for materials (and labs) with outliers.

Table D.9 — Relative reduction factors — Precision parameters, r **and** R

Part 1 **Outlier procedure**	**Reduction factor for repeatability,** r **Material 1**	**Material 2**	**Material 3**	**Material 4**	**Pooled precision redn factor**
Original database (no outliers deleted)	1,0	1,0	1,0	1,0	1,0
Alternative ISO 5725-5 robust analysis	0,60	[a]	0,86	0,94	0,89
AOT outlier replacement, option 2[b]	0,68	1,03	0,61	0,85	0,78
Outliers deleted, option 1[b]	0,71	1,03	0,41	0,72	0,65
Part 2 **Outlier procedure**	**Reduction factor for reproducibility,** R **Material 1**	**Material 2**	**Material 3**	**Material 4**	**Pooled precision redn factor**
Original database (no outliers deleted)	1,0	1,0	1,0	1,0	1,0
Alternative ISO 5725-5 robust analysis	0,92	[a]	0,76	0,97	0,92
AOT outlier replacement, option 2[b]	0,78	0,89	0,53	0,74	0,70
Outliers deleted, option 1[b]	0,80	0,76	0,28	0,72	0,64

[a] Analysis not conducted for material 2.

[b] Final precision results.

Reduction factor = (revised precision database/ orig precision database)

Pooled precision reduction factor calculated on pooled precision in Table D.8.

NOTE See Table D.7 for materials (and labs) with outliers.

Table D.10 — Level 1 and type 1 — Precision for Mooney viscosity (Measured property = ML viscosity @ 100 °C, in Mooney units)

Material	Mean	Within lab			Between labs			No. of labs [a]
		s_r	r	(r)	s_R	R	(R)	
1 SBR 1712	50,7	0,328	0,920	1,81	0,967	2,71	5,35	7
2 IIR (butyl)	68,7	0,270	0,757	1,10	0,532	1,49	2,17	8
3 NR	99,2	0,366	1,03	1,04	0,892	2,50	2,52	6
4 SBR-BMB	74,6	0,878	2,46	3,30	3,87	10,84	14,5	7
Pooled values [b]		0,321	0,90	1,31	0,80	2,23	3,34	

Notation used:

s_r = within-laboratory standard deviation (in measurement units)

r = repeatability (in measurement units)

(r) = repeatability (in percent of mean level)

s_R = between-laboratory standard deviation (for total between-laboratory variation in measurement units)

R = reproducibility (in measurement units)

(R) = reproducibility (in percent of mean level)

[a] Number of labs in the revised database after option 1 outlier deletion.

[b] Simple averages are listed for pooled values, omitting 4 (SBR-BMB).

See text of precision clause for discussion of precision results given in this table.

Table D.1-R1-OR — Mooney viscosity — AQT replacement values (in italics) for 5 % outliers

Lab No.	Material 1		Material 2		Material 3		Material 4	
	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2
1	50,8	51,9	***71,6***	***72,0***	98,0	97,5	70,3	72,2
2	53,0	53,0	70,0	70,5	95,5	96,0	67,0	68,0
3	52,4	51,9	70,1	70,6	96,7	97,6	70,6	71,6
4	***51,8***	***52,7***	70,0	70,0	***93,9***	***95,1***	***74,2***	***76,4***
5	52,3	52,1	70,5	70,5	98,2	98,4	74,0	75,1
6	54,4	54,3	71,5	71,0	97,0	97,1	78,4	80,3
7	52,8	52,8	71,5	71,4	96,9	97,4	69,8	70,4
8	53,0	53,0	71,0	70,5	102,0	101,0	74,0	74,0
9	***51,3***	***51,5***	71,0	70,6	***93,9***	***95,1***	***66,0***	***68,0***
Day avg	52,42	52,58	70,80	70,79	96,90	97,24	71,59	72,89
2-Day avg		52,50		70,79		97,07		72,24
Betw-lab S dev	1,06	0,84	0,67	0,59	2,47	1,82	3,92	4,02
Pooled betw-lab S dev		0,96		0,63		2,17		3,97

Significant replaced values at 5 % = Bold, italic.

Table D.2-R1-OR — Cell averages and cell averages squared: AOT replacements for 5 % outliers

Cell averages					Cell averages squared				
Lab No.	**Material 1**	**Material 2**	**Material 3**	**Material 4**	**Lab No.**	**Material 1**	**Material 2**	**Material 3**	**Material 4**
1	51,35	71,80	97,75	71,25	1	2 636,82	5 155,24	9 555,06	5 076,56
2	53,00	70,25	95,75	67,50	2	2 809,00	4 935,06	9 168,06	4 556,25
3	52,15	70,35	97,15	71,10	3	2 719,62	4 949,12	9 438,12	5 055,21
4	52,25	70,00	94,50	75,30	4	2 730,06	4 900,00	8 930,25	5 670,09
5	52,20	70,50	98,30	74,55	5	2 724,84	4 970,25	9 662,89	5 557,70
6	54,35	71,25	97,05	79,35	6	2 953,92	5 076,56	9 418,70	6 296,42
7	52,80	71,45	97,15	70,10	7	2 787,84	5 105,10	9 438,12	4 914,01
8	53,00	70,75	101,50	74,00	8	2 809,00	5 005,56	10 302,25	5 476,00
9	51,40	70,80	94,50	67,00	9	2 641,96	5 012,64	8 930,25	4 489,00
T_1 =	472,500	637,150	873,650	650,150	T_2 =	24 813,070	45 109,543	84 843,713	47 091,248
Cell avg	52,50	70,79	97,07	72,24					
Var cell avg	0,852 5	0,357 8	4,570 7	15,641 7					
S dev cell avg	0,923	0,598	2,138	3,955					

NOTE Variance cell avg = s^2(Y_{av}).

Table D.3-R1-OR — Cell avg dev d- and h-values: AOT replacement for 5 % outliers

Cell deviations, d					Cell h-values				
Lab No.	**Material 1**	**Material 2**	**Material 3**	**Material 4**	**Lab No.**	**Material 1**	**Material 2**	**Material 3**	**Material 4**
1	–1,15	1,01	0,68	–0,99	1	–1,25	1,68	0,32	–0,25
2	0,50	–0,54	–1,32	–4,74	2	0,54	–0,91	–0,62	–1,20
3	–0,35	–0,44	0,08	–1,14	3	–0,38	–0,74	0,04	–0,29
4	–0,25	–0,79	–2,57	3,06	4	–0,27	–1,33	–1,20	0,77
5	–0,30	–0,29	1,23	2,31	5	–0,32	–0,49	0,57	0,58
6	1,85	0,46	–0,02	7,11	6	2,00	0,76	–0,01	1,80
7	0,30	0,66	0,08	–2,14	7	0,32	1,10	0,04	–0,54
8	0,50	–0,04	4,43	1,76	8	0,54	–0,07	***2,07***	0,45
9	–1,19	0,01	–2,57	–5,24	9	–1,19	0,01	–1,20	–1,32
					***h*(crit) 2 % signif level at indicated p:**				
All-lab cell avg	52,50	70,79	97,07	72,24	***p* =**	9	9	9	9
S dev cell avg	0,923	0,598	2,138	3,955	***h*(crit)**	2,00	2,00	2,00	2,00
					Lab No. > *h*(crit)	none	none	8	none

$h = d/s(Y_{av})$, where d = avg cell i – (avg all cells) and $s(Y_{av})$ = std dev of cell avgs.

Significant value = Bold and italic.

Table D.4R-R1-OR — Cell ranges and cell ranges squared: AOT replacement for 5 % outliers

Cell ranges					Cell ranges squared				
Lab No.	Material 1	Material 2	Material 3	Material 4	Lab No.	Material 1	Material 2	Material 3	Material 4
1	1,100	0,400	0,500	1,900	1	1,210	0,160	0,250	3,610
2	0,000	0,500	0,500	1,000	2	0,000	0,250	0,250	1,000
3	0,500	0,500	0,900	1,000	3	0,250	0,250	0,810	1,000
4	0,900	0,000	1,200	2,200	4	0,810	0,000	1,440	4,840
5	0,200	0,000	0,200	1,100	5	0,040	0,000	0,040	1,210
6	0,100	0,500	0,100	1,900	6	0,010	0,250	0,010	3,610
7	0,000	0,100	0,500	0,600	7	0,000	0,010	0,250	0,360
8	0,000	0,500	1,000	0,000	8	0,000	0,250	1,000	0,000
9	0,200	0,400	1,200	2,000	9	0,040	0,160	1,440	4,000
Range	0,333	0,322	0,678	1,300	T_3 =	2,360 0	1,330 0	5,490 0	19,630 0

T_3 = Sum “cell ranges squared”

Table D.4S-R1-OR — Cell standard deviations and variances: AOT replacement for 5 % outliers

Cell std deviations					Cell variances				
Lab No.	Material 1	Material 2	Material 3	Material 4	Lab No.	Material 1	Material 2	Material 3	Material 4
1	0,778	0,283	0,354	1,344	1	0,605 0	0,080 0	0,125 0	1,805 0
2	0,000	0,354	0,354	0,707	2	0,000 0	0,125 0	0,125 0	0,500 0
3	0,354	0,354	0,636	0,707	3	0,125 0	0,125 0	0,405 0	0,500 0
4	0,636	0,000	0,849	1,556	4	0,405 0	0,000 0	0,720 0	2,420 0
5	0,141	0,000	0,141	0,778	5	0,020 0	0,000 0	0,020 0	0,605 0
6	0,071	0,354	0,071	1,344	6	0,005 0	0,125 0	0,005 0	1,805 0
7	0,000	0,071	0,354	0,424	7	0,000 0	0,005 0	0,125 0	0,180 0
8	0,000	0,354	0,707	0,000	8	0,000 0	0,125 0	0,500 0	0,000 0
9	0,141	0,283	0,849	1,414	9	0,020 0	0,080 0	0,720 0	2,000 0
	0,362	0,272	0,552	1,044	T_4 =	1,180 00	0,665 00	2,745 00	9,815 00
Pooled variance						0,131 1	0,073 9	0,305 0	1,090 6

Table D.5-R1-OR — k-values: AOT replacement for 5 % outliers

Lab No.	Material 1	Material 2	Material 3	Material 4
1	***2,15***	1,04	0,64	1,29
2	0,00	1,30	0,64	0,68
3	0,98	1,30	1,15	0,68
4	1,76	0,00	1,54	1,49
5	0,39	0,00	0,26	0,74
6	0,20	1,30	0,13	1,29
7	0,00	0,26	0,64	0,41
8	0,00	1,30	1,28	0,00
9	0,39	1,04	1,54	1,35
Pooled S dev	0,362	0,272	0,552	1,044
***k*(crit) 2 % significance level at n = 2, indicated p:**				
***p* =**	9	9	9	9
***k*(crit) =**	2,09	2,09	2,09	2,09
Lab No. > *k*(crit)	1	none	none	none
Significant value = Bold and italic.				
$k = s(i)/s_r$, where $s(i)$ = indiv cell std dev and s_r = pooled all-lab std dev.				