



Designation: D 4483 – 99

Standard Practice for Determining Precision for Test Method Standards in the Rubber and Carbon Black Industries¹

This standard is issued under the fixed designation D 4483; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 This practice presents guidelines for preparing clear, meaningful precision statements for test method standards under the jurisdiction of ASTM Committee D-11 on Rubber Testing and for ASTM Committee D-24 on Carbon Black Testing. It explains the potential uses for standard test methods and gives the requirements for interlaboratory programs needed in precision formulation, the calculation algorithms for precision, and the format for expressing precision.

1.2 Test methods are used in many ways in technology. This broad usage requires careful consideration in assessing their general precision and, where pertinent, their accuracy. Clearly outlining the objectives and the uses of test methods prior to the determination of test precision is essential. A critical requirement for this is the development of a standardized nomenclature system. This practice addresses these and other issues important in evaluating precision for test method standards.

1.3 This practice is divided into the following sections:

	Section
Terminology	3
Significance and Use	4
General Principles	5
Organizing an Interlaboratory Precision Program	6
Analysis Concepts for Interlaboratory Test Data	7
Calculating the Precision Parameters	8
Format for Precision and Bias Section (Clause) of Standard	9
Statistical Model for Precision Testing	Annex A1
Practice E 691 Calculations for Cell Average Outliers: <i>h</i> -values	Annex A2
Practice E 691 Calculations for Cell Standard Deviation Outliers: <i>k</i> -values	Annex A3
Establishing a Functional Relationship Between <i>r</i> , <i>R</i> , and <i>M</i>	Annex A4
Procedure for Carbon Black Precision Evaluation	Annex A5
Spreadsheet Calculation Formulas for Precision Parameters	Annex A6
An Example of Precision Calculation—Mooney Viscosity	Annex A7

1.4 *This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety and health practices and determine the applicability of regulatory limitations prior to use.*

2. Referenced Documents

2.1 ASTM Standards:

D 1646 Test Methods for Rubber—Viscosity, Stress Relaxation, and Pre-Vulcanization Characteristics (Mooney Viscometer)²

E 691 Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method³

2.2 ISO Standard:

ISO 5725 Precision of Test Methods—Determination of Repeatability and Reproducibility by Interlaboratory Tests⁴

3. Terminology

3.1 *Definitions of Terms Specific to This Standard:*—This section gives descriptions for the important terms used in this practice. However, Section 5 should be reviewed simultaneously with this section for a more complete understanding of the need for certain terms.

NOTE—The descriptions of terms are given in a logical development sequence rather than alphabetical order.

3.1.1 *accuracy, bias, precision*—to set the stage for the more specific terminology to follow, three general terms are given. Although this practice does not address the issue of accuracy or bias, these terms are presented to clearly show the difference between these two and precision.

3.1.2 *accuracy*—the degree of correspondence between an average measured value and an accepted reference or standard value for the material or phenomenon under test.

3.1.2.1 *Discussion*—The reference value may be established by theory, by reference to an *accepted* standard, to another test method, or in some cases the average that could be obtained by applying the test method to all of the sampling units comprising a lot of the material.

3.1.3 *bias*—the difference between the average measured test result and the accepted reference value.

3.1.3.1 *Discussion*—High accuracy implies a small or negligible bias and when bias exists increased testing does not increase accuracy but merely enhances the knowledge of the degree of bias.

¹ This practice is under the jurisdiction of ASTM Committee D-11 on Rubber and is the direct responsibility of Subcommittee D11.16 on Application of Statistical Methods.

Current edition approved Nov. 10, 1999. Published December 1999. Originally published as D 4483 – 85. Last previous edition D 4483 – 98.

² *Annual Book of ASTM Standards*, Vol 09.01.

³ *Annual Book of ASTM Standards*, Vol 14.02.

⁴ Available from American National Standards Institute, 11 W. 42nd St., 13th Floor, New York, NY 10036.

3.1.4 *precision*—a measurement concept that expresses the ability to generate test results that agree with each other in absolute magnitude.

3.1.4.1 *Discussion*—The degree of agreement is normally measured inversely by the standard deviation; high precision corresponds to a low (small) standard deviation.

3.1.4.2 *Discussion*—High precision may exist simultaneously with a large bias or poor accuracy.

3.2 The following specific descriptions are given for terms that will be required to accommodate Committee D-11 and Committee D-24 test methods. The three time scales of repeatability and reproducibility discussed in 5.2 are reduced to two for the sake of simplification. Two preliminary terms, which define the “numbers” produced by test methods, are required. These are given first.

3.2.1 *determination*—the application of the complete test procedure to one test piece, specimen, or test portion to produce one numerical (test) measured value to be used to form an average or median.

3.2.2 *test result*—the average (mean or median) of a specified number of determinations; it is the reported value for a test.

3.2.3 *repeatability, r*—an established value, below which the absolute difference between two “within-laboratory” test results may be expected to lie, with a specified probability.

3.2.3.1 *Discussion*—The two test results are obtained with the *same* method on nominally identical test materials under the *same* conditions (same operator, apparatus, laboratory, and specified time period), and in the absence of other indications the probability is 95 %.

3.2.3.2 *Discussion*—The “established value” may also be called a “critical difference.”

3.2.4 *reproducibility, R*—an established value, below which the absolute difference between two “between-laboratory” test results may be expected to lie, with a specified probability.

3.2.4.1 *Discussion*—The two test results are obtained with the *same* method on nominally identical test materials under *different* conditions (different laboratories, operators, apparatus, and in a specified time period), and in the absence of other indications the probability is 95 %.

3.2.4.2 *Discussion*—The essential characteristic of reproducibility is the different laboratories in which the testing is conducted.

3.2.5 *repeatability (short-term), r*—a repeatability estimate obtained under a short or brief time period.

3.2.5.1 *Discussion*—The time period may be minutes, hours, or days and needs to be specified for each test method.

3.2.6 *repeatability (long-term), r*—a repeatability estimate obtained over a long time period.

3.2.6.1 *Discussion*—The time period may be weeks or months and needs to be specified for each test method.

3.2.6.2 *Discussion*—Events that influence long-term repeatability are the use of different operators, environmental factors (such as seasonal variations in temperature, humidity, etc.), and the recalibration or adjustment, or both, of equipment.

3.2.7 *reproducibility (short-term), R*—a reproducibility estimate obtained over a short time period.

3.2.7.1 *Discussion*—The time period may be minutes,

hours, or days and needs to be specified for each test method.

3.2.8 *reproducibility (long-term), R*—a reproducibility estimate obtained over a long period of time.

3.2.8.1 *Discussion*—The time period may be weeks or months and needs to be specified for each test method.

3.2.8.2 *Discussion*—Events that influence long-term reproducibility are different operators, environmental factors (such as seasonal variations in temperature, humidity, etc.), and the recalibration or adjustment, or both, of equipment.

3.2.9 *repeatability (Type 1)*—Type 1 *r*, a repeatability estimate obtained in an interlaboratory program where the material(s) distributed to all laboratories is (are) in a prepared state ready for testing (with perhaps some minimal preparation steps required), such as Class I or II. See 5.2.1.

3.2.10 *reproducibility (Type 1)*—Type 1 *R*, a reproducibility estimate obtained in an interlaboratory program where the material(s) distributed to all laboratories is (are) in a prepared state ready for testing (with perhaps some minimal preparation steps required), such as Class I or II. See 5.2.1.

3.2.11 *repeatability (Type 2)*—Type 2 *r*, a repeatability estimate obtained in an interlaboratory program where some or all of the materials(s) distributed to all laboratories require a specified operation or series of operations, to produce the final test samples, portions, or test pieces prior to applying the test method to the material(s) or item(s) under test, to produce one test result (value), such as Class III.

3.2.12 *reproducibility (Type 2)*—Type 2 *R*, a reproducibility estimate obtained in an interlaboratory program where some or all of the material(s) distributed to all laboratories require a specified operation or series of operations, to produce the final test samples, portions, or test pieces prior to applying the test method to the material(s) or item(s) under test, to produce one test result (value), such as Class III.

3.2.13 *relative repeatability and reproducibility*—It is often appropriate to express repeatability and reproducibility on a relative basis, as a percent of a certain mean value. This is analogous to a coefficient of variation. Such expression is important when *r* and *R* vary with the mean level of the property being measured. Relative values for *r* and *R* cannot be unambiguously expressed as percentages alongside the actual measured values in usual test result units because some test methods have “percent” as their units, for example, % Cu, % elongation. To avoid this ambiguity the following symbols are defined by the use of the parentheses.

3.2.14 (*r*)—repeatability estimate expressed as percentage of the mean of the property for which the estimate was obtained.

3.2.15 (*R*)—reproducibility estimate expressed as percentage of the mean of the property for which the estimate was obtained.

3.2.16 *acceptance difference, (duplicate determinations), AD₂*—an established value, below which the difference between two “within-laboratory” *determinations* may be expected to lie, with a specified probability.

3.2.16.1 *Discussion*—The two *test determinations* are obtained at the “same” time (side-by-side) with identical test material, operators, and apparatus, and in the absence of other indications the probability is 95 %.

3.2.16.2 *Discussion*—If the calculated difference lies (below) the acceptance difference, the two values are accepted for averaging and the average is reported as the *test result*; if the calculated difference exceeds the acceptance difference, additional determinations are made to produce acceptable data.

3.2.17 *acceptance difference (x determinations), AD_x* —an established value, below which the maximum range (maximum value–minimum value) of a specified number of determinations (within a given laboratory) may be expected to lie, with a specified probability.

3.2.17.1 *Discussion*—The specified number of determinations are obtained at the “same” time (side-by-side) with identical test material, operators, and apparatus, and in the absence of other indications the probability is 95 %.

3.2.17.2 *Discussion*—If the calculated maximum range lies within the critical range or below the acceptance difference, all of the determinations are accepted for averaging or selection of a median value and the average or median is reported as the *test result*; if the maximum range exceeds the acceptance interval, additional determinations are made to produce acceptable data.

4. Significance and Use

4.1 Tests are conducted using established (standardized) test methods to generate test data. Test data are generated to make technical and scientific decisions for commercial processes and technical operations. It follows that the precision of a particular test method is an important quality characteristic of the test method and also of the decision process that involves the data. Therefore all test methods should be evaluated for precision.

4.2 Any evaluation of the precision of a test method is normally conducted with a group of typical materials or items subjected to measurement. The evaluation therefore represents “a snapshot in time” of the precision; the results are frequently unique to the materials, the participating laboratories, and the time period of the evaluation. A repeat of the entire evaluation at a later time with different materials and participants may not give good or exact agreement with any previous evaluation. This characteristic of precision evaluation should be clearly understood when reviewing precision data from various programs and at various time periods.

4.3 Although the evaluation of a test method for precision is an important quality characteristic of the method, the resulting precision parameters (r , R) have to be interpreted with caution if there is any thought of applying them across a broad range of material testing especially for consumer-producer product acceptance testing. Product acceptance testing protocols should be developed on the basis of precision data obtained in special programs that are specific to the commercial products or items and to the laboratories of the interested parties in this type of testing.

4.4 The application of this practice is limited to test methods 1) that have test results expressed in terms of quantitative continuous variable, and 2) that are fully developed and in routine use in a number of laboratories.

5. General Principles

5.1 This practice is prepared to accommodate a broad range of test methods. It may seem overly complex for test methods that occupy a narrow part of this broad spectrum of uses.

Therefore, make use of those portions of this practice that are applicable and ignore those parts that do not directly apply.

5.1.1 Although the terminology for repeatability and reproducibility is given in Section 3 of this practice, a general discussion is repeated here.

5.1.1.1 Repeatability refers to the ability of the *same* laboratory to obtain similar (test) results under certain specified conditions.

5.1.1.2 Reproducibility refers to the ability of *different* laboratories to obtain similar test results under certain specified conditions.

5.1.1.3 If test results closely agree, then good repeatability or good reproducibility exists.

5.1.2 The precision of a test method does not of necessity characterize a test with regard to how sensitive it is in measuring the basic property it is intended to measure. Precision may be good simply because the test method is insensitive to the basic property. A concept called “test sensitivity” has been defined in statistical literature as the ratio of the responsiveness of the test measurement to finite variations in the basic property in question, to the precision of the measurement. This practice does not address this issue.

5.1.3 Both repeatability and reproducibility should be determined under realistic or typical laboratory conditions. If extraordinary care is exercised (extremely homogeneous materials) the resulting precision is overly optimistic. Also as ordinarily determined, repeatability has both a test apparatus variability as well as a material variability. The sum of these two components is the repeatability as normally quoted.

5.2 *Interlaboratory Distribution Scheme (Test Pieces, Specimens, and Materials)*:

5.2.1 A key concept that must be clearly understood when contemplating interlaboratory precision testing is the matter of *what is distributed* to the participating laboratories. The “what” may be classified as follows:

5.2.1.1 *Class I—Fully prepared test pieces, specimens, (or test portions)*, requiring no further processing (preparation or adjustments) prior to testing (example—died-out, gaged dumbbells for stress-strain testing).

5.2.1.2 *Class II—Intermediate prepared materials*, that require some minimal processing prior to action by the test machine (example—cured rubber-sheets that must have dumbbells cut from them with subsequent gaging, prior to final stress-strain testing).

5.2.1.3 *Class III—Specified (quantities of) raw materials*, that must be processed into final samples, or specimens by a standardized procedure (example—rubber, curatives, carbon black, oils, and antioxidants that must be mixed, processing steps taken, cured sheets prepared, dumbbell test pieces cut and gaged prior to stress-strain testing).

5.2.2 The primary purpose of an interlaboratory program dictates which scheme, Class I, II, or III, is selected. If the attention is on the apparatus or test machine(s) in the various laboratories, how well these agree when testing the supplied test specimens, then Class I or perhaps Class II (both Class I and Class II being quite similar) would be selected.

5.2.2.1 However, if it is the *total operational sequence of a test*, such as mixing, processing, curing, *Documents*

gaging that is of interest, then Class III would be selected. Material distribution in accordance with Class III would frequently be called for in interlaboratory precision programs where producer-user acceptance testing of raw materials is of direct importance. An example would be carbon black or synthetic rubber.

5.2.2.2 In each case (Class I, II, or III) it is necessary that the distribution of items or materials is made from a uniform source or lot, with a nominally good uniformity or homogeneity.

5.2.3 The amount of “within-laboratory” preparation or processing, after arrival of the circulated items or material, increases in the order Class I, II, and III. Analytical chemistry and other simple physical tests often require no or very little “within-laboratory” preparation upon arrival of test portions and, therefore, make use of a Class I distribution scheme. Conversely, what may be called actual or quasi-performance tests require more complex “within-laboratory” preparation or processing and, therefore, require a Class III distribution. Performance implies the attainment of a certain minimal level of some critical property, tensile strength, or modulus, in a standard compound for a raw material like carbon black or a synthetic rubber.

5.2.4 The type of test method will often indicate the scheme of interlaboratory distribution; SBR is a typical example. The “quality” of SBR may be ascertained by (a) certain analytical tests such as fatty acid content, (b) certain simple physical tests, such as Mooney viscosity, or (c) by certain performance tests, (minimum) tensile strength, modulus, or cure rate. Here categories (a), (b), and (c) correspond respectively to Class I, II, or III distribution schemes.

5.3 Discussion of Repeatability (Very Short, Short, and Long Term):

5.3.1 In 5.2 attention was focused on interlaboratory precision; within-laboratory precision (repeatability) is now discussed. There are at least three different viewpoints that have been expressed with regard to repeatability.

5.3.1.1 *View 1*—The smallest possible or “very short” time period is used to estimate the variation. The same material, apparatus, and operator is used, and repeated determinations are made within a period measured in minutes or at most within a period measured in hours.

5.3.1.2 *View 2*—A “short” time period is used for the repeated operations that produce test results. The same material and same operator (or set of operators) is employed but the time period for the repeat operations is most frequently measured in days.

5.3.1.3 *View 3*—A “long-term” time period is used for the repeated operations that produce test results within a laboratory. This may be weeks or months. In this sense, although it may be possible to use the same material, different operators are often employed and due to the long-term nature certain other changes, such as recalibration of the test apparatus, may have taken place. These changed conditions produce increased variability.

5.3.2 The time period *must be specified* as each particular test method is taken up for consideration.

5.3.3 An important added feature is the concept of “accep-

tance differences” for individual sets of determinations. These may be called “checking limits.” Such acceptable difference values can have useful applications in analytical or other quickly repetitive operations, such as testing individual tensile-strength test specimens (dumbbells or rings). They permit the exclusion of outliers among the determinations.

5.3.4 It is anticipated that the “acceptable difference” repeatability will be calculated for determinations in the same way that ordinary repeatability is calculated for test results. Therefore, an extra set of calculations can be performed for individual determinations to permit estimates of AD_2 or AD_x to be obtained.

5.3.5 For any given test method a task group or subcommittee will normally choose one type of repeatability and reproducibility whether short term or long term.

6. Organizing an Interlaboratory Precision Program

6.1 *Task Group*—A task group of qualified people should be organized to conduct the program: a chairman, a statistical expert, and members well-experienced with the standard in question. The chairman should ensure that all instructions of the program are clearly communicated to all laboratories in the program. A supervisor in each laboratory should be chosen.

6.2 *Type of Precision*—The task group should make the following initial decisions.

6.2.1 The type of precision to be obtained (Type 1 or Type 2).

6.2.2 The time period of the repeatability and reproducibility estimate; short (minutes, hours, or days) or long (weeks or months). Define the time period.

6.2.3 Whether acceptance intervals are desired or needed.

6.2.4 These decisions set the stage for important but secondary decisions that naturally evolve from the structure of the program.

6.3 Laboratories and Materials:

6.3.1 The number of laboratories should be determined. The number of materials, each comprising a different level of the measured property, should be selected.

6.3.2 The number of laboratories available is seldom large, and if the test method is complex, or expensive to run, the problem is complicated further. Therefore, the problem is finding and obtaining the cooperation of enough qualified laboratories to produce meaningful estimates of precision, rather than a selection from a group of available laboratories.

6.3.3 At least ten participating laboratories are recommended. Practical considerations usually require that fewer than ten laboratories participate in the study. However, an interlaboratory study that involves fewer than six participating laboratories may not lead to reliable estimates of the reproducibility of the test method.

6.3.4 The number and type of materials to be included will depend on the range of the property and how precision varies over that range, the different types of materials to which the test method is applied, the difficulty (expense) in performing the tests, and the commercial or legal need for obtaining a reliable estimate of precision.

6.3.5 An interlaboratory study should include at least *three* materials, and for development of broadly applicable precision statements, *five* or more materials should be included. The term

“materials” is used in a broad generic sense. Materials may be raw or natural substances, manufactured products, etc. For each level of material, an adequate quantity (sample) of homogeneous material should be available for subdivision and distribution by random allocation to the participating laboratories. This supply of sample material should include a reserve of 50 % beyond the requirements for possible later use in retesting in one or more laboratories. When the material(s) to be tested is (are) not homogeneous, it is important to prepare the samples in the manner prescribed by the test method, preferably starting with one batch of commercial material for each level. Some modifications may be necessary to ensure that the amount of material available is sufficient to cover the experiment and keep a stock in reserve.

6.3.6 At each level, p , separate containers (the number of laboratories) should be used where there is any danger of the material deteriorating when the container has once been opened. In the case of unstable materials, special instructions on storage and treatment should be prescribed.

6.4 *Actual Organization of the Tests*—The interlaboratory test plan is as shown in Table 1, a table that indicates the laboratories, materials, and replicates. With q levels and n replicates, each participating laboratory among the p total laboratories has to carry out qn tests. A decision is necessary (for each test method) as to whether a “replicate” is to be a “determination” or a “test result” as defined in this practice. The performance of these tests should be organized and the operators instructed as follows:

6.4.1 All qn tests should be performed by one and the same operator or operator set, using the same equipment throughout.

6.4.2 Each group of n tests belonging to one level must be carried out under repeatability conditions, in a specified interval of time.

6.4.3 It is essential that a group of n tests under repeatability conditions be performed independently as if they were n tests on different materials.

6.4.4 The number of replicates, n , must be specified. Each replicate may be *one* test result or *one* determination in accordance with the requirements of the test method. Normally, n is two. A larger number may be specified if necessary.

6.4.5 In on-line statistical process control situations, a single determination is often considered a test result, particularly if the precision of a duplicate determination test result does not show cost effective improvement over that of a single determination test result. This can be helpful information for many users of a test method. It is at this planning stage that the decision has to be made whether or not to have the precision statement present both the precision of a single determination test result and a duplicate determination test result. If the decision is made to run duplicate determinations, the minimum testing required for each test material consists of two sets of duplicate determinations conducted on each of two different days.

6.5 Instructions to Operators:

6.5.1 The operators should receive no instructions other than those contained in the test method; these should suffice.

6.5.2 Prior to testing, the operators should be asked to comment on the standard and state whether the instructions contained in it are sufficiently clear.

6.5.3 All participating laboratories should report their test results to one more significant figure than is customary or prescribed in this practice.

6.6 *Reporting the Test Results*—Each laboratory supervisor should write a full report on the tests containing the following particulars:

6.6.1 The final test results (avoid transcription and typing errors).

6.6.2 The original individual observations and determination values from which the final results were derived. This is required if “acceptable difference” parameters (AD_2 or AD_x) are to be calculated.

6.6.3 The date(s) on which the samples were received and the date(s) and time(s) on which they were tested.

6.6.4 Comments and information about irregularities or disturbances that may have occurred during the test.

6.6.5 Information about the equipment used, and other relevant information.

6.7 The results should be reported using the format given in Table 1.

7. Analysis Concepts for Interlaboratory Test Data

7.1 The analysis of interlaboratory data to evaluate test method precision is conducted as a “one-way” analysis of variance for each level or material in the test program. Annex A1 gives the basic statistical model for such an analysis. This annex should be reviewed to become familiar with the potential sources of variation in the database being investigated and to better appreciate the results of the precision calculations. This annex also gives the basic expressions for r and R .

7.1.1 *Outliers*—Outliers are test result and derived test result values, that deviate so much from the bulk of the data (for a certain level) that they are considered to be irreconcilable with remainder of the data. Although some care must be exercised in handling outliers, experience has shown that a certain small fraction of laboratories in any interlaboratory test program may produce outlier values. The most frequent causes are either testing blunders or inadequate control over internal testing conditions (poor test procedures, test machine maintenance, calibration). The outlier problem is addressed on the

TABLE 1 Original Test Results^A

Laboratory	Level	1	2		j		q
1							
2							
i					y_{i1} ... y_{ijk}		
p							

^AThe following notation is used:

(a) Laboratories, there are p as a total

$L(i = 1, 2 \dots p)$

(b) Materials or levels, there are q as a total

$m_j (j = 1, 2 \dots q)$

(c) Replicates, there are n as a total in each cell or L, m_j combination. There

are normally an equal number of n values (usually 2) in each cell.

(d) y_{ijk} is a single test result value.

Example—Cell (i, j) contains n_{ij} results $y_{ijk} (k = 1, 2, \dots n_{ij})$.

basis of h and k values as developed in Practice E 691. See Annex A2 and Annex A3 for background on the development of these h and k statistics and the rationale for the 95 % confidence level used for outlier rejection.

7.2 Preliminary Analysis—A preliminary analysis of the database consists of the following two initial steps:

7.2.1 Tabulate the data in the format as given in Table 1. In this table the number of laboratories is designated by p , the number of materials (levels) by q , and the number of test result replicates by n . The table contains pq “cells,” each cell containing n replicates for the usual condition of an equal number of replicates per cell. In most interlaboratory test programs for precision, $n = 2$.

7.2.2 Inspect the data for any unusual results detectable by simple review. If any unusual data values are discovered make a note and proceed as described as follows.

7.3 Full Analysis—The full analysis of the precision data is normally conducted in two parts. Part 1 is an analysis of all of the data as reported by all participating laboratories. This analysis as described below, will generate additional tables that are used to identify any outliers in the database. If no outliers are found, the required precision parameters are calculated from the (original) database.

7.3.1 Outlier Rejection—If outliers are present, outlier rejection techniques are used to eliminate the indicated data values. After the outliers are removed and replaced by data values in accordance with 7.5 (handling outlier and missing values), a Part 2 reanalysis is conducted on the adjusted database. This Part 2 analysis yields the precision parameters that are used to prepare the precision section of the standard.

7.4 Part 1 Analysis—Conduct a Part 1 analysis in accordance with Practice E 691 calculation algorithms (these are given in Section 8 and Annex A2 and Annex A3 of this practice) on the data as it exists in a Table 1 format. This is done using either (1) the Practice E 691 computer (software) program,⁵ or (2) typical spreadsheet calculation procedures. Four main steps in accordance with 7.4.1 to 7.4.4 are required. The Practice E 691 computer program performs all four steps and generates the required tables in addition to subsequent calculations for r and R . If spreadsheet calculations are performed, separate table generation steps may be required as follows:

7.4.1 Calculate the *average* of each cell in the Table 1 layout and tabulate the averages as in Table 2. For this standard

7.4.2 Calculate the *standard deviation* for each cell in the layout as shown in Table 1. (See Note 1.) Tabulate the calculated standard deviations as in Table 3.

7.4.3 Calculate the *h-value* for each cell in the Table 1 layout. See Annex A2 for calculation and other details. Prepare a table of h -values in the same format as Table 2, if a spreadsheet calculation is used.

7.4.4 Calculate the *k-value* for each cell in the Table 1 layout. See Annex A3 for calculation and other details. Prepare a table of k -values in the same format as Table 3, if a spreadsheet calculation is used.

NOTE 1—Many spreadsheet calculation algorithms for standard deviation use n , the number of values in the calculation, as a divisor for the sum of squares in the calculation of a standard deviation. The divisor should be $(n - 1)$. If n is used, correct the spreadsheet standard deviations by multiplying them by $[n/(n - 1)]^{1/2}$.

7.4.5 Review of Calculations—Review the tables of h -values and k -values in accordance with the procedures in Annex A2 and Annex A3. Reject any cell averages, (h -values), that are significant at the 95 % ($p = 0.05$) confidence level. Reject any cell standard deviations, (k -values), that are significant at the 95 % ($p = 0.05$) confidence level.

7.4.6 If no cell averages or cell standard deviations are rejected, the Part 2 analysis is not required and the calculations for S_r , S_R , r , R , (r), and (R) may be made in accordance with Section 8.

7.5 Blank or Missing Cell Values—If any outlier rejections are made, or if there are missing data in the original database, the problem of blank cells in the Table 1 format must be addressed. The recommended method to replace any blank cells is the use of a special or average value for the missing cell value in accordance with the instructions as given in the next section.

7.5.1 Cell Replacement for Practice E 691 Computer Analysis—If the Practice E 691 computer analysis is used, the blank cell replacement values must be inserted into the database in the Table 1 format and a reanalysis conducted. The Practice E 691 computer program is not structured to accommodate ‘blank’ data cells. The *replacement* test result values must be inserted into any *cell* so that *both* the recalculated average or the original average and recalculated standard deviation (variance) if both are observed for that level or the original standard deviation, are preserved or unchanged by the addition of the replacement values.

7.5.1.1 The *recalculated average*, is the (new) average calculated after removing the cell average outlier value(s) from the indicated cell(s).

7.5.1.2 The *recalculated standard deviation (and variance)*, is the standard deviation (variance) calculated after removing

TABLE 2 Cell Averages^A

Laboratory	Level	1	2		j		q
1							
2							
i					$\bar{y}_{ij}$		
p							

^A $\bar{y}_{ij}$ = cell average.

“average” refers to arithmetic mean.

TABLE 3 Cell Variance or Standard Deviation^A

NOTE 1—Uniform-Level Experiment

Laboratory	Level	1	2	j		q
1						
2						
i					S_{ij}	
p						

^ASymbols are defined as follows:
 s_{ij} = cell standard deviation.

⁵ The software for the Practice E 691 analysis may be purchased from ASTM, 1916 Race St., Philadelphia, PA 19103. Request PCN:12-506910-34.

the cell standard deviation outlier value(s) from the indicated cell(s). The technique for cell value replacement under the stipulations in accordance with 7.5.1, is described in 7.5.2 for spreadsheet analysis and also in Annex A7.

7.5.2 Cell Replacement for Spreadsheet Analysis—If a spreadsheet analysis is used a number of intermediate tables will be needed in the spreadsheet in addition to the five tables as specified in 7.4. In addition to Table 1, Table 2, and Table 3 and the *h*-value and *k*-value tables, the following tables for the Part 1 analysis are recommended—Cell Average Deviation, *d*, and Cell Standard Deviation Squared (that is, Variance). The cell average deviation table is used in the construction of the *h*-value table. The cell standard deviation squared table is used in the calculation of the pooled *S_r* and in the operation to replace blank cell standard deviation values.

7.5.2.1 In Annex A7 an example of precision analysis is given for the spreadsheet approach. This annex illustrates how a number of supplementary spreadsheet calculations are made. Refer to this annex for additional details on cell replacement operations and for some general comments on the outlier problem in precision analysis.

7.5.3 If more than one outlier of a given type (cell average or cell standard deviation) is rejected for a particular laboratory and if the cell values for other materials in general appear to be out-of-line (although not officially rejected) with results of the other laboratories, serious consideration should be given to totally eliminating the laboratory from the database for analysis.

7.6 Part 2 Analysis—After all blank cells have been replaced with appropriate averages after (1) any outlier rejection operations, or (2) missing cell values have been allowed for, the adjusted database shall be subjected to a Part 2 analysis. From this second analysis, calculate *S_r*, *S_R*, *r*, *R*, (*r*), and (*R*) in accordance with Section 8.

7.7 Preparation of Research Report for Precision Evaluation—All precision evaluation programs shall be well documented by the preparation of a research report that shall be placed on file at ASTM Headquarters. This report shall contain important information concerning the interlaboratory program as follows:

- 7.7.1 Test method designation,
- 7.7.2 Number and identification of participating laboratories,
- 7.7.3 Materials used, identification or formulations, or both,
- 7.7.4 Type of precision evaluated; time period of precision (hours, days, weeks),
- 7.7.5 Dates of test program,
- 7.7.6 Basic (raw) data obtained, in Table 1 format,
- 7.7.7 Calculations performed for evaluating precision parameters, including method used for outlier rejection and method used for replacing missing values,
- 7.7.8 Results of precision calculations in Table 4 format, and
- 7.7.9 Any unusual outcome of the test program.

8. Calculating the Precision Parameters

8.1 Although Annex A1 gives substantial background and discussion on the repeatability variance and standard deviation, on the between-laboratory variance of cell averages and on the

**TABLE 4 Example—ASTM XXXX Type 1—Precision^A
(Measured Property = XXXX in MPa)**

NOTE 1—If *AD₂* or *AD_x* is determined, the results may be given in a table similar to Table 4.

NOTE 2—Pooled or average values for all tabulated parameters may be given if appropriate.

NOTE 3—*p* = xx, *q* = 4, *n* = 2.

Material	Mean Level, (MPa)	Within Laboratories ^B			Between Laboratories ^B		
		<i>S_r</i>	<i>r</i>	(<i>r</i>)	<i>S_R</i>	<i>R</i>	(<i>R</i>)
A	XX	X	X	X	X	X	X
B	XX	X	X	X	X	X	X
C	XX	X	X	X	X	X	X
D	XX	X	X	X	X	X	X
Pooled or Average Values	XX	X	X	X	X	X	X

^AThe time period for precision is days.

^BSymbols are defined as follows:

S_r = within-laboratory standard deviation.

r = repeatability (in measurement units).

(*r*) = repeatability (in percent).

If actual measurement units are %, these values represent percent relative, such as, percent of a percent.

S_R = standard deviation for total between-laboratory variability.

R = reproducibility (in measurement units).

(*R*) = reproducibility (in percent).

reproducibility variance and standard deviation, the basic calculation algorithms for these parameters are given in this section. The calculations apply to each material.

8.1.1 Repeatability Variance, Standard Deviation—For any material, the repeatability variance designated by (*S_r*)² is calculated in accordance with Eq 1.

$$(\bar{S}_r)^2 = \sum (1 \text{ to } p)(Si)^2/p \quad (1)$$

where:

(*Si*)² = cell variance for Laboratory *i*, and
p = total number of laboratories.

The repeatability standard deviation is given in Eq 2.

$$\bar{S}_r = [\sum (1 \text{ to } p)(Si)^2/p]^{1/2} \quad (2)$$

8.1.2 Between-Laboratory Variance—A derived intermediate parameter is the term called the 'between-laboratory' variance, designated by (*S_L*)². This is evaluated from the variance of the 'cell averages,' (laboratory averages for any level), designated by (*S²*)_{*L*} $\bar{x}$, and the repeatability variance.

$$(\bar{S}_L)^2 = (\bar{S}^2)\bar{x} - (\bar{S}^2)/n \quad (3)$$

The term (*S_L*)² is used in the calculation of the reproducibility variance and standard deviation in accordance with 8.1.3. It can also be used as an indicator of the inherent variation between laboratories without the influence of the within-laboratory variation. Experience has shown, however, that the within-laboratory variation is substantially smaller than between-laboratory variation. In certain circumstances (*S_L*)² may calculate to less than zero; if this occurs (*S_L*)² is set equal to zero. This less than zero situation may occur when there is substantial within cell variation of such a nature that when laboratory cell averages are calculated, they agree quite well.

8.1.3 Reproducibility Variance, Standard Deviation—The (total) variance among all the values for a given material is defined as the reproducibility variance, in accordance with Eq 4.

$$(S)_R^2 = (S)_L^2 + (S)_r^2 \quad (4)$$

Substituting for $(S)_L^2$ produces Eq 5.

$$(S)_R^2 = (S)^2\bar{x} - (S)_r^2/n + (S)_r^2 \quad (5)$$

Simplifying and taking the square root produces Eq 6.

$$(S)_R = [(S)^2\bar{x} + (S)_r^2(n-1)/n]^{1/2} \quad (6)$$

8.2 The calculations for the above parameters are provided as part of the output of the Practice E 691 computer software program. For spreadsheet analysis the usual spreadsheet calculation procedures may be used as well as specific calculations set up in the form of macro commands. Annex A6 also contains computational formulas that may prove to be beneficial for spreadsheet precision calculations. This annex contains the formula for unequal numbers of n replicates per cell.

8.3 Annex A4 describes the calculations for discovering whether a functional relationship exists between r , R , (r) or (R) , and the mean level M .

8.4 Annex A5 is addressed to carbon black testing. It describes a special treatment of within-cell test values (test results) and their review for data consistency or outlier behavior. It also specifies a special procedure for selecting the mode of precision parameter expression, either absolute or relative, for both reproducibility and repeatability.

8.5 Annex A7 previously discussed, is an example of a typical precision evaluation for Mooney viscosity. All calculations are included in this example.

9. Format for Precision and Bias Section (Clause) of Standard

9.1 The results of the formal analysis shall be contained in a specific section or clause of the test method entitled "Precision and Bias."

9.2 *Introductory Subclauses*—The precision and bias section shall begin with two paragraphs giving important details on the interlaboratory program.

9.2.1 A statement citing that Practice D 4483 is the reference document for the precision section.

9.2.2 A caveat statement on the general applicability of the precision results, in accordance with 9.2.2.1.

9.2.2.1 The precision results in this precision and bias section give an estimate of the precision of this test method with the materials (rubbers, etc.) used in the particular interlaboratory program as described below. The precision parameters should not be used for acceptance or rejection testing of any group of materials without documentation that they are applicable to those particular materials and the specific testing protocols of the test method.

9.3 A second subclause shall consist of one or more paragraphs that give details on the interlaboratory program followed by one or more tables of results of the precision testing. The introductory paragraphs should answer the following questions:

9.3.1 What type precision was estimated, Type 1 or Type 2?

9.3.2 What is the time period for repeatability, reproducibility—short term (define), long term (define)?

9.3.3 What is a test result? How many determinations? Average (mean) or median?

9.3.4 How many laboratories participated (p)?

9.3.5 How many materials (q)?

9.3.6 How many replicates (n)? What is a replicate?

9.3.7 At what time was the interlaboratory program conducted (month, year)?

9.3.8 Are there any unusual results that the reader should be aware of?

9.3.9 How do r and R vary as the mean level of the measured property varies? Can these variations be described by a simple mathematical relationship (linear, log, etc.)? See the Annexes.

9.4 *Table of Precision Parameters*—A table with the general format such as Table 4 should be prepared. This includes the following information:

9.4.1 ASTM test method designation and year of issue,

9.4.2 Type of precision; time period used for r and R ,

9.4.3 Measured property,

9.4.4 Materials, with mean level and units of measurement, and

9.4.5 r , (r) , R , (R) , and for completeness of record the within and between laboratory standard deviation, s_r and S_R .

9.5 *Pooled Values for Table 4 Format*—If pooled or average values, or both, for the precision parameters set up in the format of Table 4 are desired, use the following procedure.

9.5.1 *Average*—The average applies to the column of mean level values only. The (arithmetic) average is calculated in the normal manner.

9.5.2 S_r and S_R —For these two parameters, the pooled values are the square root of the mean variance of each column (of standard deviation values).

9.5.3 r and R —These parameters are equal to their respective standard deviations multiplied by 2.83 (standard deviation times a constant) and therefore are to be pooled by the same procedure as for S_r and S_R .

9.5.4 (r) and (R) —There are two options for calculating the pooled values for these two relative (percent) precision parameters.

9.5.4.1 *Option 1*—For each row of the table, these parameters are also equal to a standard deviation times a constant. But the constant $[2.83 \times (1/\text{mean level value}) \times 100]$ changes for each row of the table. Therefore one pooling method is to obtain the square root of the mean value of each row value squared, in the (r) column and the (R) column.

9.5.4.2 *Option 2*—The alternative pooling method is to calculate (r) and (R) by dividing the respective r and R by the average mean level value (bottom of Column 1 mean level value) and multiplying by 100.

9.5.5 Experience shows that the two options do not give exact agreement. The recommended method is Option 2. The option adopted is not really very critical; the pooled value is simply a *general indicator* of overall precision and minor differences are of no substantial consequence.

9.6 *Significant Figures in Precision Table*—Computer calculations frequently generate several figures or decimal places after the decimal point. All the values placed in the precision table should be rounded to the number of figures after the