Designation: D 4483 – 99

Standard Practice for Determining Precision for Test Method Standards in the Rubber and Carbon Black Industries[1]

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:

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 Method[3]

2.2 *ISO Standard:*

ISO 5725 Precision of Test Methods—Determination of Repeatability and Reproducibility by Interlaboratory Tests[4]

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), AD_2*—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*—f 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 acceptancetesting 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, die-cutting, and

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 "acceptance 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.

TABLE 1 Original Test Results[A]

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

[A]The following notation is used:
(*a*) Laboratories, there are *p* as a total $L_i (i = 1, 2 \ldots p)$

(*b*) Materials or levels, there are *q* as a total $m_j (j = 1, 2 \ldots q)$

(*c*) Replicates, there are *n* as a total in each cell or $L_i m_j$ 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} ($k = 1, 2, \ldots n_{ij}$).

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

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

TABLE 2 Cell Averages[A]

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

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

"average" refers to arithmetic mean.

[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.

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 3 Cell Variance or Standard Deviation[A]

NOTE 1—Uniform-Level Experiment

Level / Laboratory	1	2	*j*		*q*
1					
2					
i			s_{ij}		
p					

[A] Symbols are defined as follows:
s_{ij} = cell standard deviation.

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

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

NOTE 1—If AD_2 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]		
		s_r	*r*	(*r*)	S_R	*R*	(*R*)
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

[A] The time period for precision is days.

[B] Symbols 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).

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

$$(S)^2_r = \Sigma\,(1\ to\ p)(Si)^2/p \tag{1}$$

where:
$(Si)^2$ = cell variance for Laboratory i, and
p = total number of laboratories.

The repeatability standard deviation is given in Eq 2.

$$S_r = [\Sigma\,(1\ to\ p)(Si)^2/p]^{1/2} \tag{2}$$

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

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

The term $(S^2)_L$ 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 $(S^2)_L$ may calculate to less than zero; if this occurs $(S^2)_L$ 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.

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

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

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

Simplifying and taking the square root produces Eq 6.

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

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 × (1/mean level value) × 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

decimal point that is realistic from the standpoint of the measurement capability of the test method. This is very frequently only one or two decimal places not counting any leading zeros for values smaller than unity. The relative precision parameters (*r*) and (*R*), should be given to only one figure after the decimal point for values below 100 and to no figures (whole numbers) for values above 100.

9.7 *Statements for Precision*:

9.7.1 Typical statements for the precision section or clause of a standard shall be listed in accordance with one of two styles, either 9.7.1.1 and 9.7.1.2 or 9.7.2.1 and 9.7.2.2.

9.7.1.1 The *difference* between two single test results (or determinations) found on identical test material under the repeatability conditions prescribed for a particular test will exceed the *repeatability* on an average of not more than once in 20 cases in the normal and correct operation of the method.

9.7.1.2 The *difference* between two single and independent test results found by two operators working under the prescribed reproducibility conditions in different laboratories on identical test material will exceed the *reproducibility* on an average of not more than once in 20 cases in the normal and correct operation of the method.

9.7.1.3 These two statements apply to either a particular mean level in a precision table (see Table 4) or to an overall level common to a standard or table, which is designated as a ′pooled’ value, that is, a special average value (see 9.5). The statement should make it clear which type of precision value is addressed, individual mean levels in a table or a pooled value.

9.7.2 Alternatively, statements of the following form may be prepared for use in the Precision clause of any test method.

9.7.2.1 *Repeatability*—The repeatability of test *xxxx* has been established as *xxxx*. Two single test results (or determinations) that differ by more than *xxxx* (expressed in appropriate terms) must be considered suspect, that is, to have come from different sample populations. Such a decision dictates that some appropriate action be taken.

NOTE 2—Appropriate action may be an investigation of the test method procedure or apparatus for faulty operation or the declaration of a significant difference in the two materials, samples, etc., which generated the two test results.

9.7.2.2 *Reproducibility*—The reproducibility of test *xxxx* has been established as *xxxx*. Two single test results (or determinations) produced in separate laboratories that differ by more than *xxxx* (expressed in appropriate terms) must be considered as suspect, that is, that they represent different sample populations. Such a decision dictates that appropriate investigative or technical/commerical actions be taken.

9.7.2.3 These two statements apply to either a particular mean level in a precision table (see Table 4) or to an overall level common to a standard or table, which is designated as a ′pooled’ value, that is, a special average value (see 9.5). The statement should make it clear which type of precision value is addressed, individual mean levels in a table or a pooled value.

9.7.2.4 Repeatability and reproducibility expressed as a percentage of the mean level, (*r*) and (*R*), have equivalent application statements as above for *r* and *R*. For the (*r*) and (*R*) statements, the difference between the two test results is expressed as an arithmetic mean (average) of the two test results.

9.7.3 *Bias Statement*—For most test methods bias cannot be determined. In that case, the following statement is recommended:

9.7.3.1 *Bias*—In test method terminology, bias is the difference between an average test value and the reference (true) test property value. Reference values do not exist for this test method since the value or level of the test property is exclusively defined by the test method. Bias, therefore, cannot be determined.

9.7.3.2 For those test methods where bias can be determined, a statement as to its magnitude should be included.

9.8 *Modification of Precision Table Format*—If for certain technical reasons, the precision table format as specified above is considered inappropriate for any particular test method standard, a modified format may be used. If a modified format is adopted, a paragraph shall be inserted into the precision section, clearly documenting the need for the modified format and explaining the modifications made.

10. Keywords

10.1 accuracy; interlaboratory study; precision; repeatability; reproducibility; statistics

ANNEXES

(Mandatory Information)

A1. STATISTICAL MODEL FOR PRECISION TESTING

A1.1 *Basic Statistical Model:*

A1.1.1 For any established measurement system, each measurement *y*, can be represented as indicated by Eq A1.1.

$$y = M + \Sigma d(j) \quad \text{(A1.1)}$$

where:

M = value obtained for a measurement when all deviations, d(j), are zero, that is, the ideal outcome of a measurement, and

$\Sigma d(j)$ = (algebraic) sum of (j) individual deviations or measurement perturbations, generated by whatever "system-of-causes" that exists for the measurement system.

A1.1.2 The term M is expressed in practice, for any measurement system, as the average of all y-values in the overall measurement program; it is also called the level of the property. (The term M is used in this Annex in place of μ, frequently defined as the true value). A more useful format is obtained when Eq A1.1 is expressed as an expanded model in Eq A1.2, where $\Sigma d\ (j)$ is replaced by a series of terms appropriate to interlaboratory testing.

$$y = M + B_i + B_m + B_L + B_g + e_b(l) + e_b(s) + e_w(l) + e_w(s) + e(g) \quad \text{(A1.2)}$$

where:

B_i = inherent bias or systematic deviation, characteristic of the design of the measurement system; it exists under all measurement conditions,

B_m = bias (systematic deviation) contributed by the measuring machine; it is unique to a particular machine,

B_L = bias contributed by the laboratory; it is unique to conditions in a particular laboratory,

B_g = general (generic) bias of a "to be specified" nature (certain measurement systems may require more than one such term),

$e_b\,(l)$ = between-laboratory random deviation of long-term nature, that is, over a period of several weeks or months,

$e_b\,(s)$ = between-laboratory random deviation of short-term nature, that is, over a period of days,

$e_w\,(l)$ = within-laboratory random deviation of a long-term nature (weeks, months),

$e_w\,(s)$ = within-laboratory random deviation of a short-term nature (days), and

$e\,(g)$ = general (generic) random deviation of a "to be specified" nature (certain measurement systems may require more than one such term).

A1.1.3 In a perfect measurement world all biases and random deviations of Eq A1.2 would be zero. In the real world of measurement, these terms take on certain values and the sum of their collective values acts as a perturbation of the M value for each measurement. Both the actual value and the variance of each of these terms are important when considering testing and precision programs. Tests to determine the significance of these individual terms usually involve a statistical comparison of the variances attributed to the terms.

A1.2 *The (B) or Bias Terms:*

A1.2.1 The value of the (B) terms is dependent on the measurement system or the system-of-causes, for the generation of the biases. The (B) terms in the model may be either fixed or variable as well as plus or minus, depending on the measurement system under consideration. For any system, the variable (B) terms are typically a non-random finite distribution and therefore the values for a particular bias term will not of necessity sum to zero over the population constituting the system. Bias terms that are fixed under one system of causes may be variable under another different system of causes and vice-versa.

A1.2.2 The inherent bias B_i is characteristic of the overall design of the machine or apparatus. This type of bias is frequently of importance in chemical tests for certain constituents whose theoretical content can be calculated, for example, percent chlorine in sodium chloride. A given test device may always be low or high due to unique design features.

A1.2.3 One or more generic bias terms, B_g, may be included in the model to allow for any (non-inherent) unique systematic deviation not attributable to test machines or laboratories.

A1.2.4 The bias terms B_m and B_L apply to most types of testing. As an example, for a particular laboratory (with one test machine) both of these bias terms would be constant or fixed. For a number of test machines, all of the same design in a given laboratory, B_L would be fixed but B_m would be variable, each machine having a unique value. For a measurement system consisting of a number of typical laboratories, both B_m and B_L would be variable for the multilaboratory measurement system but of course both B_m and B_L would be constant for each laboratory.

A1.3 *The (e) or Random Terms:*

A1.3.1 The (e) terms represent random deviations, plus or minus values that have an expected mean of zero (over the long run) and a variance equal to var(e). The distribution of the (e) terms is assumed to be approximately normal but in practice it is usually sufficient if the distribution is unimodal. The (random) value of each of the (e) terms influences the measured y-value on an individual measurement basis. However in the long run when y-values are averaged over a number of measurements, the influence of the (e) terms is greatly diminished or eliminated since the terms average out to zero (or approximately zero) and the y-value (and M) is perturbed by the (B) terms only. This long run zero-average character stands in contrast to the behavior of the fixed (B) terms where an increased number of measurements increases the knowledge (accuracy) of the actual (B) value.

A1.3.2 To make the model building as accurate as possible as in the case of the bias terms, one or more generic random deviation terms, $e\,(g)$, may be included in the model to account for any potential source of special random deviations not attributable to the general or common 'within' or 'between' laboratory categories.

A1.4 *Relating the (B) Terms to Measured Precision:*

A1.4.1 The expanded series of (B) terms in Eq A1.2 gives insight into the potential individual sources of bias between laboratories. However to express the between laboratory test results in relation to the (B) terms, it is convenient to use a collective (B) term designated as (B)Total, which is the (algebraic) sum of all (B) terms. The variance of (B)Total is the between-laboratory bias variance. The total between-laboratory variance is the sum of the between-laboratory bias variance and the between-laboratory random variance, e_b, (either long or short) and is given by Eq A1.3.

$$\text{Var}[(B)\text{Total}] + Var[e_b] = (\sigma^2)_L$$

= between−laboratory variance (A1.3)

The between-laboratory variance does not include the random within-laboratory variation. The value of $(\sigma^2)_L$ (for any material), is estimated in accordance with Eq A1.4, from the between-laboratory variance of cell averages, $(S^2)\bar{x}$, diminished by the adjusted value of $(S^2)_r$, the pooled within-cell variance. See Section 8.

$$(\sigma^2)_L = (S^2)\bar{x} - (S^2)_r/n = (S)^2_L \quad \text{(A1.4)}$$

The normal pooled within-cell variance, $(S^2)_r$, is adjusted or divided by n, the number of test values per cell, to put both of the variances in the equation on the same basis, that is, averages of n values.

A1.4.2 In Eq A1.4 and those to follow, population statistics are represented by Greek letter symbols and the estimates of the statistics are represented by English letter symbols. In Eq A1.4 the estimate $(S)^2_L$, is equated to the population statistic $(\sigma^2)_L$.

A1.5 *Relating the (e) Terms to Measured Precision*—The expanded series of random (*e*) terms gives insight into the individual sources of random deviations (or errors) that perturb the *M* value. However as in the case of the (*B*) terms, for any specific precision program with a defined time period for repeat tests, it is easier to relate the test results to precision evaluation by selecting one e_b and one e_w term, that is, commonly either a (*l*) long or (*s*) short time period; other time periods may be specified if needed.

A1.5.1 *Within-Laboratory (e) Term Evaluation*:

A1.5.1.1 Within a single laboratory, repeated testing on a given material generates a series of values for $e_w(l)$ or $e_w(s)$ depending on the time scale for measurements. From the series of such repeat measurements the simplest expression of within-laboratory variance of e_w is given by Eq A1.5. For simplicity the (*l*) and (*s*) notations will be dropped and e_w alone used with the assumption that either time span can be used in the developed relationships.

$$\text{Var}[e_w] = (\sigma)^2 = \text{simple within−laboratory variance} \quad \text{(A1.5)}$$

This applies to a particular laboratory and to a particular material.

A1.5.1.2 It is the general practice in precision analysis to assume that $(\sigma)^2_w$ will be approximately equal from laboratory to laboratory for any well-established and standardized test method and on this basis the individual cell estimates of $(\sigma)^2_w$ can be pooled for any material to obtain a collective value representing all laboratories. However the skill and internal control procedures used in conducting test measurements varies among even well-experienced laboratories and this will be reflected in the pooled $(\sigma)^2_w$ variance for any given material.

A1.5.1.3 This varying testing skill situation can be addressed by use of the generic term $e_w(g)$. Thus a more realistic estimate of within-laboratory variance for any given laboratory is expressed by Eq A1.6, a variance specific to a given laboratory.

$$\text{Var}[e_w] + \text{Var}[ew(g)] = (\sigma)^2_w$$

= specific within−laboratory variance (A1.6)

where:

$Var[e_w]$ = basic within-laboratory variance, a variance that is characteristic of routine use of the test method, that is, uniform over all laboratories, and

$Var[e_w(g)]$ = an added within-laboratory variance (component) specific or unique to a particular laboratory; it is approximately zero for good laboratories.

The simple variance of A1.5.2.1 has been redefined as a basic variance. The specific within-laboratory variance defined by Eq A1.6, which contains two components, may also be called the specific repeatability variance, $(\sigma)^2_r$, unique to any one laboratory, and is given by Eq A1.7.

$$\text{Var}[e_w] + \text{Var}[e_w(g)] = (\sigma)^2_r$$

= specific repeatability variance (A1.7)

When the individual repeatability variances for all laboratories are pooled the relationship is expressed by Eq A1.8, where the estimated value, $(S)^2_r$, is used in the equality.

$$\text{Pooled } (\sigma)^2_r = (S)^2_r = \text{repeatability variance} \quad \text{(A1.8)}$$

Since in typical interlaboratory programs there is usually only 1 degree of freedom (DF) estimate of $(\sigma)^2_r$ for each laboratory and material, the pooled $(S)^2_r$ is the parameter of direct importance.

A1.5.2 *Between-Laboratory (e) Term Evaluation*—The term e_b, either long or short time span, represents random variations between (among) a group of laboratories that measure a common material and as such e_b is one component of the overall laboratory variation. Interlaboratory test programs do not ordinarily provide a direct estimate of e_b in the same sense that e_w is evaluated.

A1.6 *Combined (B) and (e) Term Between-Laboratory Evaluation* —The total variation among between-laboratory test results (for any material) which is defined as the reproducibility variance, $(\sigma)^2_R$, is the sum of four sources or components of variance, for any selected time period, as given by Eq A1.9.

$$\text{Var}[(B)\text{ Total}] + \text{Var}[e_b] + \text{Var}[e_w(g)] + \text{Var}[e_w] = (\sigma)^2_R \quad \text{(A1.9)}$$

The estimate of this variance, $(S)^2_R$, is equal to the total variance or mean square, among all the values for each material in the interlaboratory program. Recall that (*B*)Total represents a number of potential separate sources of bias as given in Eq A1.2. Interlaboratory testing experience has shown that the order of the variance terms in Eq A1.9 (left to right), is the approximate order of magnitude of these terms.

A1.7 *Relationship Between (B) and (e) Terms and Precision Parameters r and R:*

A1.7.1 Repeatability, *r*, is defined by Eq A1.10 in terms of the estimated statistic rather than the population statistic.

$$\text{repeatability} = r = \phi\,(2)^{1/2}S_r \quad \text{(A1.10)}$$

A1.7.2 Reproducibility, *R*, is defined by Eq A1.11 on the same basis.

$$\text{reproducibility} = R = \phi\,(2)^{1/2} S_R \quad \text{(A1.11)}$$

A1.7.3 The coefficient $(2)^{1/2}$ is derived from the fact that *r* and *R* are equal to the difference between two (single) test results. The factor ϕ depends on both the total degrees of freedom (number of test results available) in the estimation of the variances $(\sigma)^2_r$ and $(\sigma)^2_R$ and on the shape of the distributions of the variable bias terms and the (*e*) terms. The normal assumptions for these terms are (1) unimodal distributions, (2) the number of test results not too small (approximately 20), and (3) a confidence level ($p = 0.05$) of 95 %. Under these assumptions the value of ϕ is approximately 2 and therefore Eq A1.10 and Eq A1.11 may be rewritten as

$$\text{repeatability} = r = 2.83\, S_r \quad \text{(A1.12)}$$

$$\text{reproducibility} = R = 2.83\, S_R \quad \text{(A1.13)}$$

A2. PRACTICE E691 CALCULATIONS FOR 'CELL AVERAGE' OUTLIERS: *h*-VALUES

A2.1 *General Background*—Practice E 691 was originally introduced in 1979 as the basic document for performing precision analysis for all ASTM test method standards. It was most recently revised in 1987. The fundamental calculation algorithms for *r* and *R* used in Practice E 691 are the same as found in Practice D 4483 (1989 and current version), in ISO TR 9272 used by ISO TC45 and in the generic ISO standard, ISO 5725.

A2.2 Practice E 691 differs however from all of these other standards in the way it addresses outliers or potential outliers. The other standards evaluate potential outliers on the basis of (1) Cochran's test for within-cell variances, across all laboratories for each material, and (2) Dixon's test for within-cell averages, across all laboratories for each material. Practice D 4483 in its 1989 version allowed for the use of an alternative test for within-cell averages, a procedure called Tiejten-Moore's test (see discussion in A2.4). The Practice E 691 approach makes use of two new parameters called" consistency statistics," designated by the symbols *h* and *k*. The general philosophy of the Practice E 691 approach will be described in this annex as well as the calculation algorithms for the *h*-value table. Calculation procedures and some additional discussion specific to the *k*-values will be given in Annex A3.

A2.3 *Defining the h Statistic*—The between-laboratory consistency statistic, *h*, is defined as follows for *each material*:

$$h = d/(S)\bar{x} \quad \text{(A2.1)}$$

where:

d = $[\bar{y}_i - \bar{Y}]$,
$\bar{y}_i$ = cell average (being tested), for any laboratory,
Y = average of all cells, and
$(S)\bar{x}$ = standard deviation of cell averages.

A2.3.1 The *h*-value is the ratio of the deviation *d*, of the cell average for any laboratory *i*, from the overall cell average of all laboratories, divided by the standard deviation among all the cell averages. The special parameter *h* may be considered as a standardized variate (or *z*-function) with a mean of zero and a standard deviation of *l*.

A2.3.2 Large *h*-values (+ or −) indicate considerable discrepancy from the overall average on the basis of a multiple of the cited standard deviation. Practice E 691 calculates an *h*-value for *each laboratory* for all materials, in distinction to the other precision standards that restrict their attention and calculation to suspiciously large within-cell standard deviations or to suspiciously small or large, within-cell averages for each material. The Practice E 691 procedure generates two additional tables that are analyzed for significantly high *h* (and *k* values; see Annex A3) indicating laboratories that are not consistent with the remainder (bulk) of the laboratories.

A2.4 *Benefits of the General Practice E 691 Outlier Approach*—The Practice E 691 technique of using *h* (and *k*) values is superior to the technique used by both Cochran's and Dixon's tests that uses the difference between the most extreme value (small or large in the case of Dixon's) and the value next in magnitude, as the basis for a test of significance for rejection of the most extreme value as an outlier. For situations where two extreme values lie close to each other and together they depart significantly from the remainder of the values, both the Cochran's and Dixon's tests fail to show the departure of the two values from the remainder of the non-suspect values. This was one of the advantages of the Tiejten and Moore test discussed above, since it looks at any number of suspicious values at the same time and avoids the masking effect of two (or more) outliers lying close to each other.

A2.5 *Decision on Significant h-values*—Practice E 691 takes an overly conservative approach on the issue of what is to be declared as a significant *h*-value (or *k*-value); it uses a 99.5 % confidence level to make this decision. This philosophy is based in part on a customary view held by statisticians, that outliers should rarely be eliminated from any interlaboratory test program (ITP). This view is based in large part on the supposition that the ITP is being done at a preliminary stage in the development of a test method and that rejecting the outliers gives a false impression of the quality or capability of the method. This view has merit for the initial phases of development for any new method and has some justification for an ITP with only a few laboratories and a few materials since it is often difficult to decide if outliers for any laboratory are indeed different from the other laboratories.

A2.5.1 For well-established test methods however, the existence of a gradation of skill and general testing competency in any large group of laboratories, argues for a modified approach to the outlier issue. For precision evaluation of established test methods with a reasonably large number of participating laboratories with several materials, there is justification to reject outliers for a particular laboratory on the basis of the more typical and universally used 95 % confidence level

rather than a 99.5 % level. This is the approach as adopted in this practice.

A2.5.2 The 95 % confidence level approach will in general, reject the results of laboratories that have poor internal testing control and are in need of improved operating procedures. Allowing these "poor" laboratories to inflate the precision results (obtained if their results are not rejected) gives a false indication of the merit or inherent quality of any test method as used by those laboratories that take the time and effort to conduct testing with proper internal control. The precision of the group of "good" laboratories (usually the majority of participating laboratories) should be the benchmark of test quality for any test method.

A2.6 *Calculating Critical h-values*—The critical value for *h*, *h* (crit), depends on the number of laboratories in the ITP and at any confidence level it may be calculated in accordance with the following equation:

$$h(\text{crit}) = (p - 1)t/\{p(t^2 + p - 2)\}^{1/2} \quad \text{(A2.2)}$$

where:

p = number of laboratories in ITP, and
t = Student's t at selected confidence level, with DF = $(p - 2)$ (a two-tailed t value).

A2.7 *Table of Critical h-values*—Table A2.1 gives calculated *h*-values, *h* (crit), at the 95 % confidence level ($p = 0.05$). These are the values as specified for the analysis of precision evaluation in accordance with this practice. If for well justified reasons another confidence level is desired for precision evaluation, it should be noted as a footnote in the precision table, the value of the alternative confidence level should be given and the reason for its adoption.

TABLE A2.1 Critical *h*-Values, *h*(crit) at 95 % Confidence Level

Number of Laboratories (*p*)	*h*(crit)
3	1.15
4	1.43
5	1.57
6	1.66
7	1.71
8	1.75
9	1.78
10	1.80
11	1.82
12	1.83
13	1.84
14	1.85
15	1.86
16	1.87
17	1.87
18	1.88
19	1.88
20	1.89
21	1.89
22	1.89
23	1.90
24	1.90
25	1.90
26	1.90
27	1.91
28	1.91
29	1.91
30	1.91
31	1.91
32	1.91

A3. PRACTICE E691 CALCULATION FOR CELL STANDARD DEVIATION OUTLIERS: *k*-VALUES

A3.1 The within-laboratory consistency statistic, designated as a *k*-value, is an indicator of how the within-laboratory variability (individual cell standard deviation, under repeatability conditions) for any selected laboratory, compares to the overall or pooled standard deviation. This comparison is done on a material (or level) by material basis. Values substantially greater than one indicate greater within-laboratory variation (for that cell) compared to the average for all laboratories.

A3.2 The usual approach to tests of significance for variability statistics, is the use of an *F*-ratio, a ratio of two variances. Therefore for the basic derivation of the *k*-value and the development of tables of significant or critical *k*-values, the variance is used rather than standard deviation.

A3.3 The *k*-value is expressed as a ratio of two standard deviations because the ratio of standard deviations is easier to comprehend in reviewing data. The units for standard deviation are the same as the units of measurement for the test.

A3.4 In the usual *F*-ratio approach, the significance of any one-cell variance to the pooled variance of all cells excluding the one cell being tested is given by Eq A3.1

$$F = (S)^2 / [(\Sigma(Si)^2 / (p - 1)] \quad \text{(A3.1)}$$

where:

$(S)^2$ = cell variance being tested for potential significance,
$\Sigma(Si)^2$ = sum of cell variances other than one being tested, and
p = number of laboratories.

The within-laboratory consistency statistic, *k*, as calculated in the Practice E 691 computer program or as it should be calculated for a spreadsheet analysis, is defined for any selected cell by Eq A3.2

$$k = (S) / (S)_r \quad \text{(A3.2)}$$

where:

(S) = cell standard deviation of the cell being tested, and
$(S)_r$ = repeatability standard deviation (for any selected material) (this is the pooled value over all laboratories).

A3.5 For purposes of calculating critical *k*-values to evaluate potential significance for any selected cell, the following development is presented. The repeatability variance is given by Eq A3.3

$$(S)^2_r = [\Sigma(Si)^2 + (S)^2] / p \quad \text{(A3.3)}$$

Combining Eq A3.1, Eq A3.2, and Eq A3.3 gives Eq A3.4:

$$k = \{[p / (1 + (p - 1) / F)]\}^{1/2} \quad \text{(A3.4)}$$

The degrees of freedom (DF) for F in Eq A3.4 are $(n - 1)$ for the numerator and $(p - 1)(n - 1)$ for the denominator.

A3.6 Eq A3.4 may be used to calculate critical k-values, k (crit), for any values of p and n, at any selected confidence level, by reference to the applicable F value at the indicated DF values. Table A3.1 gives critical k-values at the 95 % confidence level ($p = 0.05$) for various numbers of laboratories, for $n = 2$ and 3, cell replicate values.

TABLE A3.1 Critical *k*-Values, *k*(crit), at 95 % Confidence Level

Number of Laboratories, *p*	Number of Replicates, *n*		
	2	3	4
3	1.65	1.53	1.47
4	1.76	1.59	1.50
5	1.81	1.62	1.53
6	1.85	1.64	1.54
7	1.87	1.66	1.55
8	1.88	1.67	1.56
9	1.90	1.68	1.57
10	1.90	1.68	1.57
11	1.91	1.69	1.58
12	1.91	1.69	1.58
13	1.92	1.69	1.58
14	1.92	1.70	1.59
15	1.93	1.70	1.59
16	1.93	1.70	1.59
17	1.93	1.70	1.59
18	1.93	1.71	1.60
19	1.93	1.71	1.60
20	1.94	1.71	1.60
21	1.94	1.71	1.60
22	1.94	1.71	1.60
23	1.94	1.71	1.60
24	1.94	1.71	1.60
25	1.94	1.71	1.60
26	1.94	1.72	1.60
27	1.94	1.72	1.60
28	1.94	1.72	1.60
29	1.94	1.72	1.61
30	1.94	1.72	1.61
31	1.95	1.72	1.61
32	1.95	1.72	1.61

A4. ESTABLISHING A FUNCTIONAL RELATIONSHIP BETWEEN *r* (OR *R*) and *M*

A4.1 A functional relation between r (or R) and M may or may not exist. The reasoning and computational procedures presented as follows may apply to r, R, (r), and (R). They are presented for r only. Only three types of relationships will be considered:

A proportionality relation:

$$r = \nu M \quad \text{(A4.1)}$$

A linear relation:

$$r = u + \nu M \quad \text{(A4.2)}$$

A logarithmic relation:

$$\log r = c - d \log M \tag{A4.3}$$

or its equivalent:

$$r = CM^d \tag{A4.4}$$

A4.2 Eq A4.3 and also A4.4 when $d > 0$ (general case) will then lead to $r = 0$ for $M = 0$, which may seem unacceptable from an experimental point of view. Frequently, the values of M encountered in practice will have a lower limit larger than zero such that these equations can be used without introducing serious systematic errors.

A4.2.1 For $u = 0$ and $d = 1$, Eq A4.2 and Eq A4.3 will be identical to Eq A4.1, and when u lies near zero or d, or both, lies near unity. Two or all three of these equations may yield practically equivalent fits. In that case, Eq A4.1 should be preferred because it involves only one parameter and, therefore, permits a simple statement.

A4.2.2 If, in a plot of r_j against M_j, or log r_j against log M_j, the set of points is found to lie reasonably close to a straight line, a line drawn by hand may provide a satisfactory solution, but if for some reason a numerical method of fitting is preferred, the procedure of Eq A4.3 is recommended.

A4.3 The fitting of a straight line is complicated by the fact that both M and r are estimated. Since the slope, v, is usually small, of the order of 1 or less, errors in M have little influence and the errors in r predominate. The purpose is to derive values of r for given values of M; therefore, a regression of r on M is appropriate. This should be a weighted regression because the standard error or r is proportional to the value of r. With weights W_j for r_j, the computational formulas are as follows:

$$S_1 = \sum_j W_j, S_2 = \sum_j W_j M_j, S_3 = \sum_j W_j M_j^2, \tag{A4.5}$$

$$S_4 = \sum_j W_j r_j, \text{ and } S_s = \sum_j W_j M_j r_j \tag{A4.6}$$

Then, for Eq A.1

$$v = S_5 / S_3 \tag{A4.7}$$

and for Eq A4.2,

$$u = \frac{S_3 S_4 - S_2 S_5}{S_1 S_3 - S_2^2} \tag{A4.8}$$

$$v = \frac{S_1 S_5 - S_2 S_4}{S_1 S_3 - S_2^2} \tag{A4.9}$$

A4.4 The weights, W, must be proportional to r^{-2}, but the values of r_j are subject to errors; the same will hold for the weights. To correct for these and reduce the errors in the final equation, the following iterative procedure is recommended:

A4.4.1 Writing r_{oj} for the original values of r obtained by one of the calculation procedures, apply the above equations for u or v with weights:

$$W_{oj} = r_{oj}^{-2} \; (j = 1,2, \ldots q) \tag{A4.10}$$

which results in equations

$$r_{1j} = v_1 M_j \text{ or } r_1 = u_1 + v_1 M_j \tag{A4.11}$$

From these are computed adjusted values of r_j,

$$r_{1j} = v_1 M_j \text{ or } r_{1j} = u_1 + v_1 M_j \; (j = 1,2, \ldots q) \tag{A4.12}$$

and the computations are then repeated with the adjusted weights $W_{1j} = r_{1j}^{-2}$ giving

$$r_2 = v_2 M \text{ or } r_2 = u_2 + v_2 M \tag{A4.13}$$

A4.4.2 The step from W_{oj} to W_{1j} is effective in eliminating gross errors in the weights, and the equations r_2 should be considered as the final result.

A4.5 The standard error of log r is approximately proportional to $V(r)$, the coefficient of variation of r. Since the standard error of r is proportional to the value of r, the standard error of log r will be independent of r and an unweighted regression of log r on log M is appropriate when Eq A4.3 is considered.

A4.5.1 For Eq A4.3 the computational formulas are as follows:

$$S_1 = \sum_j \log M_j, \qquad S_2 = \sum_j (\log M_j)^2, \tag{A4.14}$$

$$S_3 = \sum_j \log r_j, \qquad S_4 = \sum_j (\log M_j)(\log r_j). \tag{A4.15}$$

and

$$c = \frac{S_2 S_3 - S_1 S_4}{q S_2 - S_1^2} \tag{A4.16}$$

$$d = \frac{q S_4 - S_1 S_3}{q S_2 - S_1^2} \tag{A4.17}$$

A5. PROCEDURE FOR CARBON BLACK PRECISION EVALUATION

A5.1 *Introduction*—The evaluation of precision for the test methods of Committee D-24 on Carbon Black shall be conducted in accordance with the procedure outlined in this annex. This procedure differs from the requirements as set forth in the main test of this practice. Each cell of the basic precision format table (Table 1 of this practice) contains four values as described as follows. The cell averages and cell standard deviations are used to examine outlier characteristics of the interlaboratory database by means of a protocol that differs from the basic Practice D 4483 protocol. Additionally, special calculations are made in this annex to select the mode of precision expression (absolute or relative) that is most free of influence by the level of the measured property. This special annex procedure is used so that (*1*) all carbon black test method

precision programs are conducted in the same manner, and (*2*) precision results can be compared across the tests normally employed in the carbon black manufacturing industry.

A5.2 *Terminology*—The terminology used for D-24 precision sections shall be in harmony with the terminology as used in Practice D 4483. The word 'sample' shall not be used in place of the word 'material' when discussing the number of labs, materials, days and replicates for any ITP. Samples in the context of Practice D 4483 are representative portions (or pieces) of a material scheduled for testing that are sent out to each laboratory in the ITP.

A5.3 *Materials Selected and Data Collection*—The number of materials (carbon blacks) for the precision program shall be selected based on the recommendations of Section 6. For the operations as described in this annex it is recommended that at least five materials be selected. This number of materials provides for four degrees of freedom in evaluating the significance of the coefficient of determination as described in A5.5. Tests on the selected 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. A test result is the average or median of a number of individual determinations (measurements) as specified by the method. Record all values as indicated in Table A5.1 for each material and laboratory. Each set of four values in the Table A5.1 arrangement, constitutes one cell of the final data tabulation of the entire interlaboratory test program when all the data are arranged in the basic Practice D 4483 Table 1 format. All testing shall be conducted on the same test machine or apparatus.

A5.4 *Table A5.1 Data Review and Calculations:*

A5.4.1 For each material and each laboratory calculate the average, designated as the cell average and the standard deviation, designated as the cell standard deviation, of the four values as listed in Table A5.1 format. These two statistics (cell average, cell standard deviation) are used to review the laboratories for internal testing consistency (outlier behavior) on a material-by-material basis. Although both of these statistics contain two undifferentiated components of variation, that is, between tests-between days and between tests within a day, each statistic serves as a useful index for the internal consistency (outlier) comparison.

A5.4.2 *Reviewing the Cell Averages*—Arrange the data for all laboratories and materials in the format of Table A5.2. For each material calculate *h*-values for the column of cell averages as specified by the procedures outlined in Annex A2. Also in accordance with the procedures of Annex A2, calculate the 95 % confidence level critical *h*-value, *h* (crit).

A5.4.3 *Reviewing the Cell Standard Deviations*—For each material calculate the *k*-value for the column of cell standard deviations as specified by Annex A3. Also calculate for each material the 95 % confidence level critical *k*-value, *k* (crit), in accordance with Annex A3.

TABLE A5.1 Data Format for Each Material (Each Laboratory)

	Material 1		
Date	Test Result 1	Test Result 2	Technician
Day 1	xx	xx	xxxxxxx
Day 2	xx	xx	xxxxxxx

TABLE A5.2 Format for Laboratory Precision Data

Laboratory Number	Material 1		Material 2		Material *q*	
	Cell Average	Cell Standard Deviation	Cell Average	Cell Standard Deviation	Cell Average	Cell Standard Deviation
1	xx	xx	xx	xx	xx	xx
2	xx	xx	xx	xx	xx	xx
p	xx	xx	xx	xx	xx	xx

A5.4.4 The determination of outlier laboratories is done independently for average, using the *h*- and *h* (crit) values and standard deviation, using the *k*- and *k* (crit) values. Outlier laboratories are determined by comparing the calculated *h*- or *k*-value to the *h* (crit) or *k* (crit) value, respectively. The absolute value of the calculated *h*-value is used for this comparison. Laboratories are rejected in order from highest to lowest absolute calculated *h*- or *k*-value exceeding the *h* (crit) or *k* (crit) value, respectively, for each material, until:

(a) all outliers have been rejected and the number of remaining laboratories is twenty, or greater, OR

(b) only twenty un-rejected laboratories remain, including some within the lower range of *h*- or *k*-values exceeding the *h* (crit) or *k* (crit) value, respectively.

If twenty or fewer laboratories participate in the study, reject only one laboratory for each material for average or standard deviation. If no laboratories exceed *h* (crit), retain all average data. If no laboratories exceed *k* (crit), retain all standard deviation data.

A5.4.5 After the review of data as specified in A5.4.2 to A5.4.4, the issue of blank cells (missing values) in the basic Practice D 4483 Table 1 format needs to be addressed. Refer to 7.5 and 7.6 for this as well as Section 8 for the precision calcuations.

A5.5 *Relationship Between Reproducibility Precision Parameters and M:*

A5.5.1 This section gives the necessary instructions to select the type of reproducibility precision parameter; either the absolute, *R*, expressed in measurement units or the relative (*R*) expressed in percent, that gives the most general expression of precision. General expression of precision is defined as that mode of expression that has the least dependence on the measured property level, *M*, the average material value over all laboratories.

A5.5.2 Calculate the precision parameters as specified in Section 8, on the Table A5.2 database remaining after applying the procedures of 7.5 for missing values. Plot the values of *R*, and (*R*) versus *M*. Perform a least squares regression for each of the two parameters, *R* and (*R*), and record the coefficient of determination, designated for this practice as CD, for each parameter.

A5.5.3 Select for the mode of precision expression, the parameter with the lowest CD, that is, either *R* for absolute expression or (*R*) for percent expression. This establishes which of the two modes of expression is to be used in preparing a table of precision parameters in the precision section of the

test method standard. If *R* has the lowest CD, use the absolute mode; if (*R*) has the lowest CD, use the relative mode. The mode of expression selected applies to both the reproducibility and repeatability parameters of the table of precision results.

A5.5.4 Allowing for the decision on precision parameter selection made in A5.5.3, follow the instructions as set forth in Section 9 for general guidance in preparing the final table(s) of precision results and the accompanying precision statements for the test method standard. In preparing these statements, it should be made clear whether the precision applies to individual mean levels in a table or to pooled values.

A6. SPREADSHEET CALCULATION FORMULAS FOR PRECISION PARAMETERS

A6.1 *With n = 2 replicates per cell:*

$$T_1 = \Sigma \bar{y}_i \quad \text{(A6.1)}$$

$$T_2 = \Sigma (\bar{y})^2 \quad \text{(A6.2)}$$

$$T_3 = \Sigma (W_i)^2 \quad \text{(A6.3)}$$

$$T_4 = \Sigma (S_i)^2 \quad \text{(A6.4)}$$

NOTE A6.1—Use either T_3 or T_4.

$$S_r^2 = \frac{T_3}{2p} = \frac{T_4}{p} \quad \text{(A6.5)}$$

$$S_L^2 = \frac{pT_2 - T_1^2}{p(p-1)} - \left|\frac{S_r^2}{2}\right| \quad \text{(A6.6)}$$

$$S_R^2 = S_L^2 + S_r^2 \quad \text{(A6.7)}$$

$$M = T_1 / p \quad \text{(A6.8)}$$

$$r = 2.83 \sqrt{s_r^2} \quad \text{(A6.9)}$$

$$R = 2.83 \sqrt{S_R^2} \quad \text{(A6.10)}$$

NOTE A6.2—If s_L^2 is negative, substitute $s_L^2 = 0$ in Eq A6.7.

NOTE A6.3—Symbols used:

$\bar{y}_i$ or $\bar{y}$ = average cell (test result) value,
W_i = range of cell *y* values (for *n* = 2 only),
S_i = cell standard deviation,
M = average of all *y* values (for each level), and
p = number of laboratories.

See Section 8 for other symbols used.

A6.2 *With n > 2 (a constant value over all cells)*—The computational equations are identical to A6.1 except that the value of *n* is used in place of 2 in the denominator of the second term of Eq A6.6. The value of s_r^2 is obtained by means of the T_4/p expression of Eq A6.5.

A6.3 *With unequal numbers of n replicates per cell:*

$$T_5 = \Sigma n_i \bar{y}_i \quad \text{(A6.11)}$$

$$T_6 = \Sigma n_i (\bar{y}_i)^2 \quad \text{(A6.12)}$$

$$T_7 = \Sigma n_i \quad \text{(A6.13)}$$

$$T_8 = \Sigma (n_i)^2 \quad \text{(A6.14)}$$

$$T_9 = \Sigma (n_i - 1)(S_i)^2 \quad \text{(A6.15)}$$

$$S_r^2 = \frac{T_9}{(T_7 - p)} \quad \text{(A6.16)}$$

$$S_L^2 = \left(\frac{T_6 T_7 - T_5^2}{T_7(p-1)} - S_r^2\right)\left(\frac{T_7(p-1)}{T_7^2 - T_8}\right) \quad \text{(A6.17)}$$

$$S_R^2 = S_L^2 + S_r^2 \quad \text{(A6.18)}$$

Calculate *M*, *r*, and *R* in accordance with A6.1 using:

$$M = \frac{T_5}{T_7} \quad \text{(A6.19)}$$

A7. AN EXAMPLE OF PRECISION CALCULATIONS—MOONEY VISCOSITY TESTING

A7.1 *Introduction*—The calculations illustrated in this Mooney viscosity example are performed using the spreadsheet analysis technique rather than the Practice E 691 computer analysis. This approach can better demonstrate the operations required for the various analysis steps. The data in this example, which were obtained in an interlaboratory test program (ITP) in 1982, are the same as used for the example in the previous version of this practice, that is, Practice D 4483 – 89. Although the precision calculation algorithms have not changed for this current version of Practice D 4483, the outlier rejection technique has changed, that is, it is conducted by means of the Practice E 691 *h*value and *k*-value analysis. This is in contrast to the previous Practice D 4483 – 89 technique of using the Dixon's Outlier test for cell averages and the Cochran's Maximum Variance test for cell variances (standard deviations).

A7.2 *Details on the Precision ITP*—The Mooney viscosity measurements were made in accordance with Test Methods D 1646. The ITP was conducted for seven different materials (rubbers) as illustrated in Table A7.1, which also lists the conditions of test. On each of two separate test days, one week apart, the Mooney viscosity of each of the materials was measured one time; therefore a *test result* is a single determination. In the nomenclature of a Table 1 format (see 7.2), $p = 11$, $q = 7$ and $n = 2$. The precision evaluated was a Type 1, although there were some preliminary mill-massing steps necessary for each rubber, as called for in the (1982) Section 7 specifications of Test Method D 1646, prior to viscosity measurement.

A7.3 The basic or raw data obtained in the ITP and the numerous calculations on these data are presented in a series of tables in this annex. The primary tables, starting with Table A7.2, are indicated by a table number after the annex designation, A7. Tables that are derived from a primary table, are

TABLE A7.1 Materials and Test Conditions Used

Material Number	Material Description	Test Temperature,° C	Other Details
1 (A)	SBR 1500	100	ML (1 + 4)
2 (B)	SBR 1712	100	ML (1 + 4)
3 (C)	EPDM	125	ML (1 + 4)
4[A] (D)	IIR (NIST-SRM 388j)	100	ML (1 + 8)
5 (E)	Compounded Blend 1500/1505	100	ML (1 + 4)
6 (F)	SBR Black Master Batch (1712, 65 N339, 50 HA Oil)	100	ML (1 + 4)
7 (G)	NR	100	ML (1 + 4)

[A]This IIR (butyl rubber) is a Standard Reference Material No. 388 Lot j, as furnished by the National Institute of Standards and Technology. Measurements are made on unmassed samples.

laboratories. This is given in the required Table 1 format of 7.2. At the bottom of the table are given the day averages, the 2-Day averages, the between-laboratory standard deviation of each day column and the pooled between-laboratory standard deviation over both day columns. Although these are not specified for the Table 1 format, they are easy to obtain and can be useful for data review.

A7.5 *Full Analysis—Part 1:*

A7.5.1 *Part 1: Cell Averages*—The data of Table A7.2 are used to construct Table A7.3, a table of cell averages by using the usual spreadsheet calculation operations. See Note A7.1. At the bottom of the cell average table, three parameters are

TABLE A7.2 Mooney Viscosity—Interlaboratory Test Data[A]

Laboratory Number	Level or Material													
	1		2		3		4		5		6		7	
1	46.0	47.0	51.0	51.0	68.0	67.0	69.0	69.0	68.0	69.0	76.0	76.0	99.0	101.0
2	46.8	50.4	49.2	50.0	68.4	69.6	68.3	68.3	68.9	69.6	75.8	75.2	98.0	100.0
3	46.9	46.9	48.8	49.9	68.1	67.8	70.0	70.3	69.0	69.1	72.3	74.2	100.0	99.5
4	47.0	46.0	51.0	51.0	66.0	66.0	68.0	68.5	70.0	70.0	69.0	70.0	97.5	98.0
5	45.6	46.5	50.4	49.9	65.1	65.8	68.1	68.6	68.3	67.5	72.6	73.6	98.7	99.6
6	48.5	47.0	51.0	49.5	67.0	66.0	68.0	68.0	68.5	67.0	79.0	75.5	98.0	95.0
7	46.2	46.3	50.3	50.1	68.0	68.5	68.5	68.5	68.7	68.1	76.0	77.1	100.2	100.4
8	48.2	48.9	52.4	52.3	69.0	70.0	69.5	69.0	69.2	70.2	80.4	82.3	99.0	99.1
9	46.0	46.4	50.8	50.8	69.0	69.7	69.5	69.4	68.9	69.3	71.8	72.4	98.9	99.4
10	42.0	42.5	51.0	51.0	70.0	71.0	69.0	68.5	71.0	70.5	76.0	76.0	104.0	103.0
11	46.0	45.4	48.1	48.3	70.0	66.7	69.0	68.6	68.3	67.0	63.6	61.6	93.0	91.2
Day Average[B]	46.3	46.7	50.4	50.3	68.1	68.0	68.8	68.8	69.0	68.8	73.9	74.0	98.8	98.7
2 Day Average		46.5		50.4		68.0		68.8		68.9		73.9		98.8
B-Lab Std[C]	1.70	1.97	1.22	1.04	1.53	1.85	0.68	0.63	0.86	1.26	4.73	5.13	2.59	3.18
Pooled B-Lab Std		1.84		1.13		1.70		0.65		1.08		4.93		2.90

[A]Tabulated data: Mooney units ML 1 + 4.
[B]First column each material = Day 1; Second = Day 2.
[C]B-Lab Std = Between-laboratory standard deviation.

indicated by the table number of the primary table with an appended letter designation to distinguish the derived or secondary table from the primary table. Thus tables with the same root number but with different letter designations, that is, a, b, etc. are directly related to each other.

A7.4 *Preliminary Analysis Data Review*—Table A7.2 lists the Day 1–Day 2 data for the seven materials and the eleven

calculated for each material: the material cell average (average of all cell averages); the cell average standard deviation; and cell average variance, that is, $(S)\,\bar{x}$ and $(S)\,\bar{x}^2$, given in the table by the symbols STD and VAR.

NOTE A7.1—The spreadsheet calculations were carried out with the @Avg, @Stds, @Vars, and the @Sum functions as called for in the

TABLE A7.3 Cell Averages—Mooney Viscosity

Laboratory Number	Level or Material						
	1	2	3	4	5	6	7
1	46.5	51.0	67.5	69.0	68.5	76.0	100.0
2	48.6	49.6	69.0	68.3	69.3	75.5	99.0
3	46.9	49.4	68.0	70.2	69.1	73.3	99.8
4	46.5	51.0	66.0	68.3	70.0	69.5	97.8
5	46.1	50.2	65.5	68.4	67.9	73.1	99.2
6	47.8	50.3	66.5	68.0	67.8	77.3	96.5
7	46.3	50.2	68.3	68.5	68.4	76.6	100.3
8	48.6	52.4	69.5	69.3	69.7	81.4	99.1
9	46.2	50.8	69.4	69.5	69.1	72.1	99.2
10	42.3	51.0	70.5	68.8	70.8	76.0	103.5
11	45.7	48.2	68.4	68.8	67.7	62.6	92.1
AVG (Average)	46.48	50.35	68.03	68.80	68.91	73.93	98.75
STD (Standard Deviation)[A]	1.71	1.08	1.57	0.63	0.99	4.86	2.80
VAR (Variance)[B]	2.939	1.173	2.450	0.398	0.975	23.647	7.829

[A]Standard Deviation = $(S)\,\bar{x}$.
[B]Variance = $(S)\,\bar{x}^2$.

various steps, in addition to other typical spreadsheet cell calculation procedures.[6]

A7.5.2 Using the material cell average (of each material), the cell deviation table was calculated by subtracting the material cell average from the individual cell average for each laboratory on a material-by-material basis (see Table A7.4). From the table of cell deviations, a table of *h-values* was calculated by dividing each cell deviation by the applicable material cell average standard deviation. This operation yields Table A7.5. The critical *h*-value, *h* (crit), is obtained from Table A2.1 (in Annex A2); for eleven laboratories *h* (crit) is 1.81. See Annex A2 for *h*-value analysis discussion.

A7.5.3 Reviewing Table A7.5 for observed *h*-values that exceed (crit) indicates that there are seven critical *h*-values; Laboratory 3-Material 4, Laboratory 8-Material 2, Laboratory 10-Material 1, Laboratory 10-Material 5, Laboratory 11-Material 2, Laboratory 11-Material 6, and Laboratory 11-Material 7. Laboratories 10 and 11 do not agree well with the overall ′average’ viscosity values.

A7.5.4 *Part 1: Cell Standard Deviations*—A table of standard deviations was generated by applying the appropriate standard deviation calculation function to the Day 1–Day 2 values of Table A7.2. This operation yields Table A7.6. At the bottom of the table the variance and standard deviation are given as pooled values (over the eleven cell values for each material). Table A7.7 is generated by squaring each cell value of Table A7.6 to give a table of cell standard deviations squared, that is, variances. At the bottom of Table A7.7 the pooled cell variance is given as $(Sr)^2$. The square root of this is used next to generate Table A7.8, a table of *k-values* obtained by dividing each *individual* material cell standard deviation by the *pooled* material cell standard deviation.

A7.5.5 Table A7.7 is not required to calculate a table of *k*-values as described in A7.5.4 because Table A7.6 has the necessary information to calculate the *k*-values, that is, individual cell standard deviations and the pooled ′cell standard deviation’ for each material. It is given at this point in the analysis because it will be needed (in modified format) in the process of replacing the to be rejected outlier values as described in A7.6.

[6] A Quattro Pro V4 or equivalent has been found suitable for this purpose.

A7.5.6 Reference to Annex A3, Table A3.1, for $p = 11$ and $n = 2$, yields a critical *k*-value, *k* (crit), of 1.91. A review of Table A7.8 indicates that there are five observed *k*-values that exceed *k* (crit); Laboratory 2-Material 1, Laboratory 6-Material 2, Laboratory 6-Material 6, Laboratory 6-Material 7, and Laboratory 11-Material 3. Laboratory 6 demonstrates a marked inability to repeat the viscosity measurements on the indicated Day 1–Day 2 basis.

A7.5.7 Although the next step is not strictly required for an analysis of precision, it is included in this example to illustrate the difference in the final values of the precision parameters *Sr* and *SR*, calculated (1) on the original database (no outliers rejected), and (2) on the adjusted database after all outliers are rejected and replaced with the special average values. Table A7.9, Part A, lists the values of the primary calculated variances (Sr^2 and $(S)\,\bar{x}^2$ for each material) that are required to calculate the intermediate parameter $(S^2)_L$ and the calculations that yield the final parameters *Sr* and *SR*. Refer to Section 8 for the governing equations. The original database pooled values are: $Sr = 0.82$ and $SR = 2.44$.

A7.6 *Rejection and Replacement of (Spreadsheet) Outlier Values:*

A7.6.1 *Cell Average Replacement*—The rejected cell averages have been indicated in A7.5.3. Table A7.10 is a table generated by replacing the rejected cell averages by special cell averages that preserve the *recalculated cell average* and the *recalculated cell standard deviation*. This is done on a material-by-material basis. The recalculated cell average is the average of all cells (for that material) omitting the outlier cell average value. The recalculated cell average can be easily obtained in spreadsheet calculations by erasing the outlier cell value in a table, producing a null or missing cell value (but not a zero or 0.0 value).

A7.6.2 The location of the rejected outliers are indicated in Table A7.10 by an underline. Each underlined value is a replacement value that is equal to the recalculated material cell average. Compare the recalculated values of Table A7.10 with the original database values of Table A7.3.

A7.6.3 *Cell Standard Deviation (Variance) Replacement*—The rejected cell standard deviation values have been indicated in A7.5.6. Table A7.11 has been generated by replacing the rejected cell standard deviations squared by the special cell

TABLE A7.4 Cell Average—Deviation, *d*

Laboratory Number	Level or Material						
	1	2	3	4	5	6	7
1	0.02	0.65	−0.53	0.20	−0.41	2.07	1.25
2	2.12	−0.75	0.97	−0.50	0.34	1.57	0.25
3	0.42	−1.00	−0.08	1.35	0.14	−0.68	1.00
4	0.02	0.65	−2.03	−0.55	1.09	−4.43	−1.00
5	−0.43	−0.20	−2.58	−0.45	−1.01	−0.83	0.40
6	1.27	−0.10	−1.53	−0.80	−1.16	3.32	−2.25
7	−0.23	−0.15	0.22	−0.30	−0.51	2.62	1.55
8	2.07	2.00	1.47	0.45	0.79	7.42	0.30
9	−0.28	0.45	1.32	0.65	0.19	−1.83	0.40
10	−4.23	0.65	2.47	−0.05	1.84	2.07	4.75
11	−0.78	−2.15	0.32	0.00	−1.26	−11.33	−6.65
Average (*d*)	−0.00	0.00	0.00	0.00	0.00	−0.00	0.00
Standard Deviation (*d*)	1.71	1.08	1.57	0.63	0.99	4.86	2.80

TABLE A7.5 Cell *h*–Values[A]

Laboratory Number	Level or Material						
	1	2	3	4	5	6	7
1	0.01	0.60	−0.34	0.32	−0.41	0.43	0.45
2	1.24	−0.69	0.62	−0.79	0.34	0.32	0.09
3	0.25	−0.93	−0.05	2.14[B]	0.14	−0.14	0.36
4	0.01	0.60	−1.29	−0.87	1.10	−0.91	−0.36
5	−0.25	−0.19	−1.64	−0.71	−1.02	−0.17	0.14
6	0.74	−0.09	−0.97	−1.27	−1.17	0.68	−0.80
7	−0.13	−0.14	0.14	−0.48	−0.52	0.54	0.55
8	1.21	1.85[B]	0.94	0.71	0.80	1.53	0.11
9	−0.16	0.42	0.84	1.03	0.19	−0.38	0.14
10	−2.47[B]	0.60	1.57	−0.08	1.86[B]	0.43	1.70
11	−0.46	−1.99[B]	0.20	0.00	−1.27	−2.33[B]	−2.38[B]

[A] 95 % Confidence Level *h*(crit) = 1.81.
[B] Significant *h*-Value.

TABLE A7.6 Cell Standard Deviations

Laboratory Number	Level or Material						
	1	2	3	4	5	6	7
1	0.707	0.000	0.707	0.000	0.707	0.000	1.414
2	2.546	0.566	0.849	0.000	0.495	0.424	1.414
3	0.000	0.778	0.212	0.212	0.071	1.344	0.354
4	0.707	0.000	0.000	0.354	0.000	0.707	0.354
5	0.636	0.354	0.495	0.354	0.566	0.707	0.636
6	1.061	1.061	0.707	0.000	1.061	2.475	2.121
7	0.071	0.141	0.354	0.000	0.424	0.778	0.141
8	0.495	0.071	0.707	0.354	0.707	1.344	0.071
9	0.283	0.000	0.495	0.071	0.283	0.424	0.354
10	0.354	0.000	0.707	0.354	0.354	0.000	0.707
11	0.424	0.141	2.333	0.283	0.919	1.414	1.273
Pooled Variance	0.877	0.202	0.802	0.057	0.357	1.245	1.039
Pooled Standard Deviation	0.936	0.449	0.896	0.239	0.597	1.116	1.019

TABLE A7.7 Cell Standard Deviations Squared

NOTE 1—$T4$ = Sum$(Si)^2$; [$(Si)^2$ = (Si)Squared].
NOTE 2—$(Sr)^2 = T4/p = T4/11$.

Laboratory Number	Level or Material						
	1	2	3	4	5	6	7
1	0.500	0.000	0.500	0.000	0.500	0.000	2.000
2	6.480	0.320	0.720	0.000	0.245	0.180	2.000
3	0.000	0.605	0.045	0.045	0.005	1.805	0.125
4	0.500	0.000	0.000	0.125	0.000	0.500	0.125
5	0.405	0.125	0.245	0.125	0.320	0.500	0.405
6	1.125	1.125	0.500	0.000	1.125	6.125	4.500
7	0.005	0.020	0.125	0.000	0.180	0.605	0.020
8	0.245	0.005	0.500	0.125	0.500	1.805	0.005
9	0.080	0.000	0.245	0.005	0.080	0.180	0.125
10	0.125	0.000	0.500	0.125	0.125	0.000	0.500
11	0.180	0.020	5.445	0.080	0.845	2.000	1.620
SUM(= $T4$)	9.645	2.220	8.825	0.630	3.925	13.700	11.425
$(Sr)^2$	0.877	0.202	0.802	0.057	0.357	1.245	1.038636

standard deviations squared, that preserve the pooled recalculated cell standard deviations squared, designated by Sr^2. Again the locations for rejected outliers and their replacements are indicated by the underlines. Each underlined value equals the pooled Sr^2 for that material. Compare Tables A7.11 and A7.7, from which it is generated in the spreadsheet, by the recalculation process as described above.

A7.6.4 The Table A7.10 and Table A7.11 recalculations as described in A7.6.1 to A7.6.3 provide the new values for a recalculation of *Sr* and *SR* on the adjusted (outliers removed) database, using the spreadsheet analysis technique. However, since the Practice E 691 computer program does not provide for any selected automatic rejection technique, the issue of replacing any outliers in a Practice E 691 analysis must be addressed as given in the next section.

A7.7 *Rejection and Replacement of (Practice E 691) Outlier Values:*

A7.7.1 The rejection of Practice E 691 analysis outliers is the same as for the spreadsheet technique. The tables of *h*-values and *k*-values as generated by the Practice E 691 program are reviewed with the critical values evaluated and

TABLE A7.8 Cell *k*-Values[A]

Laboratory Number	Level or Material						
	1	2	3	4	5	6	7
1	0.76	0.00	0.79	0.00	1.18	0.00	1.39
2	2.72[B]	1.26	0.95	0.00	0.83	0.38	1.39
3	0.00	1.73	0.24	0.89	0.12	1.20	0.35
4	0.76	0.00	0.00	1.48	0.00	0.63	0.35
5	0.68	0.79	0.55	1.48	0.95	0.63	0.62
6	1.13	2.36[B]	0.79	0.00	1.78	2.21[B]	2.08[B]
7	0.08	0.31	0.39	0.00	0.71	0.69	0.14
8	0.53	0.16	0.79	1.48	1.18	1.20	0.07
9	0.30	0.00	0.55	0.30	0.47	0.38	1.35
10	0.38	0.00	0.79	1.48	0.59	0.00	0.69
11	0.45	0.31	2.60[B]	1.18			

[A]95% Confidence Level *k* (crit) = 1.91.
[B]Significant *k*-Value

TABLE A7.9 Precision Parameter Calculations for Each Material

Part A—All Data Values Included:

Material	$(Sr)^2$	$(S)\bar{x}^2$	$[(Sr)^2]/2$	$(SL)^2$	$(SR)^2$	*Sr*	*SR*
1	0.877	2.939	0.438	2.500	3.377	0.94	1.84
2	0.202	1.173	0.101	1.072	1.274	0.45	1.13
3	0.802	2.450	0.401	2.049	2.851	0.90	1.69
4	0.057	0.397	0.029	0.369	0.426	0.24	0.65
5	0.357	0.975	0.178	0.797	1.153	0.60	1.07
6	1.245	23.647	0.623	23.024	24.270	1.12	4.93
7	1.039	7.829	0.519	7.309	8.348	1.02	2.89
Pooled Values	0.654				5.957	0.809	2.44

Part B—Outliers Removed:

Material	$(Sr)^2$	$(S)\bar{x}^2$	$[(Sr)^2]/2$	$(SL)^2$	$(SR)^2$	*Sr*	*SR*
1	0.317	0.973	0.158	0.815	1.131	0.56	1.06
2	0.109	0.310	0.055	0.255	0.365	0.33	0.60
3	0.338	2.450	0.169	2.281	2.619	0.58	1.62
4	0.057	0.197	0.029	0.169	0.226	0.24	0.48
5	0.357	0.604	0.178	0.426	0.783	0.60	0.88
6	0.758	9.534	0.379	9.155	9.912	0.87	3.15
7	0.692	2.964	0.346	2.618	3.310	0.83	1.82
Pooled Values	0.376				2.621	0.613	1.62
Pooled Values Excluding Material 6					1.406		1.19

TABLE A7.10 Cell Averages—Outlier Values Removed

Laboratory Number	Level or Material						
	1	2	3	4	5	6	7
1	46.5	51.0	67.5	69.0	68.5	76.0	100.0
2	48.6	49.6	69.0	68.3	69.3	75.5	99.0
3	46.9	49.4	68.0	68.7	69.1	73.3	99.8
4	46.5	51.0	66.0	68.3	70.0	69.5	97.8
5	46.1	50.2	65.5	68.4	67.9	73.1	99.2
6	47.8	50.3	66.5	68.0	67.8	77.3	96.5
7	46.3	50.2	68.3	68.5	68.4	76.6	100.3
8	48.6	50.4	69.5	69.3	69.7	81.4	99.1
9	46.2	50.8	69.4	69.5	69.1	72.1	99.2
10	46.9	51.0	70.5	68.8	68.7	76.0	103.5
11	45.7	50.4	68.4	68.8	67.7	75.1	99.4
Average	46.90	50.38	68.03	68.67	68.73	75.06	99.41
Standard Deviation[A]	0.986	0.557	1.565	0.444	0.777	3.088	1.722
Variance[B]	0.973	0.310	2.450	0.197	0.604	9.534	2.964

[A]Standard Deviation = $(S)\bar{x}$.
[B]Variance = $(S)\bar{x}^2$.

outlier values are noted and marked. Although 7.5.1 provides the criteria for replacement of the outlier values, an example of the calculation procedure will be helpful. For a Part 2 analysis (to obtain the precision parameters after outlier rejection), the Day 1–Day 2 values in a Table 1 format (Table A7.2 in this annex example) must be replaced. Thus in each cell (Table 1 format), two numbers must be inserted as a replacement for both any replacement cell average and any replacement cell standard deviation or variance.

A7.7.2 Table A7.12 is a table derived from Table A7.2 on the basis of the 7.5.1 criteria. It has both the outlier cell average and outlier cell standard deviations replaced with the special

TABLE A7.11 Cell Standard Deviations Squared—Outliers Removed

Laboratory Number	Level or Material						
	1	2	3	4	5	6	7
1	0.500	0.000	0.500	0.000	0.500	0.000	2.000
2	0.317	0.320	0.720	0.000	0.245	0.180	2.000
3	0.000	0.605	0.045	0.045	0.005	1.805	0.125
4	0.500	0.000	0.000	0.125	0.000	0.500	0.125
5	0.405	0.125	0.245	0.125	0.320	0.500	0.405
6	1.125	0.110	0.500	0.000	1.125	0.758	0.693
7	0.005	0.020	0.125	0.000	0.180	0.605	0.020
8	0.245	0.005	0.500	0.125	0.500	1.805	0.005
9	0.080	0.000	0.245	0.005	0.080	0.180	0.125
10	0.125	0.000	0.500	0.125	0.125	0.000	0.500
11	0.180	0.020	0.338	0.080	0.845	2.000	1.620
Sum(= *T*4)	3.482	1.205	3.718	0.630	3.925	8.333	7.618
$(Sr)^2$	0.317	0.110	0.338	0.057	0.357	0.758	0.693
Sr	0.563	0.331	0.581	0.239	0.597	0.870	0.832

TABLE A7.12 Mooney Viscosity: Interlaboratory Test Data—Outliers Replaced[A]

Laboratory Number	Level or Material													
	1		2		3		4		5		6		7	
1	46.0	47.0	51.0	51.0	68.0	67.0	69.0	69.0	68.0	69.0	76.0	76.0	99.0	101.0
2	48.2	49.0[B]	49.2	50.0	68.4	69.6	68.3	68.3	68.9	69.6	75.8	75.2	98.0	100.0
3	46.9	46.9	48.8	49.9	68.1	67.8	|68.5	68.8[C]	69.0	69.1	72.3	74.2	100.0	99.5
4	47.0	46.0	51.0	51.0	66.0	66.0	68.0	68.5	70.0	70.0	69.0	70.0	97.5	98.0
5	45.6	46.5	50.4	49.9	65.1	65.8	68.1	68.6	68.3	67.5	72.6	73.6	98.7	99.6
6	48.5	47.0	50.1	50.5	67.0	66.0	68.0	68.0	68.5	67.0	76.7	77.9	95.9	97.1
7	46.2	46.3	50.3	50.1	68.0	68.5	68.5	68.5	68.7	68.1	76.0	77.1	100.2	100.4
8	48.2	48.9	|50.4	50.5	69.0	70.0	69.5	69.0	69.2	70.2	80.4	82.3	99.0	99.1
9	46.0	46.4	50.8	50.8	69.0	69.7	69.5	69.4	68.9	69.3	71.8	72.4	98.9	99.4
10	|46.7	47.2	51.0	51.0	70.0	71.0	69.0	68.5	|68.3	68.8	|76.0	76.0	|104.0	103.0
11	46.0	45.4	|50.3	50.5	68.0	68.8	69.0	68.6	68.3	67.0	74.1	76.1	98.5	100.3
1 Day Average[D]	46.8	47.0	50.3	50.5	67.9	68.2	68.7	68.7	68.7	68.7	74.6	75.5	99.1	99.8
2 Day Average		46.9		50.4		68.0		68.7		68.7		75.1		99.4
B-Lab Std[E]	1.03	1.11	0.72	0.44	1.39	1.81	0.56	0.38	0.56	1.14	3.07	3.17	2.02	1.54
Pooled B-Lab Std		1.07		0.60		1.62		0.48		0.90		3.12		1.79

[A] Tabulated data—Mooney viscosity units, ML 1 + 4. Outliers replaced with either 'cell average' or mean 'cell variance' (Standard Deviation).
[B] Cell standard deviation replacement = _____.
[C] Cell average replacement = |_____.
[D] First column each material = Day 1 Test result; Second column = Day 2.
[E] B-Lab Std = Between-laboratory standard deviation.

averages. The replaced values are indicated by two types of underline as indicated in the table footnotes. The technique for doing this can be demonstrated by referring to Material 1. For this material there is one cell average replaced and one cell standard deviation replaced.

A7.7.3 *Cell Standard Deviation Replacement (n = 2)*—Laboratory 2 has an outlier cell standard deviation. Two values must be inserted in this cell, that have (1) a cell variance equal to 0.317 (see Table A7.11), and (2) an average of 48.6 (see Table A7.10). The technique to do this is reasonably straightforward for $n = 2$ (2 replicates). Two values are inserted that have the specified average (48.6) and that have a range, *w*, equivalent to a variance of 0.317 or a standard deviation of 0.563. For data pairs, the range, *w*, is related to the standard deviation of the two values, (*Si*), by Eq A7.1.

$$w = (2)^{1/2}\,(Si) \tag{A7.1}$$

In general, the data pair to be inserted in any cell, may be calculated by Eq A7.2 and Eq A7.3, with (Avg), being the average for the cell.

$$\text{Data Value 1} = (\text{Avg}) - (w/2) \tag{A7.2}$$

$$\text{Data Value 2} = (\text{Avg}) + (w/2) \tag{A7.3}$$

For this cell therefore, a standard deviation of 0.563 equals a range of 1.41 × 0.563 = 0.794 and rounding 0.794 to 0.80 the two values are; 48.6 − 0.40 = 48.2 and 48.6 + 0.40 = 49.0. This procedure is repeated on a cell-by-cell basis until all outlier cell standard deviations have been replaced.

A7.7.4 *Cell Average Replacement (n = 2)*—Laboratory 10 has an outlier cell average value (for Material 1) that must be replaced. The two replacement values must (1) be equal to the recalculated material average of 46.9 (see Table A7.10), and (2) have a range equivalent to the standard deviation of that *particular cell*, since that cell was not a cell standard deviation outlier. The cell standard deviation is 0.354 and $w = 1.41 \times 0.354 = 0.50$. Therefore the two values are 46.9 − 0.25 = 46.65 = 46.7 and 46.9 + 0.25 = 47.15 = 47.2. This procedure is repeated on a cell-by-cell basis always using the particular cell standard deviation to calculate the range used for the dual value calculation.

A7.7.5 *Cell Standard Deviation and Average Replacement (n > 2)*

If there are more than two replicates per cell, outliers may be replaced in the outlier cells with two inserted cell values by the same technique as described in A7.7.3 and A7.7.4. This replacement however unbalances the Table 1 format database,

producing unequal replicates among the cells. The analysis of this type of database may be conducted by way of the equations of A6.3 in Annex A6.

A7.7.6 *Comparing the Outlier Adjusted Databases: Practice E 691 versus Cochran Test* —The previous version of Practice D 4483 (1989) made use of Cochran's max variance test to eliminate cell standard deviation (or variance) outliers. Reference to the 1989 version Table A8.3 , shows that only two cells had significant outliers at the 95 % confidence level; Laboratory 2-Material 1 and Laboratory 11-Material 3. The Practice E 691 *k*-value analysis at the same confidence level eliminated five cells; Laboratory 2-Material 1, Laboratory 6-Material 2, Laboratory 6-Material 6, Laboratory 6-Material 7, and Laboratory 11-Material 3. The Cochran analysis completely missed the poor performance of Laboratory 6.

A7.7.7 *Comparing the Outlier Adjusted Database: Practice E 691 versus Dixon's Test* —The previous version of Practice D 4483 made use of Dixon's Test for cell average outlier analysis. Reference to Table A8.4 of the 1989 version of Practice D 4483, shows that only two cell averages were rejected at the 95 % confidence level; Laboratory 10-Material 1 and Laboratory 11-Material 7. The Practice E 691 *h*-value analysis eliminated seven cell averages at the same confidence level; Laboratory 10-Material 1, Laboratory 8-Material 2, Laboratory 11-Material 2, Laboratory 3-Material 4, Laboratory 10-Material 5, Laboratory 11-Material 6, and Laboratory 11-Material 7. The poor performance of Laboratory 11 was missed by the Dixon's Test as well as the very marginal performance of Laboratory 10.

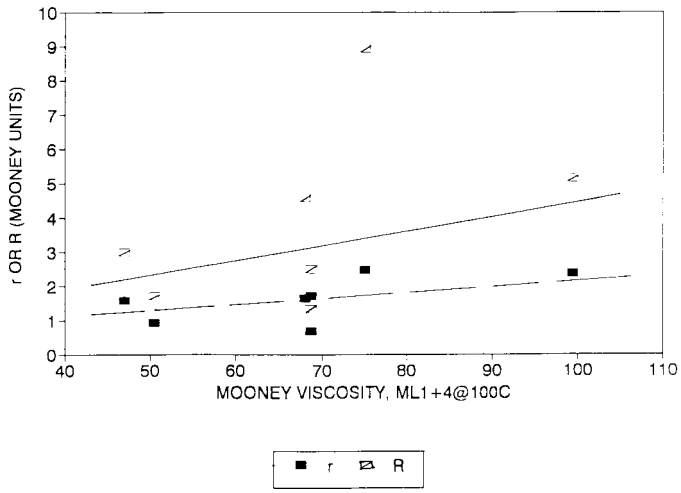
A7.8 *Full Analysis—Part 2:*

A7.8.1 Using the Table A7.12 adjusted database (outliers replaced) and the Practice E 691 computer program, the Part 2 analysis may be conducted. For the spreadsheet precision analysis, Table A7.10 and Table A7.11 are used to perform the calculations as indicated in Part B of Table A7.9.

A7.8.2 The results of the precision calculations are given in the standard Practice D 4483 format in Table A7.13. One material stands out with very poor between-laboratory precision, Material 6-SBR (BMB). This is a carbon black filled black masterbatch material. Testing programs conducted subsequent to the date of this ITP have shown that one important source of the poor between-laboratory precision is the viscosity variation introduced by the mill-massing operation that was part of the preliminary treatment of all the Mooney test specimens. (The black masterbatch material is sensitive to this mill-massing). The other rubbers of this ITP are clear rubbers and are not as sensitive to this operation. Specimen preparation options have been recently introduced into Test Method D 1646 to avoid some of these problems. At the bottom of Table A7.13, between-laboratory pooled values have been calculated that omit Material 6; these pooled values are more representative of clear rubbers.

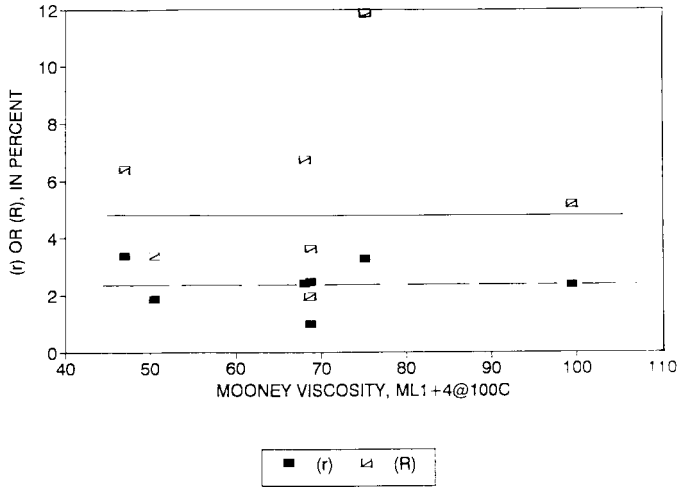
A7.8.3 All of the values in precision Table A7.13 are representative of some average or typical laboratory operation. As a rough approximation, three grade levels of testing skill and degree of internal test control contribute to the *collective results* of the table—Good, Intermediate, and Poor. Although some of the poor results have been removed from the database by the Practice E 691 *h* and *k* analysis, certain marginal data are still part of the adjusted database.

A7.8.4 Fig. A7.1 illustrates plots of *r* and *R* versus average Mooney viscosity and Fig. A7.2 is a similar plot of (*r*) and (*R*) versus Mooney viscosity. Visually fitted regres-sion lines have been drawn as indicated ignoring the point for Material 6 for the *R* and (*R*) lines. There is a very mild dependence of *r* and *R* on viscosity with however substantial scatter for the *R* points. The relative (percent) expression of precision, (*r*) and (*R*), shows no dependence on viscosity.



NOTE 1—Dashed line for *r*; solid line for *R*.

FIG. A7.1 Repeatability, *r* , and Reproducibility, *R* , Versus Mooney Viscosity



NOTE 1—Dashed Line for (*r*); Solid Line for (*R*).

FIG. A7.2 Relative Repeatability (*r*), and Reproducibility (*R*) Versus Mooney Viscosity

TABLE A7.13 Precision Parameters for Test Method D 1646—Mooney Viscosity (Type 1 Precision)[A]

NOTE 1—
Sr = Repeatability, standard deviation,
r = Repeatability (measurement units),
(r) = Repeatability (relative basis, percent),
SR = Reproducibility, standard deviation,
R = Reproducibility (measurement units), and
(R) = Reproducibility (relative basis, percent).

Material	Mean Level[B]	Within -Laboratories			Between-Laboratories		
		Sr	*r*	(*r*)	*SR*	*R*	(*R*)
1. SBR1500	46.9	0.56	1.58	3.38	1.06	3.00	6.40
2. SBR1712	50.4	0.33	0.93	1.85	0.60	1.70	3.37
3. EPDM	68.0	0.58	1.64	2.41	1.62	4.58	6.74
4. BUTYL (IIR388)	68.7	0.24	0.68	0.99	0.47	1.33	1.94
5. SBR BLEND	68.7	0.60	1.70	2.47	0.88	2.49	3.63
6. SBR (BMB)	75.1	0.87	2.46	3.28	3.15	8.91	11.87
7. NR	99.4	0.83	2.35	2.36	1.82	5.15	5.18
Average	68.2						
Pooled Values[C]		0.61	1.73	2.54	1.62	4.58	6.72
Pooled Values[C] Excluding Material 6					1.18	3.35	4.91

[A]Short-term, days, with *p* = 11, q = 7, and n = 2, and outliers in database removed.
[B]In Mooney torque units.
[C]Option 2 for (*r*), (*R*).