Designation: D4483 – 14a

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

This standard is issued under the fixed designation D4483; 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.

INTRODUCTION

The primary precision standard for ASTM test method standards is Practice E691; a generic standard that presents the fundamental statistical approach and calculation algorithms for evaluating repeatability and reproducibility precision. However, certain parts of Practice E691 are not compatible with precision as evaluated in the rubber manufacturing and carbon black industries over the past four decades. Thus a separate standard is required for precision in these two industries. This practice is being issued as a major revision of Practice D4483, which has been used for precision evaluation by Committee D11 since 1985. The basic Practice D4483 precision calculation algorithms, the same as in Practice E691, are unchanged. This new revised Practice D4483, organized to accommodate the requirements of the rubber and carbon black manufacturing industries, has three new features that provide for a more formal and structured analysis of interlaboratory test program (ITP) data.

First it addresses the overriding issues with precision evaluation over the past several decades—the frequent discovery that reproducibility for many test methods is quite poor. Experience has shown that frequently poor reproducibility is caused by only a few laboratories that differ from the remainder that give good agreement. A new procedure designated as *robust analysis* provides an improved method for detecting outliers that cause poor precision, especially poor between laboratory agreement. Second, after outlier detection the new standard provides two options; (*1*) outlier deletion or (*2*) outlier replacement. When outliers are deleted the revised standard provides a way to retain the non-outlier laboratory data. This allows for a broader database for precision calculation. The current ASTM Committee E11 computer program for calculating precision does not allow for outlier deletion in this way. Third, when exercising outlier Option 2, the standard gives a procedure for calculating special replacement values for deleted outliers in ITPs that have only a few participating laboratories. The replacement values are obtained in a way that preserves the observed data distribution of the non-outlier data. This is important since many ITPs are in the *limited number of participating laboratories* category.

1. Scope

1.1 This practice covers guidelines for evaluating precision and serves as the governing practice for interlaboratory test programs (ITP) used to evaluate precision for test methods as used in the rubber manufacturing and the carbon black industries. This practice uses the basic one way analysis of variance calculation algorithms of Practice E691. Although bias is not evaluated in this practice, it is an essential concept in understanding precision evaluation.

1.2 This practice applies to test methods that have test results expressed in terms of a quantitative continuous variable. Although exceptions may occur, it is in general limited to test methods that are fully developed and in routine use in a number of laboratories.

1.3 Two precision evaluation methods are given that are described as *robust statistical* procedures that attempt to eliminate or substantially decrease the influence of outliers. The first is a *General Precision* procedure intended for all test methods in the rubber manufacturing industry, and the second is a specific variation of the general precision procedure designated as *Special Precision*, that applies to carbon black testing. Both of these procedures use the same uniform level experimental design and the Mandel *h* and *k* statistics to review the precision database for potential outliers. However, they use

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

Current edition approved May 1, 2014. Published May 2014. Originally approved in 1985. Last previous edition approved in 2014 as D4483 – 14. DOI: 10.1520/D4483-14A.

slight modifications in the procedure for rejecting incompatible data values as outliers. The *Special Precision* procedure is specific as to the number of replicates per database cell or material-laboratory combination.

1.4 This practice is divided into the following sections:

1.5 Six annexes are presented; these serve as supplements to the main body of this practice. Annex A1 and Annex A2 are given mainly as background information that is important for a full understanding of precision evaluation. Annex A3 – Annex A5 contain detailed instructions and procedures needed to perform the operations as called for in various parts of the practice. The use of these annexes in this capacity avoids long sections of involved instruction in the main body of this practice. This allows for a better presentation and understanding of the central concepts involved in the evaluation of precision. Annex A6 is also important; it gives a complete example of precision evaluation that illustrates all of the procedures and options likely to be encountered in any precision evaluation, from the simple to the most complex.

1.6 *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:*[2]

D1646 Test Methods for Rubber—Viscosity, Stress Relaxation, and Pre-Vulcanization Characteristics (Mooney Viscometer)

D6600 Practice for Evaluating Test Sensitivity for Rubber Test Methods

E691 Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method

2.2 *ISO Standard:*[3]

ISO 289 Determination of Viscosity of Natural and Synthetic Rubbers by the Shearing Disk Viscometer

3. Terminology

3.1 A number of specialized terms or definitions are defined in a systematic sequential order, from simple terms to complex terms. This approach allows the simple terms to be used in the definition of the more complex terms; it generates unambiguous definitions. Thus the definitions do not appear in the usual alphabetical sequence.

3.1.1 This terminology section contains explanatory notes for many of the definitions as well as discussion on the connection between some of the terms and the various ways the terms are used in testing and precision evaluation. For special emphasis, a few terms are defined in the main text of this practice where certain precision concepts are discussed.

3.1.2 Annex A1 is included as part of this practice with two objectives: (*1*) Annex A1 presents new more comprehensive definitions drafted with substantial tutorial content, and (*2*) Annex A1 presents some ancillary definitions that may promote a better understanding of precision.

3.2 *Testing Terms:*

3.2.1 *balanced uniform level design, n*—the plan for an interlaboratory test program for precision, where all laboratories test all the materials selected for the program and each laboratory conducts the same number of repeated tests, on each material.

3.2.2 *element, n*—the entity that is tested or observed, to evaluate a property or characteristic; it may be a single object among a group of objects (test pieces, and so forth) or an increment or portion of a mass (or volume) of a material.

3.2.2.1 *Discussion*—The generic term *element* has a number of synonyms: test piece, test specimen, portion, aliquot part, subsample, and laboratory sample.

3.2.3 *element class (or class of elements), n*—the category or descriptive name for a group of elements that have a common origin or have nominally identical properties.

3.2.3.1 *Discussion*—The term *nominally identical* implies that the elements come from a source that is as homogeneous as possible with regard to the property being measured.

[2] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

[3] Available from International Organization for Standardization (ISO), 1, ch. de la Voie-Creuse, Case postale 56, CH-1211, Geneva 20, Switzerland, http://www.iso.ch.

3.2.4 *test result, n*—the value of a characteristic obtained by carrying out a specified test method.

3.2.4.1 *Discussion*—The test method should specify that one or a number of individual measurements, determinations, or observations be made and their average or another appropriate function (median or other) be reported as the test result.

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

3.2.5.1 *global testing domain, n*—a domain that encompasses two or more locations or laboratories, domestic or international, typically used for producer-user testing, product acceptance, and interlaboratory test programs.

3.2.5.2 *local testing domain, n*—a domain comprised of one location or laboratory as typically used for quality control and internal development or evaluation programs.

3.3 *Material and Sampling Terms:*

3.3.1 *independent tests, n*—a set of measurements (or observations) for a defined testing domain, where, in relation to the measurement process, there is no influence of any selected measurement on any other measurement in the set.

3.3.1.1 *Discussion*—The word *independent* is used throughout this practice as an adjective to indicate the concept of independence, for samples, test pieces, and so forth, as well as tests.

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

3.3.2.1 *Discussion*—A lot may be generated by a common production (or other natural) process in a restricted time period and usually consists of a finite size or number. A lot may be a fractional part of a population (Interpretation 2 of population, see Annex A1). A recognized property range implies that some rough approximation is available.

3.3.3 *material, n*—a specific entity or element class to be tested; it usually exists in bulk form (solid, powder, or liquid).

3.3.3.1 *Discussion*—Material is used as a generic term to describe the *class of elements* that is tested, that is, a material may be a rubber, a rubber compound, a carbon black, a rubber chemical, and so forth. A material may or may not be homogeneous. In product testing the term material may be used to describe the *class of elements* or type of rubber products such as O-rings, hose assemblies, motor mounts, and so forth. See also 5.1.4.1.

3.3.4 *sample (data), n*—the number of test or observation values (n = 1, 2, 3, and so forth), obtained from (one or more) physical samples, by the application of a specific test (observation) method.

3.3.5 *sample (physical), n*—the number of elements or the specified mass of a material, selected according to a particular procedure, used to evaluate material, lot, or population characteristics.

3.3.5.1 *Discussion*—The term *sample* should not be used as a synonym for *material*, see 3.3.3, or *target material*, see 5.1.4.1. Ideally several *materials* are tested in any ITP with each material being different (chemically, structurally, property wise). From each material, some number of *samples* (all nominally identical) may be taken for testing. See 3.3.4.

3.3.6 *test sample, n*—that part of a (physical) sample of any type taken for chemical or other analytical testing, usually with a prescribed blending or other protocol.

3.3.6.1 *Discussion*—A test sample is usually a mass or volume that is some small fractional part of a bulk material.

3.3.7 *test specimen, n*—an object (appropriately shaped and prepared) taken from a sample for physical or mechanical testing.

3.3.7.1 *Discussion*—Other terms for test specimen are: test portion, test item, and test piece (used in ISO standards).

3.4 *Statistical Terms Relating to Precision:*

3.4.1 *estimated (true or reference) mean, n*—the mean obtained on the basis of n independent replicate measurements; the greater n the better the approximation to the true or reference mean, provided there is no systematic deviation or bias.

3.4.1.1 *Discussion*—The words *mean* and *estimated mean* are frequent synonyms for *estimated (true or reference) mean*. The value for n in typical routine testing programs is of the order 1 to 10. When bias exists, the estimated (true or reference) mean so obtained estimates $[\mu + \Sigma\ Bi]$, where μ = true or reference mean and $\Sigma\ Bi$ = algebraic sum of all bias deviation terms. Therefore, if bias exists and is unknown in magnitude, the true value or μ cannot be approximated despite increased replication. See random and bias deviations in A1.2.5 and A1.2.6. See also Annex A2.

3.4.2 *outlier, n*—a member of a set of values which is inconsistent with the other members of that set.

3.4.3 *reference value, n*—a value (usually a mean) generated by a recognized and accepted procedure that is used as a true value.

3.4.3.1 *Discussion*—Reference values are used when it is impossible or exceedingly difficult to obtain a true value. Such values are most often assigned on the basis of comprehensive testing programs sanctioned by a local or global task group, a standardization organization, or a committee devoted to domestic or international metrology.

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

3.4.4.1 *Discussion*—The word *replicate* refers to a physical object (element). It can also be used in reference to a data set, where it refers to one of a number of independent data values.

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

3.4.5.1 *Discussion*—The true value is also defined as the mean that would be obtained by testing all members of any population (see population in Annex A1). Typical *systems-of-causes* are the unavoidable fluctuations in temperature,

humidity, operator technique, fidelity of calibration, and so forth, in a controlled testing domain.

3.5 *Definitions:*In some of the following definitions, the term *figure of merit* is used. A high figure of merit is an indication of high quality or a high level of excellence or goodness for the measurement or test domain, or both. The term *figure of merit* applies to a number of test method characteristics: precision, sensitivity, bias, useful range, ruggedness and ease of operation, and rapid or automated operation.

3.5.1 *precision, n*—a *figure of merit* concept, it is proportional to the inverse of the dispersion of independent replicate (test or observed) values, as estimated by the standard deviation, for a specified class of elements and a defined testing domain.

3.5.1.1 *Discussion*—The merit of a test method depends on the precision, high merit equals high precision. However, it has become customary practice to express precision in terms of the dispersion of replicate values, that is, by the standard deviation. However, this is actually a measure of imprecision; therefore, the use of the inverse of the standard deviation in this definition. Precision may be influenced by both random and bias deviations depending on the defined testing domain. There are other *figure of merit* testing concepts. An additional one is test sensitivity; the ratio of the magnitude of the measurement response for a selected property difference to the precision or accuracy of the measurement, or both. See Practice D6600 for more details on test sensitivity.

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

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

3.5.4 *repeatability, r, n*—the precision for a defined *local testing domain*, obtained by way of *n* independent replicate tests (on nominally identical elements) expressed in terms of an interval or range that is a multiple of the standard deviation; this interval should (on basis of a 95 % probability) encompass duplicate independent test results obtained under the defined local testing domain.

3.5.4.1 *Discussion*—The *local testing domain* is defined as one laboratory, usually one instrument, one test technician with a *specified* replicate test time period. The words *nominally identical* imply elements drawn from a homogenous source with all reasonable effort taken to eliminate production variation within the source. Repeatability may be dependent on the magnitude or level of the measured property and is usually reported for particular property levels or materials or element classes (that determine the level). The repeatability time period may be minutes, hours, or days depending on the goals and scope of the testing.

3.5.4.2 *Discussion*—Although repeatability as defined in 3.5.4 applies to a local testing domain, it can be obtained in two different ways and can be used in two different contexts. It can pertain to a common community value, obtained as an average (or pooled) value from all laboratories in an ITP among *N* different laboratories. This is a *global* repeatability, that applies to a *typical laboratory*, that stands as a representative of all laboratories that are part of a global testing domain. It can also pertain to the long-term or established value for a *particular laboratory* as derived from ongoing testing in that laboratory, not related to any ITP. The second use can be referred to as a local repeatability, that is, repeatability obtained in and for one laboratory.

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

3.5.5.1 *Discussion*—Each laboratory in the global domain conducts *n* repeatability tests on a material (target material), and reproducibility is evaluated based on the mean values for the *N* laboratories for that material or element class. Reproducibility may also depend on the level of the measured property or on the materials tested and it is also usually reported for particular levels or materials. Reproducibility usually does not have the dual interpretation or use as previously discussed for repeatability, since it is a *group characteristic* that only applies across a number of laboratories in a global testing domain.

3.5.5.2 *Discussion*—It is appropriate to also express precision on a relative basis, as a percent of a certain mean value. This is analogous to a coefficient of variation. A relative expression may be important when the precision varies with the level of the property being measured. Frequently the relative precision is reasonably constant when so expressed. To avoid any confusion with measured properties that are expressed in percentages, for example, % copper, % elongation, and so forth, relative precision is expressed using parentheses that enclose the symbols for repeatability and reproducibility.

3.6 Additional terms concerning certain types of precision will be defined in 5.1. Better understanding can be gained by giving these definitions, which relate to the nature of the material to be tested, in that section.

4. Significance and Use

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

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

4.3 Experience as indicated in Refs (**1-4**)[4] and elsewhere has shown that the poor reproducibility among the laboratories of a typical ITP is almost always due to interlaboratory bias. Certain laboratories are always low or high compared to a reference as well as other laboratories in all tests. This usual outcome for many ITPs is addressed in this practice by the use of the three-step robust analysis procedures as described in Section 7.

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

5. Precision Evaluation: General Precision and Special Precision

5.1 *General Precision*—Two precision categories are described: General Precision and Special Precision. General Precision is discussed first and Special Precision is described in Section 11. General Precision evaluation follows established procedures used in the rubber manufacturing industry over the past four decades. The evaluation is usually conducted using a balanced uniform level design ITP with three or more materials sent to each of the participating laboratories with tests conducted to generate an independent *test result*, on each of two (or more) test days. The ITP database is reviewed for outliers by the Mandel *h* and *k* consistency statistics by the procedures in Annex A3.

5.1.1 *Options for Outliers*—If no outliers are found, the original database is used to develop a table of precision results. If outliers are identified, there are two options for outlier treatment; Option 1, outlier deletion, is the first choice. Option 2, outlier replacement, is chosen for an ITP with a minimum (approximately six) number of laboratories. Issues such as the number of replicate values on each test day or the number of technicians or operators used to obtain a test result, or both, which are characteristic of the particular test, are considered on a case-by-case basis by the ITP organizing committee. Outlier treatment is discussed in more detail in Annex A3 and Annex A5.

5.1.2 *Types of Test Methods*—The General Precision approach has been successfully used for the broad range of test methods characteristic of the rubber manufacturing industry; from simple physical or chemical *bench type* tests, conducted in a few minutes (hardness and pH tests) to a complex multistep test method, such as an aging test. Such a test requires preliminary property measurement, a substantial aging period (days) followed by aged property measurement to obtain a final calculated test result or performance index. For such complex tests, any realistic precision evaluation must of necessity include all of the procedural steps in arriving at the test result, the basic datum used in precision analysis, and evaluation. The procedures required for general precision are described in Sections 8 – 10.

5.1.3 *Types of General Precision*—In addition to the General Precision aging tests as previously cited, other tests also require a more complex total sequence of operations to generate a final test result. One important test of this type is a *performance-in-rubber* test; the evaluation of various rubbers, reinforcement fillers, or other compounding materials in standardized formulations. The typical stress-strain evaluation of a selected lot of a specified rubber will require (*1*) an appropriate sample of the rubber, (*2*) a standardized formulation and mixing operation to prepare a compound using standard compounding materials, (*3*) processing of this compound to prepare cured or vulcanized molded sheets at a selected time and temperature, (*4*) cutting and gaging of dumbbell (or other) test pieces, and (*5*) the testing of the lot to obtain the final test results for tensile stress (modulus), elongation, and tensile strength properties.

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

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

5.1.5 Using the term *target material*, two types of precision may be defined:

5.1.5.1 *Type 1 Precision*—A precision evaluated directly for or on, a target material; fully prepared test pieces or test portions of the target material drawn from a homogeneous source are tested, with no processing or other operations required prior to testing.

5.1.5.1.1 *Discussion*—An example is a lot comprised of died-out, gaged dumbbells for stress-strain testing.

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

[4] The boldface numbers in parentheses refer to the list of references at the end of this standard.

5.1.6 The properties of the composite material are directly related to the quality or properties of the target material. An example: To evaluate the quality of a grade of SBR, a sample of the rubber, plus curatives, filler, antioxidants, and so forth, are mixed, cured, test pieces prepared, and the resulting compound tested for specified quality properties. It is possible that a Type 1 precision program might be conducted on test pieces or portions that require some minimum processing or other simple operations prior to actual testing. This is, in a strict sense, an intermediate level of precision. However, to avoid unnecessary complications, this will be designated as a Type 1 precision.

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

6. Steps in Organizing an Interlaboratory Test Program

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

6.1.1 *Organization Committee*—An organization committee or task group and a program coordinator should be selected. One member of the committee or group should be a statistician familiar with the testing technology of the test method as well as the content of this practice. Most ITPs are organized on the basis of a balanced uniform level design for the precision program.

6.1.2 *Category and Type of Precision*—For all programs except for carbon black testing, a General Precision ITP is organized. For carbon black testing a Special Precision ITP is organized. The type of precision to be evaluated shall be selected, see 5.1.5. Type 1 precision is the most frequently evaluated. For some test methods such as rubber or polymer or other performance-in-rubber evaluations using standard formulations, a Type 2 precision is required.

6.1.3 *Test Operator or Technician Selection*—For simple General Precision testing requiring only one operator or technician, all replicate tests should be conducted by the same technician unless the effect of different technicians is part of any program. For more complex tests where several operators or technicians are required to perform a sequence of different steps to arrive at a test result, the same *operator team* should conduct testing for all replicates again unless the effect of different operator teams is part of the program.

6.1.3.1 For Special Precision testing follow the procedure of using two technicians on each of two test days. See Section 11.

6.1.4 *Test Result and Number of Replicates*—Each test method has a final value for the property under evaluation, defined as a test result. A test result may be a mean or median value of a number of individual determinations as specified by the test method. For the purposes of this practice, a replicate is defined as a test result. The number of replicate test results, *n*, within each laboratory on any material should be specified. In most ITPs this is two. For some tests, three or four replicates, as in Special Precision, may be selected. All analysis is conducted on test results.

6.1.5 *Time Period for Repeatability*—The time period between replicate tests within any laboratory should be selected. This time period is usually one of days, in the range from 1 to 7 days. For special tests (long aging periods) replicate tests may require a longer time span. For other special testing operations, shorter time periods (minutes, hours) may be selected. The primary consideration is how the test method is typically used in the industry. The selected time period shall be reported in the precision section of the test method.

6.1.6 *Number of Target Materials*—The number of target materials or classes of objects (or manufactured products) to be tested should be selected. Ideally, this should be three or four with substantially different property levels. The target materials should represent typical industry materials as normally used and subjected to test. See 5.1.

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

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

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

6.1.10 *Testing Instructions*—Although all ITPs are usually conducted for a standard test method that includes the complete set of instructions for the test, some supplemental instructions are required. One important supplemental instruction is the schedule for the testing. All tests should be performed on specified days, and all participating laboratories should conduct the test as specified by the test method. The schedule should allow for adequate material delivery time. Any special modifications of the test method should be clearly described as well as special instructions as to operators or technicians (one,

two, or more) versus replicate testing. If an ITP is to be conducted for a test method at some intermediate development level, it is essential to give all participating laboratories instructions for conducting the test method as well as all the required ITP instructions.

6.1.11 *ITP Test Data Report*—A *test report data form* should be prepared by the ITP coordinator and a copy sent to each participating laboratory along with the test materials and instructions. This form should contain locations to report the following: the name of the laboratory; the test dates as actually used; and for each target material tested, the test value (test result) for each replicate test (day), reported if possible to one more significant figure than is normally used (that is, do not truncate). The test report form should also ask for a description of the test equipment or machines used (model number, condition), comments about any unintended deviations from the standard test procedure and disclosure of any mishaps or other pertinent information. The completed test report should be returned to the ITP coordinator.

7. Overview of General Precision Analysis Procedure

7.1 *Analysis Operation Sequence*—This section gives a quick overview of the procedures required for the analysis of the ITP database and provides the user with a better appreciation of the complete analysis process. Some background on outliers is also presented in this section for a better appreciation of this topic. The General Precision procedure may require as many as three analysis operations or overall steps. The actual number will be determined by the uniformity of the data in the database. If there are no outliers, only Analysis Step 1 is used. If outliers are present, Analysis Steps 2 and 3 may be required depending on the extent of outliers in the database. Annex A4 contains instructions for all three analysis operations and also gives the details on how to layout the required tables and their interlinking that enables the automatic recalculation of the final precision parameters, *r* and *R*, when outliers are deleted or replacement values are substituted into the basic data Table 1 format. Fig. 1 is a decision tree or flow chart diagram that outlines the steps in the complete analysis process.

7.1.1 *Preliminary Data Review*—A quick numerical review of any database is important to gain a first impression of the results of any ITP. This preliminary data review is conducted after cell averages and cell standard deviations (or cell ranges) have been calculated. Part of this review is the generation of special plots of cell averages and cell standard deviations or cell ranges versus laboratory number. These plots, as described in 8.1.3, will clearly show potential outlier values.

7.1.2 *Analysis Step 1*—The original database is analyzed to generate values for repeatability and reproducibility for each material (or target material) and the h and k statistics calculated. See Annex A3. Annex A4 gives the instructions for generating six tables that yield values for the *h* and *k* statistics and the precision results for each material. The calculated *h* and *k* values are compared to the 5 % significance level critical *h* and *k* values to determine if there are any significant outlier values. If there are none, the analysis is complete and the values found for repeatability and reproducibility are used to generate a table of precision results for the test method. If there are any significant outliers, Analysis Step 2 is required.

7.1.3 *Analysis Step 2*—If there are any outliers at the 5 % significance level, the outlying values are either (*1*) deleted using Option 1 or (*2*) replaced using Option 2. See Annex A3, Annex A5, and 5.1.1. On the basis of either option, the resulting revised database, designated as Revision 1 or *R1*, is analyzed to generate new values for repeatability and reproducibility, designated as *R1* precision values. This analysis produces a new set of calculated *h* and *k* values that are compared to 2 % significance level critical *h* and *k* values to determine if there any significant outlier values at this level. If there are none, the analysis is complete and the values found for repeatability and reproducibility are used to generate a table of *R1* precision results for the test method. If there are any significant outliers, Analysis Step 3 is required.

7.1.4 *Analysis Step 3*—If any of the R1 calculated *h* and *k* values exceed the 2 % significance level critical *h* and *k* values, the outlying values are either (*1*) deleted using Option 1 or (*2*) replaced using Option 2. On the basis of either option, the resulting *R2* database is analyzed to generate new values for repeatability and reproducibility, designated as *R2* precision values. This completes the analysis sequence, and the values found for repeatability and reproducibility for each material are used to prepare a table of precision results for the test method.

NOTE 1—Although complete analysis algorithms using spreadsheet procedures are given in this practice, a special computer program has been developed by ASTM Committee E11 to calculate repeatability and reproducibility equivalent to this practice, and the software for this is available from ASTM. See Ref (**5**). However, the ASTM program is not able to accommodate databases that have blank cells. See 8.1 and Annex A4 for more details on calculation procedures.

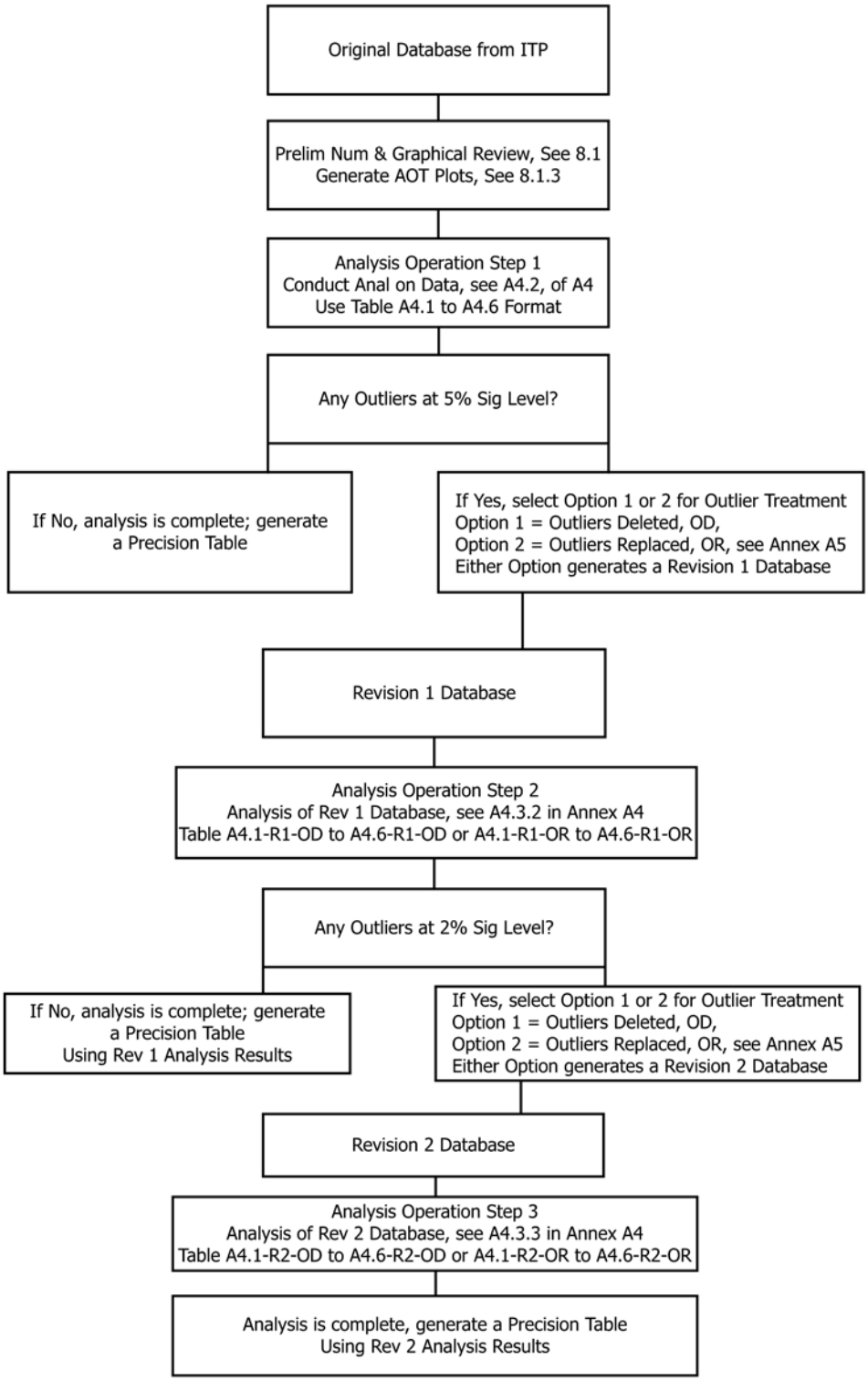
7.1.5 The General Precision part of this practice does not address the issue of attempting to fit a relationship; *r*, *R*, (*r*) or (*R*) versus the property (level) for any ITP for two reasons. First, most ITPs do not have a sufficient number of materials to produce any meaningful functionality of precision versus material level; the degrees of freedom for any obtained fit are small. Second, experience has shown that even when there are several materials in an ITP, a good fitting linear or other relationship is not obtained. It should be remembered that any ITP is *an event in time* that gives an indication of the general level of precision for three or four materials in a selected

TABLE 1 Precision Program—Basic Data[A]

Material (*j*) ==> Laboratory (*i*)	1	2	3	4	...	*q*
1						
2			*Yijk*			
3						
4						
5						
...						
p						

[A] Table layout for uniform level ITP.

Notation used:
Laboratories, a total of *p*, *L*(*i*) = 1, 2, 3, ... *p*
Materials or Levels, a total of *q*, *m*(*j*) = 1, 2, 3, ..., *q*
Replicates, a total of *n* per cell; a cell = each combination of *L*(*i*) *m*(*j*); normally *n* = 2
Yijk = a single test result value; where *k* = 1, 2, ... *n*(*ij*); see cell (23) of table for example
Cells (*i, j*); each cell contains *n* test result values



NOTE 1—Refer to Example Precision Calculations in Annex A6 for tables with data.

FIG. 1 Decision Tree Diagram for ITP Data Analysis

number of laboratories. With some occasional exceptions, the precision found is usually quite different for each material with no detectable pattern or functionality.

7.2 Annex A2 gives a statistical model that demonstrates the influence of both random and bias components of variation inherent in any precision evaluation. Section A2.5 gives the derivation of the expressions for repeatability and reproducibility in terms of the between laboratory and within laboratory variance and illustrates how both of these are related to random and bias components of variance.

7.2.1 The term S_L^2 is used in the calculation of the reproducibility variance and standard deviation in accordance with A2.5.1. Experience has shown, however, that the within-laboratory variation is substantially smaller than between-laboratory variation. In certain circumstances S_L^2 may calculate to less than zero; if this occurs, S_L^2 is set equal to zero. This less than zero situation may occur when there is substantial within cell variation of such a nature that when laboratory cell averages are calculated, they agree quite well. The analyst is cautioned to check the final (that is, after all outlier treatments, if any, have been completed) calculations to see if there are any less than zero values for S_L^2 and replace the less than zero value(s) with zero before calculating and reporting the final precision values.

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

7.4 *Outlier Appearance Patterns*—Outliers frequently occur in one of two general appearance patterns: (*1*) *None or Infrequent*—There are no outliers or there are only a few outliers; one or two for every 20 data cells in a Table 1 format or (*2*) *Extensive*—Outliers occur in greater numbers, three, four, or more for every 20 data cells and frequently in several of the cells for any laboratory. When outliers are extensive they may frequently be of substantial magnitude. There are of course some intermediate cases between these two extremes.

7.5 *Rationale 1 for Outlier Rejection*—There are two points of view on what significance level should be adopted for outlier rejection. The extremely conservative approach maintains that outliers should rarely be eliminated in any ITP. This is based in part on the concept that in the preliminary stages of test method development, outlier rejection will lead to an overly optimistic impression of the quality of the test method. This approach usually adopts a probability significance level of 0.5 % (p = 0.005), for outlier rejection. This approach has some limited merit for the initial stages of development for any test method especially when only a few laboratories participate in an ITP. This significance level is specified by Practice E691. However, this approach has some serious limitations as described as follows.

7.6 *Rationale 2 for Outlier Rejection*—For well-established test methods and any group of laboratories, experience has taught that there is a distribution of skill and testing competence, from poor to good. This capability range argues for a more realistic approach to the outlier issue; the use of a 5 % significance level, p = 0.05 (or a 95 % confidence level) for the declaration of incompatible values as outliers. This is the usual level for most statistical significance tests and will in general reject the results of laboratories that have poor quality control for internal testing and are in need of improved operating procedures.

7.6.1 Allowing a few *poor* laboratories to inflate the evaluated precision gives a false negative impression of the true precision defined by laboratories with good control of testing operations. The precision of the *good* laboratories (the majority of those participating) should be the benchmark for industry-wide precision level for any test method. The use of the robust General and Special Precision procedures to identify these poor quality control laboratories can lead to a general industry-wide improvement for any test method provided that feedback is employed to encourage the poor performing laboratories to improve testing operations.

7.7 *Sequential Review of Outliers*—Experience in outlier review at the 5 % significance level raises the issue of a subsequent review of the database once the 5 % outliers are deleted. To properly frame this operation, recall that the h and k statistics represent ratios of either individual cell averages or cell standard deviations to the *across all laboratory* standard deviation for each parameter. The influence of any outlier extends to both the outlier value itself (the numerator for h and k), as well as the standard deviation for all laboratories (the denominator for h and k).

7.7.1 The removal of 5 % significance outliers will generate a second (or Revision 1) database with substantially reduced *across all laboratories* or denominator standard deviation for either the h or k statistics, or both. When outliers are deleted the resulting revised database is one that might have been obtained had the outlying laboratories not volunteered for the ITP. The question now presents itself: Can this *R1* database be reviewed again for h and k outliers using the newly calculated *across all laboratory* h and k standard deviations.

7.7.2 For any ITP that contains six or more original laboratories, the answer to this question is yes, and the second or revised database should be reviewed for any potential outliers. However, to guard against the generation of an excessively optimistic precision, the significance level for this second review should be more rigorous than for the initial review and should be conducted at the 2 % significance level. For any ITP that contains less than 6 laboratories, the decision to conduct a second review is left to the judgment of the analyst.

7.8 *Special Case Circumstances for Outliers*—In the analysis of larger databases (12 or more laboratories) it may happen (infrequently) that there are three or more suspicious or potential h and/or k outliers in the database for Step 2. When this happens the calculated h or k values for the three or more laboratories may be close to the critical values but usually do not exceed them. If there is a fairly large difference between these calculated h or k values (for the suspicious three or more outlier values) and the remaining h or k values that constitute the bulk of the laboratories that give good agreement, it is recommended that a 5 % significance level be used for the Step 2 procedure in place of the 2 % significance level. Recent

experience (2004) in the analysis of such databases has shown that this exception to the standard Step 2 procedure will eliminate those laboratories that do not have good control over their testing operations. This prevents the results of such laboratories from inflating the precision estimate of the 'in good agreement' laboratories (the majority of the total number) that should constitute the benchmark for the industry for the test method in question.

8. General Precision: Analysis Step 1

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

8.1.1 *Calculating Cell Averages, Cell Ranges, or Standard Deviations*—A table in the format of Table 2 is prepared by calculating the average of the n replicates per cell as given in Table 1. After cell averages have been calculated they should be reviewed for any apparent outlier values as described in 8.1.3 and these noted for evaluation as given in the formal Step 1 outlier rejection procedure as described in 8.3 and 8.4. See also Annex A3.

8.1.2 A table in the format of Table 3 is prepared by calculating, for all cells, the standard deviation for the n replicates per cell. Alternatively, cell ranges, denoted by w, the absolute difference between the maximum and minimum values in each cell, may be calculated. Both the cell ranges and the cell standard deviations should also be reviewed for any apparent outlier values and these noted for evaluation as given in the formal Step 1 outlier rejection procedure as described in 8.3 and 8.4. See Annex A3.

TABLE 2 Precision Program—Cell Averages[A]

Material (j) ==> Laboratory (i)	1	2	3	4	...	q
1						
2			avg $Yijk$			
3						
4						
5						
...						
p						

[A] Table layout for uniform level ITP.

Notation sed:
Laboratories, a total of p, $L(i)$ = 1, 2, 3, ... p
Materials or Levels, a total of q, $m(j)$ = 1, 2, 3, ..., q
Replicates, a total of n per cell; a cell = each combination of $L(i)$ $m(j)$; normally $n = 2$
avg $Yijk$ = average of cell (ij) for n test results

TABLE 3 Precision Program—Cell Std Deviations[A]

Material (j) ==> Laboratory (i)	1	2	3	4	...	q
1						
2			SD $Yijk$			
3						
4						
5						
...						
p						

[A] Table layout for uniform level ITP.

Notation used:
Laboratories, a total of p, $L(i)$ = 1, 2, 3, ... p
Materials or Levels, a total of q, $m(j)$ = 1, 2, 3, ..., q
Replicates, a total of n per cell; a cell = each combination of $L(i)$ $m(j)$; normally $n = 2$
SD $Yijk$ = standard deviation of cell (ij) for n test results

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

8.1.3.1 Using the prepared tables, for each laboratory-material pair of columns, sort the cell averages (or cell ranges) in ascending order (across all laboratories) retaining the laboratory number with the cell value in the sorting operation. For each parameter (cell average or cell range), plot the parameter value versus the laboratory number in ascending parameter value order, using a line plot procedure. This is designated as an *ascending order trend* or AOT plot.

8.1.3.2 For an ITP with no outliers, the cell average plot is typically a positive slope straight line with some reasonable degree of point scatter. If any outliers are present, they will be at the opposite ends of the plot, and will show substantial departure from the straight line of the central data point region. The cell range plot may contain more curvature from the low end (which may contain zero values) toward the central point region, but it will also clearly show the outliers at the high value end of the plot. Ascending order plots will be used in the operation to replace outlier values with *replacement values* as outlined in Annex A5.

8.2 *Calculation of Precision for Original Database*—Comprehensive specific instructions for this are given in Annex A4.

NOTE 2—In Sections 8 – 10 Tables A4.1 to A4.6 are discussed; these are tables that the analyst will prepare in a computer spreadsheet according to the instructions as outlined in Annex A4. There are no actual (printed) Tables A4.1 to A4.6 (with the appended letter designations) in the standard. The table letter designations R1, R2, OR, and OD appended in

pairs to the usual ASTM table identification numbers help to make the tables self-identifying. Their use improves comprehension both in table generation and in reviewing the tables during analysis. The use of these appended designations is further explained and discussed in Sections 8 – 10. See also A4.2.2 and A4.3 in Annex A4.

The test result values for the original database are entered into a table, designated as Table A4.1. This tabular format is also described as Table 1 in the main body of the standard. However, to preserve continuity between Annex A4 and the instructions of 8.2, the table identification terminology of Annex A4 will be used.

8.2.1 The next step is to set up a tabular format designated as Table A4.2 for cell averages and cell averages squared. The corresponding values in Table A4.1 are the argument values for Table A4.2.

8.2.2 Table A4.3 is generated next, cell average deviations, denoted by d and the calculated h-values. The corresponding values in Table A4.2 are used as the arguments for Table A4.3. Refer to Annex A3 for cell deviation d and h-value calculations.

8.2.3 Table A4.4R for cell ranges and cell ranges squared and Table A4.4S for cell standard deviations and cell variances (standard deviations squared) both address the same issue; the within cell variation. It is recommended that both tables be generated in the analysis.

8.2.4 Table A4.5 is used to calculate k-values for each cell in the database. The corresponding values in Table A4.4S are used as the arguments to calculate k-values in Table A4.5. Refer to Annex A3 for k-value calculations.

8.2.5 Table A4.6 is used to calculate the precision parameters, r, R, (r), and (R). Values for T_1, T_2, T_4 and n and p are required to calculate r and R. See the imbedded calculation algorithms, 1 to 5, in Table A4.6 and also Annex A4 for the details on these calculations.

8.3 *Detection of Outliers at the 5 % Significance Level Using h and k Statistics*—The calculated values for h in Table A4.3 and the calculated values of k in Table A4.5 are reviewed for potential outlier values.

8.3.1 If the Table A4.3 h-value for any cell equals or exceeds the 5 % significance level critical h-value as given in Table A3.1, that particular cell value is declared as an outlier.

8.3.2 If the Table A4.5 k-value for any cell equals or exceeds the 5 % significance level critical k-value as given in Table A3.1, that particular cell value is declared as an outlier.

8.3.3 If outliers are detected, a summary of the outliers detected is presented in the form of a sub-table at the bottom of Table A4.6 showing the laboratory numbers that had 5 % significance outliers for both h and k for each material. See Table in Annex A6 for an example. When outliers are present, a revised database is generated by the use of either Option 1, outlier deletion, or Option 2, outlier replacement, as described in 8.4.

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

8.4 *Generation of R1 Database Using Outlier Option 1 or 2*—If outliers are detected, the database is revised using either Option 1 or 2. The revision procedure is described in A4.3.

8.4.1 Option 1 is the deletion of the n cell values in Table A4.1 that are indicated as outliers and the correction of ERR indications in certain cells in Tables A4.2 to A4.6 that result from the deletion process as described in A4.3. The deletion applies to both cell averages as indicated by equal or greater than 5 % critical h-values and to cell standard deviations as indicated by equal or greater than 5 % critical k-values. Once all ERR corrections have been made the database is designated as a *R1* database. Each *R1* table designation contains the appended symbols, R1-OD, outliers deleted. This revised OD database will be reviewed again for outliers now at the more critical 2 % significance level as described in Analysis Step 2.

8.4.2 Option 2 is the replacement of the n cell values in Table A4.1 that are indicated as outliers. The replacement applies to both cell averages and to cell standard deviations as indicated by greater than 5 % critical values. For either the h or k values, the replacement is a two sequence, one- or two-stage process. All of the details for this are described fully in Annex A5. Once replacements have been generated by the Annex A5 procedure, they are inserted into the database, replacing the outlier values, to produce a *R1* database using the table identification symbol, *R1-OR*, outliers replaced. This revised OR database will be reviewed again for outliers now at the more critical 2 % significance level as described in Analysis Step 2.

8.5 *R1 Database Tables*—A second set of tables in the format of A4.1 to A4.6 is prepared for the Step 2 analysis. As previously noted, this second set should be (*1*) tables designated as A4.1-R1-OD to A4.6-R1-OD for the selection of outlier Option 1, or (*2*) tables designated as A4.1-R1-OR to A4.6-R1-OR for Option 2 outlier replacement. Once the deletions or the replacements have been made, according to the instructions in Annex A4, the new set of precision values will appear in Table A4.6-R1-OD or Table A4.6-R1-OR depending on the option chosen.

9. General Precision: Analysis Step 2

9.1 *Detection of Outliers at the 2 % Significance Level Using h and k Statistics*—The calculated values for h in Table A4.3-R1-OD or Table A4.3-R1-OR and the calculated values of k in Table A4.5-R1-OD or A4.5-R1-OR are reviewed for potential outlier values at the 2 % significance level. The calculated h and k values must be greater than the 2 % significance level for outliers to be rejected. For each of these tables, a sub-table is generated at the bottom of either table to summarize the results of the h and k comparisons of calculated values versus critical values. See Annex A6 for an example. If outliers are detected, the database is revised using either Outlier Option 1 or 2. The revision procedure is described in A4.3.

9.1.1 Option 1 is the deletion of the n cell values in Table A4.1-R1-OD that are indicated as outliers and the correction, as previously noted, of ERR indications in certain cells in Tables A4.2-R1-OD to A4.6-R1-OD that result from the deletion process. Once all ERR corrections have been made the database is designated as a R2-OD database. This revised OD database will be used for the operations of Analysis Step 3.

9.1.2 Option 2 is the replacement of the *n* cell values in Table A4.1-R1-OR that are indicated as outliers. The replacement applies to both cell averages as indicated by greater than 2 % critical values for either *h* or *k*. The replacement is a two sequence, one- or two-stage process. All of the details for this are described fully in Annex A5. Once replacements have been generated by the Annex A5 procedure, they are inserted into the database to produce a R2-OR database. This revised OR database will be used for the operations of Analysis Step 3.

10. General Precision: Analysis Step 3

10.1 *Final Precision Results*—Although the Fig. 1 decision tree diagram or flow sheet implies that Analysis Step 3 involves an analysis operation, the analysis has already been conducted automatically with the outlier treatment as described in Step 2. Step 3 is really a review of the precision results that have been obtained previously from the *R2* database. The automatic calculation procedure of the interlinked Tables A4.1 to A4.6 produces the new precision results once either outlier Option 1 (deletion) or Option 2 (replacement) have been selected and the deletion or replacement operations completed.

10.1.1 Analysis Step 3 is the end of the precision calculations when outliers have been found at both the 5 % and 2 % significance levels. The results for either Table A4.6-R2-OD or Table A4.6-R2-OR are used to generate a Precision Table for the test method under review. Refer to Section 13 on the appropriate format for a precision table, see Table 6, and the appropriate text for the precision clause or section.

11. Special Precision Analysis—Carbon Black Testing

11.1 *Background*—The evaluation of test methods for the carbon black manufacturing industry shall be conducted by the procedures as described in this section for the typical uniform level experimental design. These procedures differ from the requirements as set forth in the General Precision procedure as follows: (*1*) the number of replicates in each cell of the Table 1 format is specified as four, (*2*) the cell averages and cell standard deviations are reviewed for potential outliers by a procedure that differs from that as specified for General Precision in terms of the potential number of outliers deleted, see 11.3.1, and (*3*) special calculations are conducted to select the mode of precision expression for reproducibility (absolute or relative) that is most free of influence of the magnitude of the measured property on the reported precision value. Note also that in reviewing discordant data values as potential outliers, only the 5 % significance level *h* and *k* values in Table A3.1 are used to reject outliers.

11.1.1 The terminology as set forth in Section 3, as well as the terminology in Annex A1 shall apply to the procedures for this special precision. Frequently in the carbon black industry and elsewhere, the word *sample* is used as a synonym for the word *material* in the discussion of interlaboratory testing, that is, a grade of carbon black used in an ITP is frequently referred to as a *sample*. This can be a source of confusion and is not consistent with the terminology of this practice. To avoid confusion, the terms *material* or *target material*, or both, shall be used for what is tested (for example, a series of different grades of carbon black), in the process of organizing, reporting, and discussing interlaboratory test programs and the precision parameters as calculated from such programs.

TABLE 4 Initial Data Format for Each Material—Special Precision: Carbon Black Testing

	Material (*i*)		
Date	Test Result 1	Test Result 2	Operator or Technician
Day 1	xxx	xxx	xxxxx
Day 2	xxx	xxx	xxxxx

TABLE 5 Format for Interlaboratory Data—Special Precision: Carbon Black Testing

	Material 1		Material 2		Material *q*	
Laboratory Number	Cell Avg	Cell Std Deviation	Cell Avg	Cell Std Deviation	Cell Avg	Cell Std Deviation
1	xx	xx	xx	xx	xx	xx
2	xx	xx	xx	xx	xx	xx
...	xx	xx	xx	xx	xx	xx
p	xx	xx	xx	xx	xx	xx

11.2 *Materials Selected, Initial Data Recording*—The number of materials (or target materials), which will normally be different grades of carbon black, shall be selected as recommended in 6.1.6. It is recommended that at least five materials be selected for any ITP. This number of materials provides for at least four degrees of freedom in evaluating the coefficient of determination as described in 11.4.

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

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

TABLE 6 Example of Precision Table Organization—Type 1: Precision for ASTM XXXXX

NOTE 1—Measured Property = xxxxxx, in xx.

		Within Laboratories			Between Laboratories			
Material	Mean Level	*Sr*	*r*	(*r*)	*SR*	*R*	(*R*)	No. Laboratories[A]
A								
B								
C								
D								
Pooled or Average Values								

[A] List number of laboratories in final database, also list the Option chosen; if Option 2, indicate with number of laboratories in parentheses.

Notation used:
Sr = within-laboratory standard deviation (in measurement units)
r = repeatability (in measurement units)
(*r*) = repeatability (in percent of mean level)
SR = between-laboratory standard deviation (for total between laboratory variation in measurement units)
R = reproducibility (in measurement units)
(*R*) = reproducibility (in percent of mean level)
See text of Precision Clause for discussion of precision results of this table

11.3.1 The review process for carbon black ITP testing is based on the premise that a substantial number of laboratories participate in the ITP, some number greater than 20. For each material in the Table 5 format, calculate the *h*-value and *k*-value for each cell (or laboratory) by the procedure as specified in Annex A3. A value for *h*(crit) and *k*(crit) at the 5 % significance level is selected from Table A3.1. The calculated *h*-values and *k*-values are reviewed to determine if any are greater than *h*(crit) or *k*(crit). The rejection process is conducted on the basis of the following rules.

11.3.1.1 If there are no calculated *h* -values or *k*-values greater than *h*(crit) or *k*(crit), all cell averages or standard deviations, or both, are retained.

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

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

(1) If there are 20 or fewer laboratories in the ITP, reject only one cell average or cell standard deviation per material, with the greatest (absolute value) calculated *h* or *k* value.

(2) If there are greater than 20 laboratories in the ITP and there are several *h*-values or *k*-values, or both, greater than the respective *h*(crit) and *k*(crit), reject cell averages or cell standard deviations, or both, starting with the highest (absolute value) calculated *h* and *k* values and proceeding downward, until the number of remaining laboratories is 20, or all the *h* values greater than *h*(crit) or *k* values greater than *k*(crit) have been rejected, and use this as the database for precision evaluation.

11.3.2 If any outliers are rejected following the rules of 11.3.1, the resulting database with outlier data deleted is designated as an *R1* database. Conduct a second precision analysis on the *R1* database to generate the final table of precision parameters to be used in the operations as described in 11.4.

11.4 *Expressing the Evaluated Precision for Carbon Black Testing*—Calculate the precision parameters *r*, *R*, (*r*), and (*R*) using the formulas as specified in A4.1. The calculations shall be on the original database if there are no outliers, or on the *R1* database after any potential outlier rejection following 11.3.1. Plot the values of *R* and (*R*) versus *M* or Y_{AV}, the mean value for material measured property, for all materials in the ITP. Perform a least squares regression for both relationships, and record the coefficient of determination, designated as C_d, for each parameter *R* and (*R*).

11.4.1 Select for the mode of precision expression, the parameter, *R* or (*R*), with the lowest value for C_d. This establishes which of the two modes of expression has the least relationship to the level of the measured property or inversely which parameter is the most independent of the measurement level. This lowest C_d or most independent parameter is to be used to prepare a final precision table in the format as indicated by Table 6. The selected mode of expression applies to both repeatability and reproducibility. Follow the rules for expressing General Precision as outlined in Section 12 using, where appropriate, the designation Special Precision. The columns [*r* and *R* or (*r*) and (*R*)] for the parameter with the highest C_d may be omitted from the format of Table 6.

12. Format for Precision Table and Section or Clause in Test Method Standards

12.1 *General Precision Table*—Precision is expressed in summary form in a Table 6 format. Each summary precision table should have a heading to indicate: (*1*) use of General Precision or Special (Carbon Black) Precision, (*2*) the type of precision (Type 1 or Type 2), see 5.1.3 – 5.1.5, and (*3*) the measured property and its measurement units.

12.1.1 For each material tested, the following shall be recorded: (*1*) the material identification, (*2*) the mean level of the measured property, (*3*) the repeatability standard deviation, *Sr*, (*4*) the repeatability, *r*, (in measurement units), (*5*) the relative repeatability, (*r*), in percent of the mean level, (*6*) the reproducibility standard deviation, *SR*, (*7*) the reproducibility, *R*, in measurement units, (*8*) the relative reproducibility, (*R*), in percent of the mean level, and (*9*) the number of laboratories in the final database as used to evaluate precision.

12.1.2 If there are no outliers, the value for item (*9*) in 12.1.1 is the number of laboratories for the original database. If outliers are found and Option 1 deletion is used, the number will be less than the number for the original database. If Option 2 outlier replacement is chosen, the number of laboratories that

did not have outliers replaced, should be indicated in this column with a parentheses around the number. Explain this with a footnote to the table.

12.1.3 If the mean value of a measured parameter for any material is very close to zero, the relative precision, (*r*) and (*R*), will be very large. For these circumstances omit the relative expressions of precision from a Table 6 format. The precision table should also contain, as footnotes, an explanation of the table symbols used.

12.1.4 The calculation of pooled or average values is recommended only if the values for *r* and *R* are roughly equal for all materials. When there is a substantial difference in precision among several materials, caution should be exercised in the interpretation of a pooled or average precision. It may have very little meaningful value or applicability.

12.1.5 When there is a substantial difference in precision among the materials, the use of a pooled value may give a false impression of the overall precision. It would be better to direct the user to select a material from the table that is closest in mean value to a specific material under consideration to determine the expected precision instead of using the pooled value. Ultimately, it is the responsibility of those conducting the ITP to determine what constitutes a substantial difference among materials and the reporting of a pooled value.

12.2 *General Precision Section or Clause*—The results of the precision evaluation should be displayed in a section or clause in the test method standard entitled "Precision and Bias." The concept of bias is discussed in Annex A2. The one or more paragraphs or sub-clauses should contain information on the following issues concerning the ITP and the evaluated precision.

12.2.1 A statement that the precision ITP was conducted in accordance with Practice D4483 (the latest revision year designation), and the year the ITP was conducted. A statement that the reader should refer to Practice D4483 for terminology and other details on the precision evaluation.

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

12.2.3 A statement giving (*1*) category of the precision, that is, General Precision or Special Precision (Carbon Black), (*2*) the type of precision, Type 1 or Type 2, (*3*) the number, *p*, of laboratories participating in the ITP, (*4*) the number, *q*, and description of the materials (or target materials) used, (*5*) the number of within-laboratory replicates, *n*, (*6*) the time span for the repeatability or within-laboratory replicates, (hours, days), (*7*) the definition of a test result (average, median of *x* number of determinations or individual measurements), (*8*) the option chosen for outlier treatment, deletion, or replacement, and (*9*) any unusual features of the ITP.

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

12.2.5 Generic statements on repeatability and reproducibility shall be part of the precision clause using the recommended text as set forth as follows. A95 % confidence level (or $p = 0.05$) applies to these statements where Table xx designates the final table as inserted into the test method.

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

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

12.2.6 Bias is defined in A1.2.5 in terms of *bias deviation*, a deviation for a measured value from a true or reference value. Bias is not addressed in this practice, since for essentially all the test methods that will be evaluated for precision, the evaluation of bias is not possible because no reference or true value exists or may be determined. For all such test methods, a statement should be included as the last item in the precision clause, stating that bias is not determined. Using the word bias as a synonym for bias deviation, the suggested statement text is as follows.

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

12.3 *Special Precision Table*—The Special Precision table shall conform to the rules for General Precision.

12.3.1 If the mean value of a measured parameter for any material is very close to zero, the relative precision, (*r*) and (*R*), will be very large. For these circumstances omit the relative expressions of precision from a Table 6 format.

12.4 *Special Precision Section or Clause*—The expression for Special Precision should in general follow the rules for General Precision (12.2.1 – 12.2.5) including the recommended text in 12.2.5.1 and 12.2.5.2 taking into account the differing repeatability and reproducibility procedures as set forth in Tables 4 and 5. State if there are substantial reasons for a differing mode of expression.

13. Report for Precision Evaluation ITP

13.1 A full report of the precision evaluation shall be prepared for any ITP. This is a full comprehensive report of all ITP details, not the report that each participating laboratory prepares and returns as part of the ITP. This full report should

contain information on the details of the organization and execution of the program as follows:

13.1.1 Identify the organization committee, where located, coordinator, and date of ITP,

13.1.2 Category of precision, General Precision, or Special Precision,

13.1.3 Type of precision, Type 1 or Type 2,

13.1.4 Number of laboratories, *p*, and their names without connection to ITP laboratory number,

13.1.5 Number and description of materials or target materials, *q*,

13.1.6 Definition of a test result, number of replicates, *n*, and time span for repeatability,

13.1.7 Information on technicians conducting the testing, any special details,

13.1.8 Details on preparation of materials, how homogeneity is documented,

13.1.9 Details on packaging and delivery of materials to all ITP participants,

13.1.10 Copies of all ITP Data Reports from each participating laboratory,

13.1.11 ITP analysis report, with all tables as designated in Annex A4, full description of all analysis steps, options chosen for outlier rejection, and other required comments,

13.1.12 Table of precision results, comments on outcome, and

13.1.13 Draft of precision section for the test method.

14. Keywords

14.1 general precision; interlaboratory test program; ITP; precision; repeatability; reproducibility; special precision

ANNEXES

(Mandatory Information)

A1. DEFINITIONS FOR SELECTED TERMS CONCERNED WITH PRECISION AND TESTING

A1.1 General Background

A1.1.1 This annex gives comprehensive definitions drafted to contain substantial information content with emphasis on basic concepts. Some ancillary definitions are also given that may promote a better understanding of precision. The word *uncertainty* is used in some of the following definitions in a sense that implies the typical everyday meaning, that is, *a sense of doubt*. The more specific statistical or measurement term *uncertainty* is defined in A1.2.8.1. The definitions as presented in Section 3 of this practice (Terminology) should be understood in using this annex.

A1.2 Basic Statistical Definitions

A1.2.1 *variation, n*—the existence of deviations (differences) among measured element values for repeated independent tests (observations) for a particular class of elements; generated by perturbations produced by one or more *system-of-causes*.

A1.2.1.1 *Discussion*—Deviations are produced by some group of factors or causes, acting within a certain domain that jointly influence the independent measurement or observation output. This is called a variation *system-of-causes*. Typical *system-of-causes* are the unavoidable fluctuations in temperature, humidity, operator technique, fidelity of calibration, and so forth, in a controlled testing domain.

A1.2.1.2 *production variation, n*—variation in properties due to one or more deviation *system-of-causes* that are (*1*) inherent in the process that generates a particular material or class of elements, or (*2*) inherent in the storage or conditioning prior to testing, or both, after such generating processes are complete.

A1.2.1.3 *measurement variation, n*—variation due to one or more deviation *system-of-causes* inherent in the operation of instruments or machines that evaluate certain properties for a material or class of elements, in a defined testing domain.

A1.2.2 *distribution, n*—the characteristic dispersion (scattering) pattern of independent element values generated by one or more variation *system-of-causes* ; defined by the range (maximum to minimum) and the ordering of the element values based on their frequency of occurrence.

A1.2.2.1 *Discussion*—In a graphical sense, ordering is related to the number (or frequency) of element values in any small range (or point) along the element value axis. The independent values may be arranged along this axis in one of three general patterns; (*1*) a unimodal or symmetrical dispersion around a highest frequency central value with a decreased frequency of occurrence the greater their plus and minus difference from the central value (*2*) dispersed in a uniform frequency across a value range, or (*3*) asymmetrically dispersed above and below a central or other special value. The concept of a distribution usually applies to data values rather than physical elements although it may apply to both. Both production and measurement variation may contribute to the total variation. A distribution may be characterized by a mathematical equation called the probability density function that describes the frequency of occurrence of any value, with parameters that define the location and shape of the distribution.

A1.2.3 *normal distribution, n*—a distribution that is symmetric (unimodal) and bell-shaped; it may be defined by a unique probability density function that contains two parameters; the central value or mean and the standard deviation.

A1.2.3.1 *Discussion*—Most of the data obtained from testing, with certain exceptions, will have a unimodal distribution that is normal or approximates a normal distribution. The means of *n* values ($n = \text{or} > 4$) will have an approximate normal

distribution even when the source or individual value distribution (n = 1) is not normal.

A1.2.4 *population, n*—the distribution (collection) of independently distributed elements that constitute the totality for a defined system; it may refer to any one of the following: (*1*) one or several elements, (*2*) a finite but large number of elements, or (*3*) a hypothetical infinite number of elements.

A1.2.4.1 *Discussion*—The preceding definition is for a physical population or a collection of elements. An additional understanding is data population, the collection of all data values produced by testing (or observing) the physical population (or parts thereof). All three population interpretations imply that the elements are generated by some identifiable process and have a rough approximation available for a property range. Testing programs, defined by the testing domain and the sampling program, may vary from a very limited focus of attention, Interpretation 1, to a broad focus of attention, Interpretation 3.

A1.2.5 *random deviation, n*—a difference (plus or minus) between an independently measured or observed value and a known (or estimated) mean or an accepted reference value; the differences vary in magnitude, usually have a normal (unimodal) distribution, and for a long run series of replicates in a stable domain, the sum and mean of the differences is zero.

A1.2.5.1 *Discussion*—Increased replication reduces the random uncertainty of a mean (but not the total uncertainty which may contain a bias component, see bias deviation definition as follows) and provides a more reliable estimate of the true or reference mean property. The definition of *long run* depends on the goal of the testing. For routine testing, the number of replicates, n, may be of the order of 10. For critical testing, n may be two or three times this value. For an intermediate number of replications, the mean of the random deviations may be reduced to a small value that may be considered to be zero, depending on the scope of the testing.

A1.2.6 *bias deviation, n*—a constant difference (plus or minus), absent any random deviations, between an independently measured or observed element value and the true or accepted reference value for a defined domain.

A1.2.6.1 *Discussion*—A bias deviation is a systematic or offset difference produced by some system perturbation. For some domains the offset affects all measurements equally; for others the offset may vary with the magnitude of the measured value. When a reference value is known, the bias deviation may be evaluated by eliminating (or reducing to a negligible value) the effect of random variation by a long-run series of measurements. When the test domain is altered, the magnitude (and less likely the sign) of the bias deviation may change. Any system may have more than one source for bias, and bias deviations, unlike random deviations, do not sum to zero. The word bias is frequently used as a synonym for bias deviation.

A1.2.7 Although accuracy and trueness are not evaluated in this practice, their definitions are given here to provide additional background for a better understanding of their relationship to precision. In some of the definitions to follow, the term *figure of merit* is used. A high figure of merit is an indication of high quality or a high level of goodness of the measurement system for a given parameter of the system.

A1.2.7.1 *accuracy, n*—a test characteristic proportional to the inverse of the difference between an individual test value and the true or reference mean value for some class of elements.

A1.2.7.2 *Discussion*—When the absolute difference is small the inverse is large or high and the testing is said to have *high accuracy*. The observed difference is influenced by both random and bias deviations when both types of deviations exist.

A1.2.7.3 *trueness, n*—a test characteristic proportional to the inverse of the difference between the long-run estimated mean (for high n) and the true or reference mean value for some class of elements.

A1.2.7.4 *Discussion*—Since the estimated mean is a long-run (high n) estimate, the random deviations sum to approximately zero and the influence of random deviations is substantially reduced or eliminated. The observed difference is influenced by the sum of the bias terms only. Thus trueness is a testing concept that is intended to evaluate bias.

A1.2.8 As previously noted, the concept of uncertainty needs some attention. The definition given as follows is a definition that attempts to capture the general nature of the concept. As the definition and discussion indicate, uncertainty is local, and precision is global. It has been defined equivalently, but using different words, by a number of organizations addressing this concept.

A1.2.8.1 *uncertainty, n*—a test characteristic for a local domain; it is the magnitude of the difference between the measured (observed) element value and an accepted reference value and includes both random and bias deviations.

A1.2.8.2 *Discussion*—The word Uncertainty is capitalized in the use as defined in A1.2.8.1 to distinguish it from the ordinary use of the word. As indicated, *goodness* or *merit* and uncertainty (doubt about the measurement), are inversely related. Uncertainty is a characteristic of a local testing domain; each local domain for any defined test, may have a different uncertainty value. Precision (both repeatability and reproducibility) is a characteristic of a global testing domain; the precision values obtained in any ITP are intended for universal application, that is, to a number of laboratories as a group.

A2. STATISTICAL MODEL FOR INTERLABORATORY TESTING PROGRAMS

A2.1 Introduction

A2.1.1 Although this practice does not address the evaluation of bias or accuracy, it is important that the influence of bias in interlaboratory testing be well understood. This annex provides some background on the influence of random and bias deviations by the use of a statistical model for interlaboratory testing.

A2.1.2 In the real world, all measurements are perturbed by a *system-of-causes* that produces test deviations or error. Typical cause systems are fluctuations in atmospheric pressure, temperature, humidity, attention of test operators to the details of a test, and so forth. There are two general *deviation or variation categories* for any specified domain. These are defined by the character and source of *deviations* that perturb the testing or observed values compared to what would be obtained under ideal conditions. Two major categories of variation are:

A2.1.2.1 *Production Variation*—Variation in properties due to one or more deviation *system-of-causes* that are inherent in the process that generates a particular material or class of elements or inherent in the storage or conditioning (prior to testing), or both, after such generating processes are complete.

A2.1.2.2 *Measurement Variation*—Variation due to one or more deviation *system-of-causes* inherent in the operation of instruments that evaluate certain properties for a material or class of elements, in a defined testing domain.

A2.1.3 Within each category, deviations may be of two different types, (*1*) random, plus and minus differences about some central (true) value or (*2*) bias or systematic differences. Both types may occur in either category. The domain of the testing program determines the system-of-causes. These *cause systems* can vary from simple to complex. The production process is broadly defined; it can be (*1*) the ordinary operation of a manufacturing facility, (*2*) a naturally occurring and ongoing process, or (*3*) some smaller scale processing that generates a material or class of objects for testing. The discussion applies to both objects and materials.

A2.1.4 Objects may be discrete manufactured items or test pieces generated by a particular preparation process. Materials may be tested in a direct manner, such as the tensile stress or modulus of a polymer or in an indirect manner, such as the quality of a carbon black or other additive in a standard formulation by a performance-in-rubber test. When performance-in-rubber testing is conducted, the designation *target material* is used for the material, since a composite containing the target material is tested, not the material itself. This composite testing may involve objects or test specimens for the measurement process. These testing concepts, target material, and Type 1 and Type 2 precision are defined and discussed in 5.1.3 – 5.1.5 of this practice.

A2.2 General Model

A2.2.1 For any testing domain, each measurement, y_i, can be represented as a linear additive combination of fixed or variable (mathematical) terms as indicated by Eq A2.1. Each of these terms is an individual deviation or component of variation and the sum of all component deviations is equal to the total variation observed in the individual measurement. It is assumed that all participants test a selected number of classes of objects or different materials drawn from a common lot, employ the same type of apparatus, use skilled operators, and conduct testing according to a test method standard, in one or more typical laboratories or test locations.

$$y_i = \mu_o + \mu_j + \Sigma(b) + \Sigma(e) + \Sigma(B) + \Sigma(E) \qquad \text{(A2.1)}$$

where:

y_i = measurement value, at time (i), using specified equipment and operators, at one laboratory or location (among a total of p laboratories),

μ_o = constant term (mean value), that dictates the general magnitude of the measured parameter for the particular test,

μ_j = constant term (mean value), unique to material or object class (j),

$\Sigma(b)$ = (algebraic) sum of the number of component *bias deviations* in the *process* that produced material or object class (j),

$\Sigma(e)$ = (algebraic) sum of the number of component *random deviations* in the *process* that produced material or object class (j),

$\Sigma(B)$ = (algebraic) sum of the number of component *bias deviations*, for measurement (i), generated by the *measurement system*, and

$\Sigma(E)$ = (algebraic) sum of the number of component *random deviations*, for measurement (i), generated by the *measurement system*.

A2.2.2 An alternative approach is to use a single μ term, that is, μ_r, in place of the two terms $\mu_o + \mu_j$, where both of the characteristics defined by μ_o and μ_j are contained in the single term. Eq A2.1 indicates that there are three groups that contribute to the value of y_i, (*1*) constant terms (population mean values), (*2*) bias deviations, and (*3*) random deviations.

A2.3 Specific Model Format

A2.3.1 A more useful format is obtained when Eq A2.1 is expressed in the format of Eq A2.2 where the generic summations are replaced by a series of typical individual terms or components appropriate to interlaboratory testing on a number of different object classes or materials, over a particular time period.

$$y_i = \mu_o + \mu_j + \Sigma(b) + \Sigma(e) + B_L + B_M + B_{OP} + B_G + E_B + E_W \qquad \text{(A2.2)}$$

where:

B_L = bias deviation term unique to one laboratory or local domain,

B_M = bias deviation term unique to the specific instrument or machine,

B_{OP} = bias deviation term unique to the operator(s) conducting the test,
B_G = generic bias deviation term; to account for other bias factors,
E_B = between laboratory (global domain) random deviation term, and
E_W = within laboratory (local domain) random deviation term.

The B_L term is exclusively a between laboratory bias, the terms B_M, B_{OP}, and B_G may be either between laboratory or within laboratory components depending on the scope of the testing, that is, whether these components are part of the chosen within laboratory repeatability testing. The between laboratory random deviation term, E_B, is usually the sum of a number of subcomponents that represent typical sources of variation between laboratories.

$$E_B = E_L + E_M + E_{OP} + E_G \quad \text{(A2.3)}$$

where:

E_L = random deviation term attributable to a laboratory or location,
E_M = random deviation in the use of the specific instrument or machine,
E_{OP} = random deviation inherent in the operator's technique, and
E_G = generic random deviation term; to account for other random factors.

The within laboratory random deviation term, E_W, may also be the sum of a number of subcomponents due to varying operator(s) technique, different instruments or machines of a given design, if such factors are part of the testing domain, in addition to the time period for repeatability measurements. Typical B_G or E_G testing perturbations, may be bias and random components due to temperature, long-term time period (time of the year), and so forth.

A2.3.2 $\mu_o + \mu_j$ *Terms*—In the absence of bias or random deviations of any kind, a number of materials or object classes would have individual measured test values given by the sum of the two terms, $\mu_o + \mu_j$. The term μ_o uniquely characterizes the general magnitude of the measured parameter. Each material or object class would be characterized by the value of μ_j, which would produce a varying value for the sum $[\mu_o + \mu_j]$ across the number of materials or object classes in the test program and the sum would be the *true* or unperturbed test value.

A2.3.3 *Production Terms* $\Sigma(b) + \Sigma(e)$—There will always be some bias and random variation in the materials or object classes produced by the process that generates them. These usually unknown number of bias and random variations are designated by $\Sigma(b) + \Sigma(e)$. For testing in general, appropriate sampling and replication plans will reduce the random components to some selected level. However, increased sampling and replication does not reduce bias components; such action merely enhances the fidelity of the evaluated magnitude of these effects, if reference materials are available. Reducing or removing bias requires (*1*) special test programs to discover and eliminate the causes or (*2*) a documented correction procedure that eliminates the bias. For most precision ITPs, special care is required to ensure some minimal level of variation in the lots of materials selected for the program, that is, to make them as homogeneous as possible. Any residual production variance adds to the measurement variance or basic precision as evaluated by the ITP.

A2.3.4 *Measurement Bias Terms*—Bias deviations may be divided into two classes: *local* or *global*. A local bias is a fixed offset that applies to certain specific conditions within a larger testing domain, for example, a single test machine or laboratory among many machines or laboratories. Such biases are the principle component of between laboratory differences, that is, one laboratory or test instrument is always low or high in comparison to other laboratories or instruments.

A2.3.4.1 When the domain consists of a large number of machines or laboratories, the local biases may be *variable* (plus or minus) deviations unique to each of these machines or laboratories and the distribution may be either random with a zero mean in the long run or a nonrandom finite distribution with a nonzero mean. A global bias is either (*1*) a fixed offset that applies across the whole testing domain and is unique to a generic condition that is common within the domain or (*2*) an inherent deviation in a particular design of a test apparatus. Although more than one global bias may exist, global biases usually are not considered to have a distributional character.

A2.3.4.2 Bias terms that are fixed under one *system-of-causes* may be variable under another *system-of-causes* and vice versa. As an example, consider the bias terms B_L and B_M which apply to most types of testing. 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 potentially having a unique value. For a domain consisting of a *number of typical laboratories*, each with one machine, both B_L and B_M would be variable for the domain, but both B_L and B_M would be fixed or constant for the *system-of-causes* in each laboratory. One or more generic bias terms, B_G, may be present in any test domain. These represent unique bias effects not attributable to test machines, operators, or laboratories.

A2.3.5 *Measurement Random Terms*—These terms are deviations or components that are frequently called *error*. Random deviations are plus or minus values that have an expected mean of zero over the long run. As indicated in Eq A2.2 there are three potential sources of random variations: laboratories, test machines, and operators, in addition to the special case where another source, a generic source, is an important component. The distribution of these terms is assumed to be approximately normal but in practice it is usually sufficient if the distribution is unimodal. The value of each random term influences the measured y_i value on an individual measurement basis. However, in the long run, when y_i values are averaged over a substantial number of measurements, the influence of the random terms may be greatly diminished or eliminated depending on the sampling and replication plan, since in the long run each term averages out to zero (or approximately zero) and the *mean* y_i is essentially unperturbed.

A2.3.6 *New Term,* $M_{(j)}$—With highly replicated testing programs (both production and test measurement replication)

the average values obtained in any program are estimates of the value of a new combined term as given as follows:

$$M(j) = [\mu_o + \Sigma(b) + \Sigma(B)] + \mu_j \quad \text{(A2.4)}$$

and $M(j)$ is the mean value for the material or class of objects tested, for one laboratory or location, j, for the specific equipment and operators used during the existing time period. It contains bias components or potential bias components for all of these conditions. If all biases are fixed for any given program, the three terms in the bracket can be considered as a constant, and the *average test value* varies across the number of materials or object classes because of the varying value of μ_j. If the biases vary across the system, then both μ_j and the biases influence the average value for any candidate test and material.

A2.4 Evaluating Process and Measurement Variance

A2.4.1 Eq A2.1 may be used to illustrate how the variance of individual measurements, y_i, may be related to the terms or components of the equation. Recall that μ_o and μ_j are constants, $\Sigma(b)$ and $\Sigma(e)$ refer to the sum of bias and random components, respectively, for the production process, and $\Sigma(B)$ and $\Sigma(E)$ refer to the sum of bias and random components, respectively, for the test measurement operation. The magnitude of the individual components are ordinarily not known and the equation can be simplified by combining the bias and random components for both sources where $\Sigma(b, e)$ = sum of bias and random components for the *production process* and $\Sigma(B, E)$ = sum of bias and random components for the *measurement procedure*.

$$y_i = \mu_o + \mu_j + \Sigma(b,e) + \Sigma(B,E) \quad \text{(A2.5)}$$

The variance of any individual measurement y_i, designated by $s^2(y_i)$ is:

$$s^2(y_i) = [\Sigma\ \mathrm{Var}(b,e)] + [\Sigma\ \mathrm{Var}(B,E)] \quad \text{(A2.6)}$$

where:

$[\Sigma\ \mathrm{Var}(b,e)]$ = variance, that is the sum of individual bias and random variances, for the *production* process, and

$[\Sigma\ \mathrm{Var}(B,E)]$ = variance, that is the sum of individual bias and random variances, for the *measurement* procedure.

Eq A2.6 can be written in simplified format as:

$$s^2(y_i) = s^2(tot) = s^2(p) + s^2(m) \quad \text{(A2.7)}$$

where:

$s^2(tot)$ = total variance among the materials or object classes in a test program,

$s^2(p)$ = variance due to the production process, and

$s^2(m)$ = variance due to the measurement operation.

A2.5 Relating the Bias and Random Terms to Measurement Precision

A2.5.1 *Between Laboratory Variation*—The expanded series of B terms in Eq A2.2 gives insight into potential sources of measurement bias in any testing domain. However, to express the between laboratory test results in relation to the B terms, it is convenient to use a collective or total B term designated as $B(Tot)$, which is the algebraic sum of all B terms. The variance of $B(Tot)$ is the between-laboratory bias variance. When the results of an ITP for precision are analyzed, the total between-laboratory variance, S^2_L, is the sum of the between-laboratory bias variance plus the total between-laboratory random variance due to E_B terms, designated as $E_B(Tot)$. $E_B(Tot)$ is defined as the sum of the variance of all random E_B terms as expressed in Eq A2.2. Thus:

$$\mathrm{Var}[B(Tot)] + \mathrm{Var}[E_B(Tot)] = S^2_L \quad \text{(A2.8)}$$

where:

S^2_L = between-laboratory variance, with S^2_L evaluated for an ITP as given by Eq A2.9.

$$S^2_L = S^2(Y_i) - (S^2_r/n) \quad \text{(A2.9)}$$

where:

$S^2(Y_i)$ = variance among the cell averages across all laboratories, with Y_i defined as cell average for any laboratory, i, and

S^2_r = within cell variance pooled across all laboratories, adjusted or divided by n, the number of values per cell, to put both variances on an equivalent basis of mean values (averages of n).

As indicated by Eq A2.9, S^2_L is a special derived variance that does not include the random within-laboratory variation.

A2.5.2 *Within-Laboratory Variation*—Within any laboratory, repeated testing (for a defined test domain) on a given material or at a given level generates a series of measurement values and a series of values for E_W. The within laboratory variance, S^2_W, is given by Eq A2.10:

$$\mathrm{Var}[E_W] = S^2_W \quad \text{(A2.10)}$$

For a standardized test method, it is general practice in precision evaluation and analysis to assume that S^2_W will be approximately equal for all laboratories. On this basis, the individual values for S^2_W (one for each laboratory for each material) may be pooled to obtain a collective or global value representative of all laboratories. Therefore, for each material or level, S^2_W is a universal value characteristic of all laboratories in the ITP and by assumption, all laboratories likely to use the test method. However, experience has shown that the skill and the internal control practices used in conducting tests varies even among well-experienced laboratories.

A2.5.3 This varying testing skill and general laboratory competence can be addressed by the use of a generic within laboratory term, E_{WG}, where the double subscript denotes a within laboratory generic random deviation component. Using this, a more well-defined within laboratory variance is:

$$\mathrm{Var}[E_W] + \mathrm{Var}[E_{WG}] = S^2_W\,(\mathrm{sp}) \quad \text{(A2.11)}$$

where S^2_W (sp), the specific within laboratory variance, is equal to the sum of the universal within laboratory test variance characteristic of the test, E_W, and another variance component unique to a particular laboratory. The variance associated with E_{WG} is essentially zero for good well-controlled laboratories. Allowing for the potential existence of E_{WG} terms among laboratories, the repeatability variance, S^2_r, is defined by Eq A2.12:

$$\mathrm{Var}[E_W] + \mathrm{Var}[E_{WG}] = S^2_r \quad \text{(A2.12)}$$

where S^2_r is a pooled value across all laboratories for any material or level, each individual laboratory value having (n – 1) degrees of freedom where n = number of replicates tested.

A2.5.4 *Combined Between and Within Laboratory Variation*—The total combined variation for between and within laboratory test results for any selected time period, defined as the reproducibility variance and designated as S^2_R, is the sum of four potential sources of variation.

$$\mathrm{Var}[B(Tot)] + \mathrm{Var}[E_B(Tot)] + \mathrm{Var}[E_W] + \mathrm{Var}[E_{WG}] = S^2_R \quad \text{(A2.13)}$$

The estimate of this variance, S^2_R, is equal to the total variation among all values for each material (or level) in the ITP. Recall that $B(Tot)$ represents a number of potential sources of between laboratory bias. Interlaboratory testing experience has demonstrated that the left to right order of the variance terms in Eq A2.13 is the approximate order of magnitude of these terms.

A2.5.5 *Defining Repeatability and Reproducibility*—Repeatability and reproducibility are each equal to a range or interval that is a special multiple of the respective standard deviation. The repeatability, designated as r, is given by

$$\text{repeatability} = r = \varphi\,(2)^{1/2}\,S_r \quad \text{(A2.14)}$$

and reproducibility, designated as R, is given as:

$$\text{reproducibility} = R = \varphi\,(2)^{1/2}\,S_R \quad \text{(A2.15)}$$

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

A2.5.5.1 Thus $[(2)^{1/2}\,S_R]$ is the standard deviation of differences. The factor φ depends on both the total degrees of freedom in the estimation for either of the standard deviations and on the shape of the distribution of the variable bias terms and the E terms. The normal assumptions for these are (*1*) the distributions are unimodal, (*2*) the number of test results is sufficient (approximately 20), and (*3*) a probability level of p = 0.05 or confidence level of 95 % is chosen. Under these assumptions, the value of φ is similar to a t- value or approximately 2.0, and therefore the simplified expressions for r and R are:

$$\text{repeatability} = r = 2.83\,S_r \quad \text{(A2.16)}$$

$$\text{reproducibility} = R = 2.83\,S_R \quad \text{(A2.17)}$$

A3. CALCULATING THE *h* AND *k* DATA CONSISTENCY STATISTICS

A3.1 General Background

A3.1.1 The test results of a typical ITP when placed in a Table 2 and Table 3 format may well contain cell values that appear to be outliers. It is necessary to review the data and make a decision on how to treat these outliers. This should identify any one, two, or more potential outliers that have substantial deviations from the mean for a particular material in the database. Outlier treatment consists of rejection of all identified outliers using one of two options. Option 1 is the deletion of the outliers to generate a reduced size database. Option 2 is the replacement of the outliers by a procedure that maintains the character of the distribution of the non-outlier data.

A3.1.2 Some outlier rejection techniques use the difference between the outermost value and the adjacent value as the basis for rejection. This works well as long as potential outliers do not occur as pairs with minimal pair separation, but substantial separation from the nearest value in the database. Frequently, when this occurs, the rejection techniques fail to identify the outermost value(s) and the rejection iteration process stops.

A3.1.3 Both the General and the Special Precision sections of this practice use two particular parameters, called *consistency statistics*, to reject potential outliers, the h and k values as developed by J. Mandel and used in Practice E691. The h statistic is a parameter used to review the between-laboratory cell averages for potential outliers, and the k statistic is a parameter used to review the between-laboratory cell standard deviations for potential outliers.

A3.2 Defining and Calculating the *h* Statistic

A3.2.1 *h-value*—The between-laboratory *cell average* consistency statistic, h, is calculated using the cell averages for all laboratories and is defined as follows for each material or q level in the ITP.

$$h = d/S\,(Y_{AV}) \quad \text{(A3.1)}$$

where:

d = $Y_{AV}(i) - Y_{AV}$,
$Y_{AV}(i)$ = individual cell average, for laboratory (i),
Y_{AV} = average of *all* cells, for any material, and
$S(Y_{AV})$ = standard deviation of cell averages for any material or q level across all laboratories.

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

A3.2.2 *Calculating Critical h-values*—After an h-value is calculated for each laboratory for each material, the values are

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

$$h(\text{crit}) = (p - 1)\, t/\left[p\,(t^2 + p - 2)\right]^{1/2} \qquad \text{(A3.2)}$$

where:

p = number of laboratories in the ITP,
t = student's t at selected significance level, with df = $(p - 2)$, a 2-tailed value, and
df = degrees of freedom.

A3.3 Defining and Calculating the *k* Statistic

A3.3.1 *k-value*—The *cell standard deviation* consistency statistic, *k*, is an indicator of how the within-laboratory individual cell standard deviation for any selected laboratory, compares to the overall (or pooled across all laboratories) *cell standard deviation*. The usual approach to tests of significance for variability statistics is the use of the *F*-ratio, a ratio of two variances. However, the *k*-value is expressed as a ratio of two standard deviations since it is easier to comprehend this ratio when reviewing data. The *k*-value is developed as follows.

A3.3.2 In the usual *F*-ratio approach, the significance of any individual cell-variance compared to the pooled variance of all the cells (for any material) excluding the one cell being tested is given by:

$$F = S^2_{(i)} / \left[\Sigma\, S^2_{(p-i)} / (p - 1)\right] \qquad \text{(A3.3)}$$

where:

$S^2_{(i)}$ = cell variance being tested for potential significance, laboratory (i),
$\Sigma\, S^2_{(p-i)}$ = sum of cell variances, excluding cell (i), and
p = the number of laboratories in the ITP.

The *k*-value is defined by Eq A3.4 and is calculated for each material by:

$$k = S(i)/S_r \qquad \text{(A3.4)}$$

where:

$S(i)$ = cell standard deviation for laboratory (i), and
S_r = pooled cell standard deviation (across all laboratories), this is the initially calculated repeatability standard deviation (see Eq A3.5).

A3.3.3 *Calculating Critical k-values*—For purposes of calculating critical *k*-values, designated as *k*(crit), the following development is presented. The repeatability variance is given by Eq A3.5:

$$S^2_r = \left[\Sigma\, S^2_{(p-i)} + S^2_{(i)}\right]/p \qquad \text{(A3.5)}$$

Combining Eq A3.3, Eq A3.4, and Eq A3.5 gives Eq A3.6:

$$k(\text{crit}) = \left\{p / \left[1 + \left((p - 1) / F\right)\right]\right\}^{1/2} \qquad \text{(A3.6)}$$

The degrees of freedom, df, for *F* in Eq A3.6 are $(n - 1)$ for the numerator and $(p - 1)(n - 1)$ for the denominator, where n = number of replicates per cell. Eq A3.6 may be used to calculate *k*(crit) for any values of p and n, at a selected significance level by reference to the critical *F* value at the indicated df for numerator and denominator.

A3.4 Identification of Outliers Using the Critical *h* and *k* Values

A3.4.1 When all the *h* and *k* values have been calculated using Eq A3.1 and Eq A3.4 respectively, and tabulated for any database generated by a particular ITP, they are reviewed to determine if any of the calculated *h* and *k* values exceed the critical *h* and *k* values.

A3.4.2 Table A3.1 gives the 2 % and 5 % significance level (or $p = 0.02$, $p = 0.05$) critical values for both *h* and *k*, for various numbers of laboratories, $p = 3$ to 30, and cell replicates, $n = 2$, 3, or 4. This is used for the two-step procedure for reviewing the database for potential outliers as described in Sections 8 and 9. See especially the recommendations in 7.8.

TABLE A3.1 Critical *h*-Values and *k*-Values at 2 and 5 % Significance Level

Number Labs = *p*	5 % Critical *h*-value	5 % Crit *k*-value for *p* and *n*			Number Labs = *p*	2 % Critical *h*-value	2 % Crit *k*-value for *p* and *n*		
		$n = 2^A$	$n = 3^A$	$n = 4^A$			$n = 2^A$	$n = 3^A$	$n = 4^A$
3	1.15	1.65	1.53	1.45	3	1.15	1.69	1.59	1.52
4	1.42	1.76	1.59	1.50	4	1.47	1.85	1.68	1.59
5	1.57	1.81	1.62	1.53	5	1.67	1.94	1.74	1.67
6	1.66	1.85	1.64	1.54	6	1.80	2.00	1.77	1.65
7	1.71	1.87	1.66	1.55	7	1.89	2.04	1.79	1.67
8	1.75	1.88	1.67	1.56	8	1.95	2.07	1.80	1.68
9	1.78	1.90	1.68	1.57	9	2.00	2.09	1.83	1.69
10	1.80	1.90	1.68	1.57	10	2.00	2.11	1.84	1.70
11	1.82	1.91	1.69	1.58	11	2.07	2.12	1.84	1.70
12	1.83	1.92	1.69	1.58	12	2.09	2.13	1.85	1.71
13	1.84	1.92	1.69	1.58	13	2.11	2.14	1.86	1.72
14	1.85	1.92	1.70	1.59	14	2.13	2.15	1.86	1.73
15	1.86	1.93	1.70	1.59	15	2.14	2.16	1.87	1.73
16	1.86	1.93	1.70	1.59	16	2.15	2.16	1.87	1.73
17	1.87	1.93	1.70	1.59	17	2.16	2.17	1.87	1.73
18	1.88	1.93	1.71	1.59	18	2.17	2.18	1.88	1.73
19	1.88	1.93	1.71	1.59	19	2.18	2.18	1.88	1.74
20	1.89	1.94	1.71	1.59	20	2.19	2.18	1.88	1.74
21	1.89	1.94	1.71	1.60	21	2.20	2.18	1.88	1.74
22	1.89	1.94	1.71	1.60	22	2.20	2.19	1.88	1.74
23	1.90	1.94	1.71	1.60	23	2.21	2.19	1.89	1.74
24	1.90	1.94	1.71	1.60	24	2.21	2.19	1.89	1.74
25	1.90	1.94	1.71	1.60	25	2.22	2.19	1.89	1.74
26	1.90	1.94	1.71	1.60	26	2.22	2.20	1.89	1.74
27	1.91	1.94	1.71	1.60	27	2.23	2.20	1.89	1.74
28	1.91	1.94	1.71	1.60	28	2.23	2.20	1.89	1.74
29	1.91	1.94	1.72	1.60	29	2.23	2.20	1.90	1.74
30	1.91	1.94	1.72	1.60	30	2.24	2.20	1.90	1.74

[A] *n* = number of replicates per cell within each laboratory for each material or level.

A4. SPREADSHEET CALCULATION FORMULAS FOR PRECISION PARAMETERS, RECOMMENDED SPREADSHEET TABLE LAYOUT AND DATA CALCULATION SEQUENCE

A4.1 Calculation Formulas

A4.1.1 When a dedicated computer program is not available to calculate precision, the repeatability and reproducibility may be calculated using typical spreadsheet procedures and algorithms. The final precision calculations involve a series of sums or totals. The calculation formulas are given in this section. In A4.2 a recommended spreadsheet table layout is presented that facilitates the calculations. A4.3 gives some recommendations for setting up the table sequence and conducting the analysis. Fig. 1 is a decision tree diagram that gives guidance on the sequence of steps. Recall that p = number of laboratories in the ITP.

NOTE A4.1—The calculations were set up for this annex using Lotus 123. It is assumed that any spreadsheet program can be used, however some of the particular algorithms may be slightly different than indicated in this annex.

A4.1.2 *Uniform Level ITP Design, n = 2*—All laboratories in the ITP test all materials; each material has $n = 2$ replicates per cell and the summations are over all laboratories.

$$T_1 = \Sigma\, Y_{AV} \tag{A4.1}$$

where:

Y_{AV} = cell average for laboratory (i).

$$T_2 = \Sigma\, (Y_{AV})^2 \tag{A4.2}$$

$$T_3 = \Sigma\, (w)^2 \tag{A4.3}$$

where:

w = range of cell values, laboratory (i).

(for $n = 2$ only)

$$T_4 = \Sigma\, (S)^2 \tag{A4.4}$$

where:

S = cell standard deviation, laboratory (i).

For the calculations as outlined as follows use either T_3 or T_4.

$$S^2_r = T_3/2p = T_4/p \tag{A4.5}$$

$$S^2_L = \{[p\, T_2 - (T_1)^2]/p\, (p - 1)\} - [S^2_r/2] \tag{A4.6}$$

$$S^2_R = S^2_L + S^2_r \tag{A4.7}$$

$$M_{AV} = T_1/p, \text{ material average for all laboratories} \tag{A4.8}$$

$$r = 2.83\, (S^2_r)^{1/2} = \text{repeatability} \tag{A4.9}$$

$$R = 2.83\, (S^2_R)^{1/2} = \text{reproducibility} \tag{A4.10}$$

A4.1.3 For any ITP with n equal to more than two but with a constant number of cell replications for each material-laboratory combination, the computation equations are identical to Eq A4.1-A4.10 with the following exceptions: (*1*) the value of n is used in place of 2 in the last term of Eq A4.6, and

(2) T_3 is not calculated, the value for S^2_r is obtained by means of the T_4/p expression in Eq A4.5.

A4.1.4 For any ITP with an unequal number of replicates per cell:

$$T_5 = \Sigma\,[n_i\,(Y_{AV})_i],\ n_i = \text{number of replicates in cell } i \quad \text{(A4.11)}$$

$$T_6 = \Sigma\,(n_i)\,(Y_{AV})^2{}_i \quad \text{(A4.12)}$$

$$T_7 = \Sigma\,(n_i) \quad \text{(A4.13)}$$

$$T_8 = \Sigma\,(n_i)^2 \quad \text{(A4.14)}$$

$$T_9 = \Sigma\,(n_i - 1)\,(S^2{}_i) \quad \text{(A4.15)}$$

where:

$S^2{}_i$ = variance for cell i.

$$S^2{}_r = T_9/(T_7 - p) \quad \text{(A4.16)}$$

$$S^2{}_L = \{\,|\,[T_6T_7 - (T_5)^2]/[T_7(p-1)]\,|\, - S^2{}_r\}\,\{[T_7\,(p-1)]/[(T_7)^2 - T_8]\} \quad \text{(A4.17)}$$

$$S^2{}_R = S^2{}_L + S^2{}_r \quad \text{(A4.18)}$$

$$M_{AV} = T_5/T_7 \quad \text{(A4.19)}$$

Calculate r and R as in Eq A4.9 and A4.10.

A4.2 Table Layout for Spreadsheet Calculations

A4.2.1 *Table Organization*—This section contains a listing of all the tables required with a brief description of the linking between the tables to permit all calculations to be automatically performed to give the values for r and R, once all tables have been set up and the basic table of data has been generated. The layout is for a uniform level design with $n = 2$. The description is directed mainly to Analysis Step 1. If outliers are found for Step 1, then the calculation operations of Step 2 and perhaps Step 3 will be required. For a full understanding of these two additional steps, it is necessary to completely review the precision evaluation example in Annex A6, which gives instructions for these additional calculations.

A4.2.2 For this annex, the tables will be identified as Table A4.1, Table A4.2, and so forth. Each of these is set up for a specific calculation. However, to avoid having blank tables (with the appropriate format as discussed in this annex) added to the length of the standard, the reader is referred to Annex A6. Annex A6 contains each table as discussed in Annex A4, filled in with data from the Mooney viscosity precision example. Therefore, when the set up for Table A4.1 format is discussed in this annex, refer to the corresponding table in Annex A6, which is Table A6.1 which gives both the table format and actual data. Starting with Table A4.1, the tables differ from the format of Tables 2 and 3 in the main body of this practice, in the use of a double or side-by-side data display format. This double table setup permits a quick view of the data and calculated parameters as data is entered and processed.

A4.2.3 There are potentially three analysis operation steps for any ITP. The number of steps actually required depends on the quality or uniformity of data in the database. If outliers are found, then a second and perhaps a third analysis step will be required. Each of these analysis operations should be conducted on a separate *sheet* or tabbed page of the computer spreadsheet program. This facilitates the analysis and avoids confusion. If outliers are found for any analysis operation, there are two options to continue with the analysis.

A4.2.3.1 *Outlier Option 1: Removal by Cell Deletion*—The simplest option for outliers is the deletion of the outlier from the database as expressed in a Table A4.1 format. See A4.3.2 for more details on this.

A4.2.3.2 *Outlier Option 2: Cell Replacement Values for Outliers*—If this option is chosen, cell replacement values are calculated by the procedures as described in Annex A5. This option involves more work but it may be the only option for a limited ITP database with a small number of laboratories.

A4.2.4 The three potential analysis steps are described in Sections 8 – 10. If there are no outliers, only Analysis Step 1 is used. If outliers are present, Analysis Steps 2 and 3 may be required depending on the extent of outliers in the database. The table description outlined as follows is for Analysis Step 1, the first set of calculations for any ITP, (see Section 8), prior to the possible rejection of any incompatible values as outliers.

A4.2.4.1 The word *cell* is used in two different contexts; it is the intersection of a row with a column in a computer spreadsheet, and it is also, for any ITP, the combination of a laboratory and a material as in Table 1 in the main body of this practice. The word cell will be italicized when it refers to a computer spreadsheet. In many cases there is a dual usage or meaning, a Table 1 cell is also a spreadsheet cell.

A4.2.4.2 Although described as follows, a Table A4.1 may contain blank table *cells*. All table cells that have data must contain the number of replicate values characteristic of the design of the ITP. For most General Precision ITP, $n = 2$ and each cell must contain both values. The original database generated in some ITPs may be one where one or more laboratories report only one value for a particular material, that is, they did not fully participate and only supplied partial data. The partial data for such a laboratory cannot be used since the spreadsheet program as set up in this annex requires that all Table A4.1 *cells* (for Analysis Step 1, 2, or 3) be uniform, that is, have the required number of replicates or no values at all.

Table Number and Name	Table Description
Table A4.1 Basic Data from ITP	This is the basic Table 1 format (as discussed in the main body of this practice); Rows = Laboratories; Columns in Replicate 1, 2 format = materials. Two spreadsheet columns are required for each material. Each (double column) ITP cell contains two test results. In generating all tables beyond Table A4.1, preserve the same row-column identification for laboratories and materials. Remember, go to Annex A6, Table A6.1, for an example of a Table A4.1 format, with actual data entered.

Table Number and Name	Table Description
Table A4.2 Cell Averages, Averages Squared	This is a dual table, cell averages in left side and cell averages squared in the right side, each side preserving the laboratory-material row versus column format of Table A4.1. Totals are calculated for each material column; cell average totals = T_1, cell average squared totals = T_2. Also calculate for the left section, the grand cell average (all laboratories), the variance, and standard deviation of the cell averages (across all laboratories) . Note—Do not truncate the significant figures for any total in any of these tables. Retain 4 significant digits for all calculations.
Table A4.3 Cell Average Deviations, *d* and *h*-values	A dual table, cell deviations *d*, *d* = cell (*i*) – (all cell avg); in the left section and cell *h*-values in the right section. Review the cell *h*-values and mark all that are significant at the 5 % level in some manner appropriate for the spreadsheet being used, such as making the value bold and italic, shaded, or using color. See Annex A3 for calculation of *h*-values.
Table A4.4R Cell Ranges and Ranges Squared	A dual table, cell ranges on left and cell ranges squared on the right. For each left-hand-side cell, the cell range may be obtained from Table A4.1. Use an appropriate spreadsheet function (such as @IF or ABS) to convert those negative difference values to positive values for the cells of Table A4.4R. It is useful to obtain the average range for each material. Calculate the cell squared totals T_3 for each material.
Table A4.4S Cell Standard Deviations and Variances	A dual table, with cell standard deviations on the left and cell variances on the right. It is convenient to calculate the pooled variance for each column of standard deviations. Place these at the bottom of each left-hand-side column. Calculate the total for the cell variances; place these values at bottom of each column of variances on the right side. Total of cell variances for each material = T_4.
Table A4.5 Cell *k*-values	A single table, cell *k* -values. See Annex A3 for calculation of *k*-values. For each *k*-value that equals or exceeds the 5 % significance level value, indicate by making the value bold and italic.
Table A4.6 Calculations for Precision	A table giving the sequence of calculations for precision. The calculations are performed for each material separately, thus a column is required for each material. Insert values for T_1, T_2, and either T_3 or T_4, by means of spreadsheet linking to the appropriate preceding tables. Calculation 1 is a calculation of $(Sr)^2$, using either T_3 or T_4. Calculation 2 evaluates $(S_L)^2$ using T_1 and T_2. Calculation 3 is a calculation of $(S_R)^2$, using $(S_L)^2$ and $(Sr)^2$. Calculation 4 evaluates *r* and Calculation 5 evaluates *R*. At the bottom of Table A4.6, material means (averages) are given as well as the standard deviations *Sr* and *SR*. Also listed is a sub-table for the Step 1 and if used, Step 2 or 3 outlier review at the 5 % and 2 % significance levels. This sub-table indicates the outlying laboratories for both *h* and *k*. Note: The values for *n* and *p* in Table A4.6 can either be active or be a fill-in format. The value of *n* will be 2, but *p* will vary depending on the number of cells for laboratories deleted for either *h* or *k* values. For active *p* values, a count function should be performed for the cell values in Table A4.5-R1-OD or A4.5-R2-OD, see A4.3.1, for each material. This counts the number of laboratories after both *h* and *k* deletions. The count result enters the appropriate cell of Table A4.6. For a fill-in operation, the values in Table A4.6 must be inserted manually.

A4.2.5 *Setting up the Spreadsheet*—Begin on Sheet 1 of a spreadsheet program. This will be used for Analysis Step 1. The first set of calculations is for the original database. For any subsequent analysis operations with a complete set of recalculations after outliers are removed from the database or outliers replaced, one or more additional computer program sheets will be used. Calculations are facilitated if each table occupies a single screen area, using the *page down* command to go to the next table. Refer to Annex A6 for more details on Steps 2 and 3.

A4.2.5.1 *Link Table A4.2 to A4.1*—For Laboratory 1 and Material 1, use the appropriated spreadsheet average function (such as an @function or AVERAGE) to calculate the average for *Cell* 1 in Table A4.2, using the corresponding two adjacent (spreadsheet) *cells* on Row 1 of Table A4.1, for Laboratory 1 and Material 1, as the argument spreadsheet range. Repeat for all table *cells*. After this is completed, calculate the cell average squared values for all *cells* on the right side of Table A4.2 by the appropriate spreadsheet squared function algorithm using the left-hand-side *cell* averages.

A4.2.5.2 *Link Table A4.3 to A4.2*—For Material 1, using the appropriate spreadsheet algorithm, subtract from each laboratory *cell* average in the left-hand-side of Table A4.2 the overall *cell* average. This gives *d*. Divide each calculated *d* by the standard deviation of all *cell* averages to give the calculated *h*-value. Repeat for all materials. The calculation output for *h*-values is entered into the corresponding (row-column) *cell* in the right-hand-side section of Table A4.3.

A4.2.5.3 *Link Table A4.4 to Table A4.1*—For Laboratory 1 and Material 1, calculate the standard deviation for *Cell* 1 in Table A4.4, by means of the appropriate spreadsheet function for standard deviation, using the corresponding two adjacent

cells on Row 1 of Table A4.1 (Laboratory 1 and Material 1), as the spreadsheet argument range. Repeat for all materials or *cells*. Ensure that the divisor for standard deviation calculation is $(n - 1)$, not n, where n = number of values for standard deviation calculation for each material. In spreadsheet terminology, this is often designated as a *sample standard deviation* calculation. Using the appropriate algorithm, square each *cell* standard deviation value; the result is entered into the corresponding *cell* on the variance or the right side of Table A4.4.

A4.2.5.4 *Link Table A4.5 to A4.4S*—For Material 1, divide each individual (within) cell standard deviation, by the pooled value for (within) cell standard deviations (this is the square root of the pooled or mean variance) to obtain *k*-values. Repeat for all materials. The *k*-values are entered into the corresponding *cells* in Table A4.5.

A4.2.5.5 *Link Table A4.6 to Tables A4.2, A4.4S, or A4.4R, or Combination Thereof*—For Material 1, use the appropriate spreadsheet function or algorithm to bring the totals T_1, T_2, T_3, or T_4, or combination thereof, into Table A4.6. Repeat this for all materials. The source for each total should be the total at the bottom of each of the appropriate columns in Tables A4.2, A4.4S, or A4.4R. For Calculation 1 in Table A4.6, use the formula given in the table to calculate each of the parameters for all materials in the ITP. The formula should use the active values for *n* and *p* as well as values for that material as brought in from Tables A4.2, A4.4S, or A4.4R. When Calculation 5 of Table A4.6 is completed, the entry of values for T_1, T_2, T_3, or T_4, or combination thereof, along with values for *p* and *n* (by means of their linkages to preceding tables) will produce an immediate result for all intermediate and final precision calculations in the table.

A4.3 Sequence of Database Calculations for Precision

A4.3.1 *Outliers in Analysis Step 1 (Sheet 1)*—As previously noted, the Step 1 analysis operation or set of calculations should be performed on Sheet 1 of the computer spreadsheet program. If any incompatible values are declared as outliers at the 5 % significance level, the database shall be revised according to 8.4 to either delete outliers for any laboratory or to insert replacements into the database for those *cells* that contain outliers. If any outliers are found, it is necessary to conduct Analysis Step 2 (Sheet 2) on the *R1* database. The calculations for analysis of the *R1* database are facilitated by copying all of the executed Tables A4.1 to A4.6 on Sheet 1, onto corresponding locations in Sheet 2 of the spreadsheet, with all programmed calculations active, that is, not as values or copying Sheet 1 and renaming it as Sheet 2. These tables on Sheet 2 are now designated as (*1*) Tables A4.1-R1-OR to Table A4.6-R1-OR for replaced outliers or (*2*) Tables A4.1-R1-OD to Table A4.6-R1-OD for deleted outliers.

A4.3.2 *Outliers in Analysis Step 2 (Sheet 2): Option 1 Outlier Deletion*—All deletion operations can be facilitated by marking on a printed out Table A4.1, all table *cells* that have significant *h* and *k* values. To delete data, simply delete from Table A4.1 all the *cells* that have a 5 % significance level *h* or *k* value; that is, delete both values in each ITP design cell, which occupy two spreadsheet cells. When this is done, the typical spreadsheet program will give some *ERROR* indication at several calculation *cell* locations in Tables A4.2-R1-OD to Table A4.6-R1-OD. (*ERROR* is used generically in the following text to indicate the specific spreadsheet error flag.) This is due to the deletion of one or more argument values in Table A4.1-R1-OD and some subsequent tables as well.

A4.3.2.1 *Correcting the ERROR Cells—ERROR* notations will appear in two general locations (*1*) in columns as data entries that come from tables above them in the sequence of tables, that is, values used to calculate parameters for a particular column such as averages, standard deviations, and so forth, and (*2*) at the bottom of columns where averages, standard deviations, and so forth were previously located. To correct the tables, start with the first table that contains a spreadsheet cell that has an *ERROR* notation, and delete the *ERROR cell* that is a data entry, not an *ERROR cell* at the base of a column. Correcting the data entry value or *cell* will automatically correct the *ERROR* (calculated value) at the base of the column.

A4.3.2.2 The use of a spreadsheet *delete* operation for any *ERROR cell* will make the *cell* in question blank. Continue this for all tables until all *ERROR* indications are removed and replaced by blank values, not zeros. This will produce correct calculations for all parameters. Also remove from all tables any zero *cell* values that are generated by the deletions from any of the preceding tables. If they are not removed, the bottom of the table column calculations will be in error. For Option 1, outlier deletion, the revised precision parameters will automatically be calculated and appear in Table A4.6-R1-OD of Sheet 2, after all *ERROR* entries are removed.

A4.3.3 *Outliers in Analysis Step 2 (Sheet 2): Option 2 Outlier Replacement*—When this option is chosen, replacement values are inserted into the *cells* that contain outliers. Insert into the experimental design cells of Table A4.1 (individual) *cell* data replacement values or DRVs, as evaluated in Annex A5. These will be in cells that have a significant *h* or *k* value. Correct any possible *ERROR* occurrences, if they appear, as described in A4.3.2.1 and A4.3.2.2. For Option 2, insertion of DRVs, the revised precision parameters will automatically be calculated and appear in Table A4.6-R1-OR of Sheet 2.

A4.3.4 *Outliers in Analysis Step 3 (Sheet 3)*—The precision values for (Sheet 2) *R1* analysis are accepted as final if there are no outliers at the 2 % significance level.

A4.3.4.1 If any outliers are found at the 2 % significance level, the procedure as previously cited (for 5 % significance) is followed to either do a Option 1 deletion of all outliers to generate a *R2* OD database or select Option 2 and calculate replacement values. When these are inserted into the *R1* OR database, a *R2* OR database is generated.

A4.3.4.2 If outliers are found, copy the executed Tables A4.1-R1-OR to A4.6-R1-OR or Tables A4.1-R1-OD to A4.6-R1-OD, of spreadsheet Sheet 2 to spreadsheet Sheet 3 with active values as above or copy Sheet 2 and rename as Sheet 3. These *R2* tables, when completed as indicated as follows, will be designated as Table A4.1-R2-OR to Table A4.6-R2-OR or the corresponding Table A4.1-R2-OD to Table A4.6-R2-OD. The purpose of a Sheet 3 analysis is to

delete or replace the 2 % significance outliers and thereby generate final *R2* precision values.

A4.3.4.3 Once outlier values have been deleted from any *cell* or DRVs have been calculated (using Annex A5) and inserted into the appropriate *cells* of Table A4.1-R2-OR or A4.1-R2-OD in Sheet 3, the new precision values will appear in Sheet 3 Table A4.6-R2-OR or Table A4.6-R2-OD after any *ERROR* indications are removed. These Sheet 3 Table A4.6-R2-OR or Table A4.6-R2-OD values are the final precision parameters, *r* and *R* for the ITP.

A4.3.5 *Precision Result Rounding*—The final precision results as given in Table A4.6, Table A4.6-R1, or Table A4.6-R2 (with either outlier option) are transferred into a Table 6 format (see 12.1) for insertion into the test method. When this is done, the final precision parameters should be rounded to the number of significant digits or figures that are technically attainable in usual practice with the test method, with perhaps one more significant figure than normally employed.

A5. PROCEDURE FOR CALCULATING REPLACEMENT VALUES FOR DELETED OUTLIERS

A5.1 Introduction

A5.1.1 If outliers are found in Analysis Step 1 at the 5 % significance level, there are two options. Option 1 is to delete the outliers and thereby generate a revised or *R1* database. Option 2 is to replace the outliers in a way that essentially preserves the distribution of the non-outlier data as described in more detail in A5.2. This annex provides the algorithms to address the replacement process when outliers are found at either the 5 % or 2 % significance level.

A5.1.2 Outlier Option 2 (replacement) is usually the choice when outliers are found with a small database with a limited number of laboratories (approximately six or less). Replacing outlier values, rather than deleting them, preserves the size of the database. The procedure to calculate replacement values however must be one that is *consistent with the observed data distribution* in the database. The replacement procedure as described in this annex fulfills this objective. The procedure consists of the evaluation or calculation of two types of replacements.

A5.2 The Replacement Procedure

A5.2.1 The replacement procedure (for either Step 1 or 2) is one that replaces outliers with realistic values. The initial operation evaluates replacement values for each outlier *cell average* and each outlier *cell standard deviation*. The first type of replacement is designated as a parameter replacement value, or PRV. There are two possible types of PRVs described as follows that might be inserted into the database. Although only one is selected, both are described in order to demonstrate the merit of the selected second type of replacement.

A5.2.2 *Distribution Mean Parameter Replacement*—The first possible approach for a PRV is to insert into the database a value equal to the distribution or actual database mean for all cell values for any material. There are two types of distribution means (*1*) for cell averages or (*2*) for cell standard deviations or cell ranges. The word *mean* applies to both. If only one PRV is being considered and there are ten or more laboratories, this will not substantially change the nature of the distribution. However, if two or more outliers are being replaced and the number of laboratories is much less than ten, this may narrow the distribution and therefore give a falsely optimistic standard deviation for (*1*) the final precision results (if no further outliers are found) or (*2*) for denominator standard deviation for the *h* or *k* statistics, or both, that will be used for outlier review at the 2 % significance level. For this reason, this type of replacement is not chosen.

A5.2.3 *Ascending Order Trend (AOT) Parameter Replacement*—The alternative approach for a PRV is to use a value that substantially preserves the observed distribution as illustrated by the ascending order trend plots as discussed in 8.1.3. This is designated as an ascending order trend or AOT replacement or PRV for a cell mean. Each AOT replacement or PRV is in essence a predicted value; one that would be expected for the laboratory in question, absent the unexpected perturbation that generated the outlier illustrated by the *off-the-line* behavior in the AOT plot. This AOT replacement does not narrow the observed distribution in the same sense as a distribution *mean* value replacement.

A5.2.4 *Outlier Replacement Categories*—There are two different categories for outlier replacements, *parameter* replacements or PRVs as previously discussed, and a *data replacement value* or DRV. After PRVs have been determined for all outlier cell averages and cell standard deviations (or ranges), the next step is the calculation of DRVs for each cell of Table A4.1 format that contained a *parameter* outlier.

A5.2.4.1 The *DRVs* are required to insert into a Table A4.1 data format (to generate a Table A4.1-R1-OR) to permit a recalculation of the revised precision values based on the new *R1* database. See Annex A4 and the Table A4.1 to A4.6 series. Once the initial basic data Table A4.1 is revised to generate a Table A4.1-R1-OR, all the succeeding tables, A4.2-R1-OR to A4.6-R1-OR, are recalculated by the automatic calculation process as described in Annex A4. The procedures as described (for this Annex A5) are for uniform level designs with two cell values or $n = 2$. The procedures may be slightly amended for $n = 3$ situations. The precision example in Annex A6 on Mooney viscosity testing illustrates the entire AOT replacement process and the operations described in this annex as well as Annex A3 and Annex A4.

A5.3 *PRVs for Outliers at 5 % Significance Level*—Outlier values at the 5 % significance level shall be replaced using the AOT replacement procedure as described in A5.3.1 – A5.3.3.

These procedures apply in principle to any of three databases: the original database, the *R1* database, or the *R2* database. The *R1* and *R2* databases will potentially contain PRVs as determined by a previous outlier replacement process.

A5.3.1 *PRVs: Cell Average Outliers*—For each material, visually fit a (least squares type) straight line through the central data point region of the cell average AOT plot and extend the line to both extreme ends of the plot. Alternatively, a linear regression may be used to fit the straight line, however, do not include in the data set any questionable outlier end points. For the outlier values (low or high end of plot), determine the difference between the outlier value (plotted point) and that point on the extended line at the x-axis location of the laboratory in question. Add or subtract this difference to the outlier value to produce a new value that is *on the fitted line* at that *x*-axis location. For each outlier, this *on the line* value is the cell average PRV for that laboratory.

A5.3.2 *PRVs: Cell Range Outliers*—For each material, visually fit a straight line through the central value point region of the cell range AOT plot and extend the line to the high value end of the plot. Repeat the procedure as cited in A5.3.1 to evaluate a new value on the fitted line. For each outlier, this *on the line* value is the cell range PRV for that laboratory.

A5.3.3 *PRVs: Cell Standard Deviation Outliers*—If cell standard deviations were calculated initially rather than cell ranges, evaluate a standard deviation PRV using the same procedure as described for cell range outliers in A5.3.2. For ITP designs that have $n = 2$, the replacement cell standard deviation (SDev) can be converted to a cell range, w, by using: $w = (\text{Sdev})\ (2)^{1/2}$. In the following equations, a value for the range is required for calculating DRVs.

NOTE A5.1—The equations to calculate DRVs using PRVs for ranges as given in A5.4 can be altered for use with standard deviations rather than ranges. For ITP where $n = 2$, substitute the value for the range w, that is, (SDev)*1.414, into the equations.

A5.4 *DRVs for Outliers at 5 % Significance Level*—After PRVs have been determined for all outlier cell averages and cell standard deviations (or ranges) at the 5 % significance level, the next step is the calculation of DRVs for insertion into a Table A4.1 format. For the DRV process, procedures are used that maintain the values not declared as outliers at their observed values in the database. As an example, when only a replacement cell average is required, (that is, the cell range is not an outlier), the actual or existing cell range shall not be changed by the replacement. Also, when only a replacement cell range is required, the existing cell average shall be maintained. There are four possible combinations of PRVs that require DRVs. The procedures for these are described in A5.4.1 – A5.4.4.

A5.4.1 *Cell Average Outlier with Non-Outlier Cell Range*—For the two DRVs for a cell average outlier, add one half and subtract one half of the original or existing cell range, ECR, to and from the PRV (cell average), as obtained in A5.3.1, using Eq A5.1 and A5.2. This gives two cell values, *DRV*1 and *DRV*2 that yield the replacement cell average. Insert the replacement values into the Table A4.1 format database.

$$DRV1 = PRV(\text{cell average}) + ECR/2 \qquad (A5.1)$$

$$DRV2 = PRV(\text{cell average}) - ECR/2 \qquad (A5.2)$$

To avoid the confusion of excessive notation, all DRVs (each of four categories) are identified as *DRV*1 and *DRV*2.

A5.4.2 *Cell Average Outlier with Cell Range Outlier*—For the two DRVs for this situation, add one half and subtract one half of the AOT plot evaluated PRV(cell range), as obtained in A5.3.2, to and from the PRV(cell average) as obtained in A5.3.1, using Eq A5.3 and A5.4. This gives the two new cell data values, *DRV*1 and *DRV* 2, that yield the replacement cell average and the replacement cell range. Insert the DRVs into the Table A4.1 format database.

$$DRV1 = PRV(\text{cell average}) + PRV(\text{cell range})/2 \qquad (A5.3)$$

$$DRV1 = PRV(\text{cell average}) - PRV(\text{cell range})/2 \qquad (A5.4)$$

A5.4.3 *Cell Range Outlier with Non-Outlier Cell Average*—For the two DRVs required for this situation, add one half and subtract one half of the AOT evaluated PRV(cell range) as obtained in A5.3.2, to and from the original or existing cell average, ECA, using Eq A5.5 and A5.6. This gives the two new cell data values, *DRV*1 and *DRV*2, that yield the original cell average and the replacement cell range. Insert these into the Table A4.1 format database.

$$DRV1 = ECA + PRV(\text{cell range})/2 \qquad (A5.5)$$

$$DRV2 = ECA - PRV(\text{cell range})/2 \qquad (A5.6)$$

A5.4.4 *Cell Range Outlier with Cell Average Outlier*—Follow the same procedure as in A5.4.2. This gives two cell data values with the replacement cell average and the replacement cell range. Insert these into the Table A4.1 format database.

A5.5 *PRVs for Outliers at 2 % Significance Level*—For an Analysis Step 2 review of the revised or *R1* database, follow the instructions of A5.5 and A5.6 that apply to a significance level of 2 %.

A5.5.1 *PRVs: Cell Average Outliers*—For each material, replot the cell average data to give a new AOT plot, using the revised data of Table A4.1-R1-OR. The data in the Table A4.1-R1-OR format will have new replacement values for all 5 % significance outliers. Follow the procedure as described in A5.3.1 to determine the PRV *cell average* for outliers at the 2 % significance level.

A5.5.2 *PRVs: Cell Range Outliers*—For each material, replot the cell range data in an AOT plot, using the revised data of Table A4.1-R1-OR. Follow the procedure as described in A5.3.2 to determine the PRV *cell range* for outliers at the 2 % significance level.

A5.5.3 *PRVs: Cell Standard Deviation Outliers*—If cell standard deviations were calculated initially rather than cell ranges, calculate a replacement standard deviation using the cell range procedure as described in A5.5.2. As previously noted, for ITP designs with $n = 2$, the replacement cell standard deviation (SDev) can be converted to a cell range, w, by using: $w = (\text{Sdev})\ (2)^{1/2}$.

A5.6 *DRVs for Outliers at 2 % Significance Level*—After PRVs have been determined for all outlier cell averages and

cell standard deviations (or ranges), at the 2 % significance level, the next operation is the calculation of DRVs for Table A4.1 format. These are required to generate a Table A4.1-R2-OR format, to permit a recalculation of the revised precision values (repeatability and reproducibility) based on the new *R2* database. See Annex A4. Just as for the 5 % significance level calculations, there are four possible combinations of parameter outliers that require data replacements for a *R2* database. For A5.6.1 to A5.6.4, the outliers are at the 2 % significance levels and the database being considered for revision is the *R1* database. After 2 % significance level outliers have been replaced (both PRVs and DRVs) for a *R1* database, it becomes a *R2* database and is used to calculate the final or terminal values for repeatability and reproducibility. Refer to the flow sheet diagram in Fig. 1.

A5.6.1 For the four outlier combination categories as discussed in A5.4.1 – A5.4.4, repeat the calculations for DRVs based on evaluated PRVs using AOT plots of the *R1* database. Use the equations as cited in these sections.

A6. AN EXAMPLE OF GENERAL PRECISION EVALUATION—MOONEY VISCOSITY TESTING

A6.1 Introduction

A6.1.1 This annex presents a detailed example of the Three-Step Analysis General Precision evaluation with emphasis on how outliers are detected and how the original database is revised to obtain robust precision estimates that are free of outlier effects. All precision calculations are given, starting with a basic Table 1 (or equivalent Table A4.1) format, using the calculation formulas and other operations in the series of tables as described in Annex A4. Most of the table in this annex use a two part identification system; first a sequential table number starting with Table A6.1 and a second identification set of symbols in parenthesis that indicate the purpose of the table. The sequential number is required for computer preparation of the standard and the second identification symbol set permits a better comprehension of the context and use of each of the tables. There is a connection between the tables of Annex A4 and of the tables of this annex in terms of their context and use. This second set of symbols inside the parenthesis indicates this connection between the two annexes. Therefore the first Table A6.1 (1) of this annex is equivalent to Table A4.1 in Annex A4, and the second Table A6.2 (2) is equivalent to Table A4.2 of Annex A4, and so forth for all tables with identification symbols (3), (4R), (4S), (5) and (6). Each of the tables in the sequence (1) to (6) performs a unique function in the calculation operation. There are four final tables in this annex that are not part of the Annex A4 - Annex A6 connection and do not use this two part identification system, i.e.. Tables A6.36 to A6.39. Note that Annex A4 does not have this two part table identifying system since in this standard no Annex A4 tables have been generated. The Annex A4 table designations are specified for the user of the standard to employ in setting up a spreadsheet for any actual analysis operation.

A6.1.2 Two outlier treatment options may be chosen after outliers are detected. Option 1 is the deletion of all outliers and the calculation of precision results on the revised and reduced database. Option 2 is the replacement of outliers with AOT replacements (PRV, DRV) and the calculation of precision results on the revised database. For purposes of illustration, both of these options are given in this example. An additional feature is illustrated, the use of technical judgment by the statistical analyst to override the outcome of a particular objective outlier rejection procedure. The reasons for this are cited.

A6.1.3 The ITP for Mooney Viscosity Testing was conducted in 1982 using the version of the ASTM standard for Mooney viscosity testing, Test Methods D1646, that existed at that time. Test Methods D1646 is equivalent to ISO 289. Four materials (rubbers) were used and nine laboratories participated in the ITP. The rubbers, identified as Materials 1 to 4, and some of the details of the testing are described as follows.

Material Number	Material Description	Test Conditions
1	SBR1712 (37.5 oil ext)	ML 1+4 at 100°C
2	IIR (Butyl) NIST SRM 388	ML 1+8 at 100°C
3	SBR1712 BMB (37.5, 65 N339)	ML 1+4 at 100°C
4	NR (natural rubber)	ML 1+4 at 100°C

NIST = National Institute of Standards and Technology, the new name for the National Bureau of Standards
SRM = Standard Reference Material as developed by NIST
BMB = Black Masterbatch, 37.5 Oil + 65 of carbon black N339

A6.1.4 Samples of each of the four materials were sent out to the nine participating laboratories, and viscosity tests were conducted on two separate days one week apart. A test result is one determination (measurement) of Mooney viscosity at the indicated time and temperature. Therefore for this ITP, $p = 9$, $q = 4$ and $n = 2$. A Type 1 precision was evaluated with one additional operation just prior to testing; Materials 1, 3, and 4, were mill-massed as specified in Section 7 of the 1982 version of Test Methods D1646. Material 2, the IIR, an SRM, was not mill-massed since this was not specified in Test Methods D1646 for this reference material.

A6.1.5 *Organization of the Mooney Example Precision Evaluation*—The ordinary practice to evaluate precision for any given ITP, is to use the sequence of steps as outlined in Fig. 1 and discussed in the overview Section 7. The detailed instructions are in Sections 8 – 10. If outliers are found for Step 1, one of the two outlier options is selected and the analysis proceeds to Step 2 and on to Step 3 if needed based on this decision, see again Fig. 1. However to better illustrate precision evaluation in this example, calculations are given for both outlier options. Although outlier replacement is Option 2, the

calculations for this option will be demonstrated first as Part 1. After that, the simpler Option 1 approach of outlier deletion will be demonstrated as Part 2. The preliminary data and graphical review, given in A6.2.1, is not repeated for the Part 2 outlier deletion option.

NOTE A6.1—Eq A2.16 and Eq A2.17 use a multiplier value of 2.83. A rounded value of 2.8 was used to calculate the values in Tables A6.7 (6), A6.14 (6-R1-OR), A6.21 (6-R2-OR), A6.28 (6-R1-OD), and A6.35 (6-R2-OD). If the data in Annex A6 are used to validate a spreadsheet and a multiplier of 2.83 is used in the spreadsheet formulas, the resulting values will be slightly different from those shown in the aforementioned tables.

A6.2 Part 1: Outlier Replacement—Analysis Step 1

A6.2.1 *Preliminary Review*—Table A6.1, as set up in Sheet 1 of the computer spreadsheet program (see Annex A4), is a tabulation of the original data in a format as specified in 8.1.1 and 8.1.2. Although it is not necessary for the analysis steps to follow, it is informative to obtain averages and standard deviations of all columns in the table and the results for these calculations are illustrated.

A6.2.1.1 The next operation is to generate Tables 2 and 3. To avoid unnecessary redundant tables, the basic Tables 2 and 3 data tabulation is combined with other tabulations and calculations in a dual-table format. This dual-table format is required for the full analysis and is fully described in Annex A4. Therefore, the Table 2 format is given in the left side of Table A6.2 and the Table 3 data tabulation format is given in the left side of Table A6.4S, for within cell standard deviations or in Table A6.4R, for within cell ranges.

A6.2.1.2 The graphical examination of the ITP data is conducted using Figs. A6.1-A6.4 and Fig. A6.5. Fig. A6.1 illustrates plots of *cell average* Mooney viscosity versus laboratory number in ascending viscosity order for Materials 1 and 2 and Fig. A6.2 illustrates similar plots for Materials 3 and 4. These plots serve a dual purpose: an initial review of the original data and a second operation to calculate the Outlier Option 2 AOT replacement values for outliers as described in A5.2.2 in Annex A5.

A6.2.1.3 Fig. A6.1 indicates that there may be two potential outliers for Material 1, one low outlier for Laboratory 9 and perhaps a high outlier for Laboratory 6. These deviate from the central region linear trend line. This line will be used in the AOT replacement operation to be conducted later. For Material 2, one high potential outlier for Laboratory 1 is indicated. In Fig. A6.2, Material 3 has one low potential outlier for Laboratory 9 and Material 4 has two potential outliers, low for Laboratory 9 with a less likely high value for Laboratory 8.

A6.2.1.4 Similar plots for cell ranges in Figs. A6.3 and A6.4 are slightly different than the cell average plots. There are no low-end outliers. All low values indicate good agreement, and as a result, these plots have a low-end curvilinear nature prior to the central linear region. This is ignored in drawing the trend lines. Material 1 has two potential high-end cell range outliers for Laboratories 1 and 4 . Material 2 has no potential outliers. Materials 3 and 4 in Fig. A6.4 both have potential outliers for Laboratory 4 and perhaps one for Laboratory 9. The plots give an overall impression of the degree of data uniformity for each of the four materials. The other features will be discussed later.

A6.2.2 *Precision Calculations and Outlier Review for Original Database*—The basic Step 1 Analysis operation begins by calculating the precision values for r and R for the original database. The initial calculation for r and R using the procedures as set forth in Annex A4, establishes a foundation for comparisons of the reduction in these two parameters as outliers are deleted. The next operation is an examination of the database to detect any potential outliers at the 5 % significance level. Both of these operations will be conducted in parallel and described as each table in the sequence Table A6.1 (1) to Table A6.6 (7) is reviewed.

A6.2.2.1 Table A6.2 (2), set up in the dual format for all four materials, has cell averages on the left and cell averages squared on the right. Two totals, T_1 for *cell averages* and T_2 for *cell averages squared* (as required for final precision analysis, see Table A6.6 (7)), are obtained for each column or material in the table. Also indicated are results for the overall cell average, variance, and standard deviation for individual cell averages for all nine laboratories.

A6.2.2.2 Table A6.3 (3) contains the *cell average* deviations, d, on the left and the cell h-values on the right, where for each material:

$$d = (Y_{AV}(i) - Y_{AV}) \quad \text{(A6.1)}$$

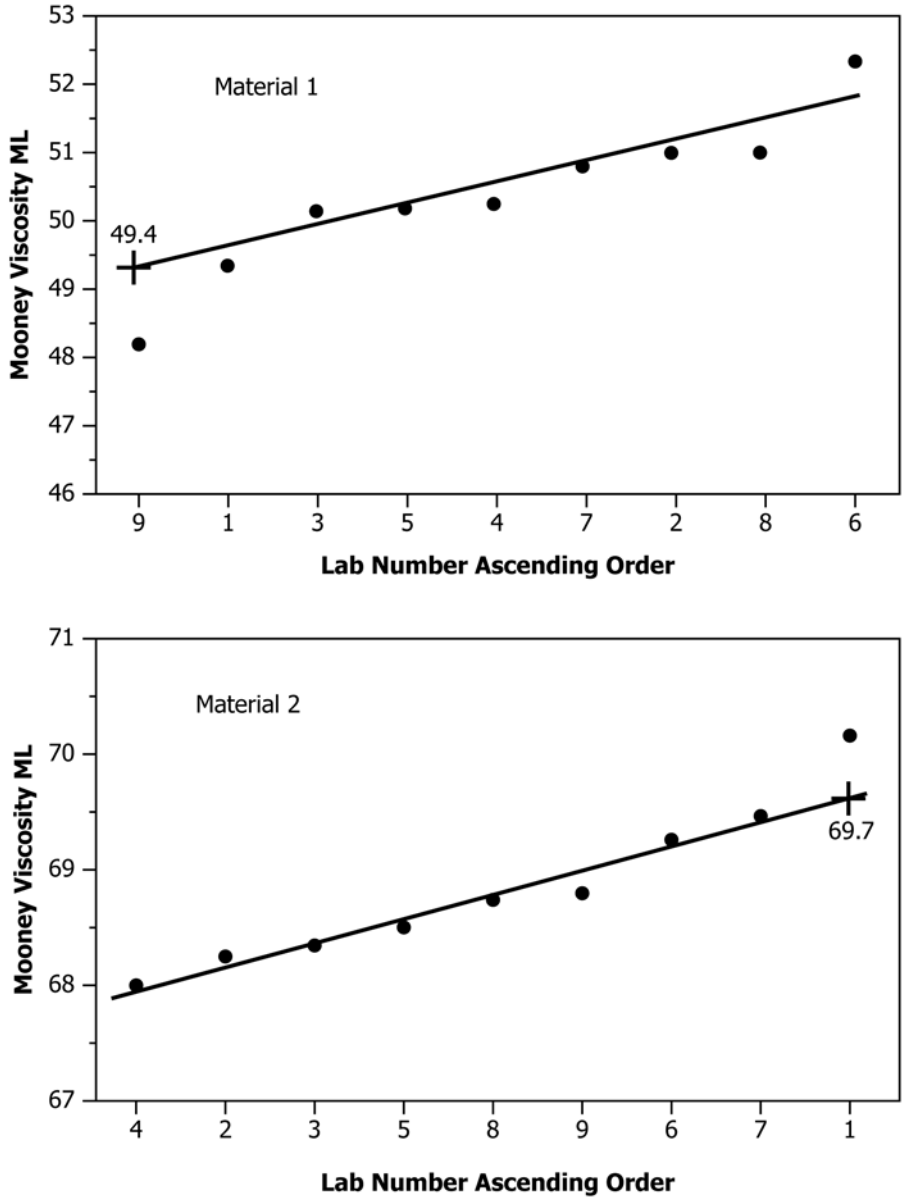
$$h = d/S_{(YAV)} \quad \text{(A6.2)}$$

where:

$Y_{AV}(i)$ = cell (i) average,
Y_{AV} = average of all cell averages, and
$S_{(YAV)}$ = standard deviation of cell averages, see Annex A3.

The values for Y_{AV} and $S_{(YAV)}$, descriptively indicated, are found at the bottom of the left section of Table A6.3 (3). Below the right side of the table, is an inset sub-table that gives the h(crit) at the 5 % significance level for the indicated number of laboratories, that is, p = 9. Critical values for both h and k are given in Table A3.1 of Annex A3. The calculated column h-values (for each material) that equal or exceed the critical value 1.78, have a bold-italic indication. There are four cells with significant h-values: Laboratory 1, Material 2, and Laboratory 9, Materials 1, 3 and 4.

A6.2.2.3 Table A6.4 (4R) and A6.5 (4S) indicate the dispersion (variation) for the day-1 versus day-2 test results. Actually only one of these two tables is absolutely needed, but both have been generated for this example. Table A6.4R contains the *within cell* ranges on the left and the cell ranges squared on the right. For each material, the *cell range* squared total, T_3, is given. Cell ranges for an ITP program with n = 2 may be converted into standard deviations by; SDev = w / $(2)^{1/2}$, where w is the range. Table A6.5 (4S) is next, it has *within cell* standard deviations on the left and variances (standard deviations squared) on the right. On the right side, the total of all variances, T_4, as well as the pooled or average variance is given for each material.



NOTE 1—With linear trend line and PRV indicated.

FIG. A6.1 AOT Plots—Original Cell Averages for Materials 1 and 2

A6.2.2.4 The analysis for outliers for cell standard deviations is conducted by means of Table A6.6 (5), the tabulation of the *k*-values for all cells for each material is generated using:

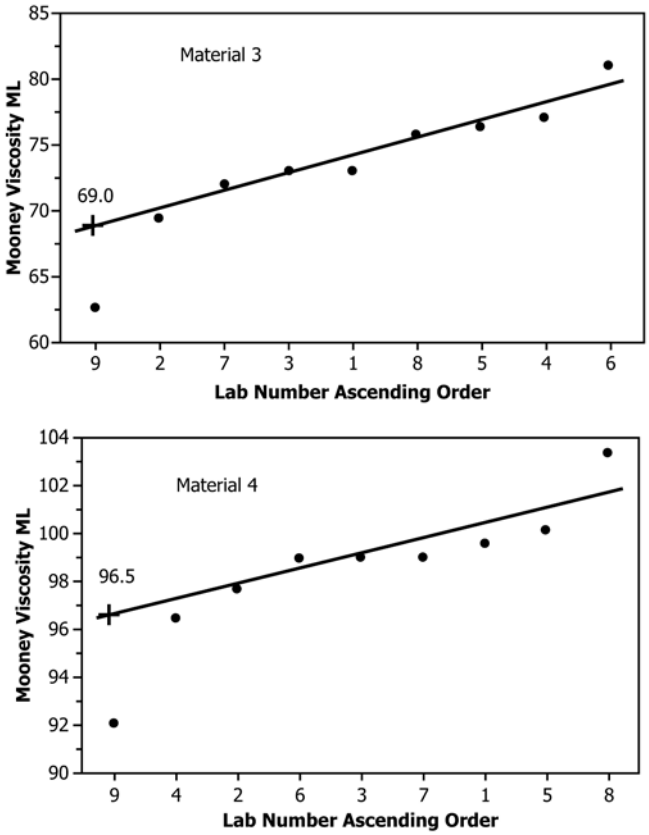
$$k = S(i)/S_r \quad \text{(A6.3)}$$

where:

$S(i)$ = cell standard deviation for Laboratory i, and
S_r = pooled cell standard deviation (across all laboratories), see Annex A3.

The pooled standard deviations (square root of pooled or average variance) are given at the bottom of both Table A6.5 (4S) and Table A6.6 (5). Part of Table A6.6 (5) is an inset sub-table that gives *k*(crit) at the 5 % significance level for p = 9 and n = 2. There are three calculated *k*-values equal to or above the critical value of 1.90, Materials 1, 3 and 4 for Laboratory 4. These cells have a bold italic indication.

A6.2.2.5 This completes Analysis Step 1. Before proceeding to Step 2 it is informative to consult Table A6.7 (6), the precision results for the original database. The *r* values span the interval from 0.74 to 3.43 and the *R* values from 1.97 to 15.15. If no outliers had been detected in the Step 1 analysis, this table would constitute the end of the analysis and the values as they appear in Table A6.7 (6) would be used to prepare a final table of precision results for entry into the test method. In addition to the five internal calculations of Table A6.7 (6) to give the final values for *r* and *R*, the table also gives, at the bottom, the mean value for each material as well as the repeatability standard deviation *Sr* and the reproducibility standard deviation *SR* and



NOTE 1—With linear trend line and PRV indicated.

FIG. A6.2 AOT Plots—Original Cell Average for Materials 3 and 4

values for (*r*) and (*R*), the relative precision in percent of the mean for each material. The results of the Step 1 outlier analysis for the *h* and *k* statistics are given in a sub-table at the bottom of Table A6.7 (6). The Step 1 outlier analysis has indicated a number of outliers at the 5 % significance level. The presence of these outliers calls for a Step 2 analysis operation on a revised ITP database.

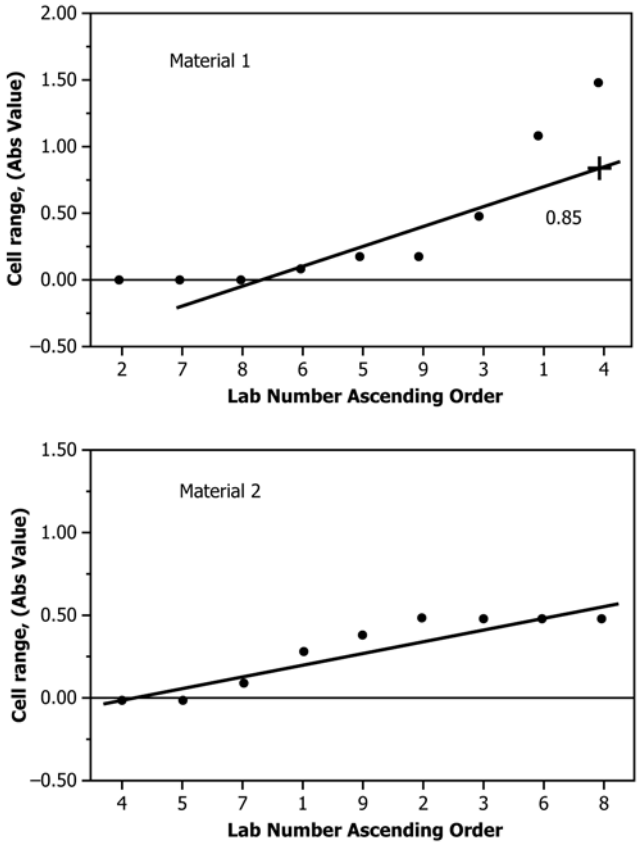
A6.3 Part 1: Outlier Replacement - Analysis Step 2

A6.3.1 *Outlier Treatment*—The Step 2 analysis process is twofold: (*1*) it generates a revised database on which the second round of calculations is conducted to obtain revised values for *r* and *R*, and other parameters, using the procedures as set forth in Annex A4, and (*2*) the revised database is examined to detect any potential outliers at the 2 % significance level.

A6.3.1.1 The Step 2 analysis is started with the calculations for Option 2 replacements for the 5 % significance outliers as detected in Step 1. In preparation for this, a second set of spreadsheet tables is generated. To make comparisons and table identification easier Step 1 vs Step 2, (and also Step 3) the table designations for Step 2 retain the (second symbol set) use of (1) to (6) with two added symbols within the parenthesis. First , the Revision 1 database symbol R1 is added and Table A6.1 (1) in Step 1 becomes Table A6.8 (1-R1) in Step 2. The second addition for Option 2 tables is the symbol, OR , where OR designates "outliers replaced". Thus to complete the identification, Table A6.8 (1-R1) becomes Table A6.8 (1-R1-OR) for Step 2, Option 2. Recall that Step 1 is conducted on the original database. This same system of additional symbols is employed for the Step 3 group of tables. In Step 3 the Revision 2 database symbol R1 is replaced with R2, thus Step 2 Table A6.8 (1-R1-OR) becomes Table A6.15 (1-R2-OR) in Step 3. There are a total of 21 tables for the three steps of the OR analysis. The same procedure is applied to the 14 tables in the "outlier deleted" or OD analysis. For the OD analysis it is not necessary to duplicate the first seven tables of the original database.

A6.3.2 *Step 2 Analysis: Replacement of 5 % Significance Outliers*—To implement Outlier Option 2, AOT replacement values must be obtained for the outliers discovered in the Step 1 analysis. Refer to Annex A5 for the AOT procedure. Basically two operations need to be performed; evaluate PRVs and then calculate DRVs for both cell averages and to cell standard deviations or ranges. Once this has been done calculation of the new set of precision values for the *R1* database can be conducted.

A6.3.2.1 *PRVs for Cell Averages*—This operation for *cell averages* is conducted, using the procedure of Annex A5 in conjunction with Figs. A6.1-A6.4 and Fig. A6.5. In Fig. A6.1 the value for Laboratory 9 was declared as an outlier in the Step 1 analysis. The PRV of 49.4 for Laboratory 9, Material 1, indicated by a cross symbol, was obtained by the A5.3.1 procedure. The cell average PRV of 69.7 for Material 2 was obtained for Laboratory 1, using the same procedure. In Fig.



NOTE 1—With dashed linear trend line and replacement values indicated.

FIG. A6.3 AOT Plots—Original Cell Ranges for Materials 1 and 2

A6.2, the cell average PRVs (69.0, 96.5) for Lab 9 for both materials were calculated in the same manner. In Fig. A6.3, the cell range PRV for Laboratory 4 is evaluated as 0.85. In Fig. A6.4 the cell range PRVs of 2.20 and 1.20 were obtained for Laboratory 4 for Materials 3 and 4, respectively, using the same procedure. The PRVs for cell averages are tabulated as Item 1 in Part A of Table A6.36, and the PRVs for cell ranges are tabulated as Item 2 in Part A of Table A6.36.

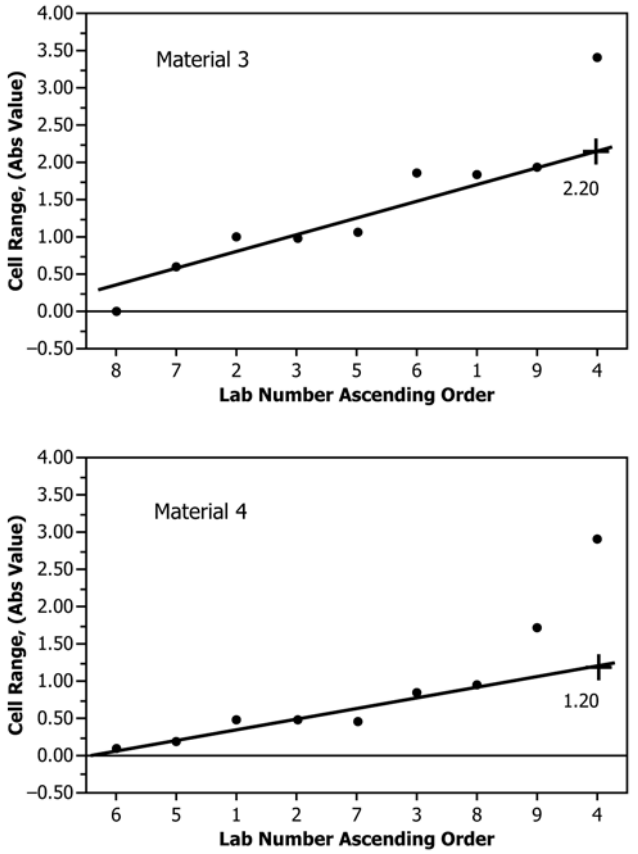
A6.3.2.2 *DRVs*—The next operation is to convert these cell PRVs into cell DRVs using the procedures of A5.4. The cell DRVs are required for entry into a Table A6.1 (1) format to generate a new Table A6.8 (1-R1-OR).

(1) DRVs for Cell Average—There are two types of cell average DRVs as outlined in A5.4. For this example, all cell average DRVs are the first type as described in A5.4.1, that is, for *Cell Average Outlier with Non-Outlier Cell Range*. The cells scheduled for replacement do not have accompanying cell range outliers. The DRVs for this first type can be calculated using the PRV (for cell averages) obtained in A6.3.2.1, and the existing cell range for that cell, using Eq A5.1 and A5.2 in A5.4.1. The data entries in Item 3 Part B of Table A6.36 were obtained using these two equations with PRVs (cell average) in Part A and the cell ranges that exist for the four cells in question (these are listed in parentheses next to the replacement averages in Part A). The calculated *cell average* DRVs are shown in Item 3 of Part B of Table A6.36.

(2) DRVs for Cell Range—The cell range PRVs, as listed in Item 2 of Part A in Table A6.36, need to be converted to cell range DRVs. All three of these are of the third type, that is, *Cell Range Outlier with Non-Outlier Cell Average*, see A5.4.3. The conversion from PRV to DRVs (duplicate data values) is achieved for any selected cell, using (*1*) the PRV range obtained in A6.3.2.1, and (*2*) the existing cell average for that cell and Eq A5.5 and A5.6. The results of these calculations are shown in Item 4 of Part B of Table A6.36.

A6.3.3 *Step 2 Analysis: Precision for Revised Database with Outlier Replacements*—Once the outlier replacements have been calculated and tabulated in Table A6.36, the revised database can be reanalyzed. This begins with Table A6.8 (1-R1-OR). The DRVs of Table A6.36 are substituted for the individual cell outlier values in Table A6.8 (1-R1-OR); these are indicated with italics. Once the replacement values for all cells have been entered into Table A6.8 (1-R1-OR), the *R1* precision results appear in Table A6.14 (6-R1-OR).

A6.3.3.1 Table A6.14 (6-R1-OR) indicates that the repeatability *r* has been reduced, values now span the interval 0.76 to 2.92 and *R* spans the interval 1.76 to 11.27. On an overall or pooled basis the repeatability has been improved for *r* by a reduction factor of 0.88 (that is, 12 % less for *r*) and the reproducibility for *R* has been improved by a reduction factor



NOTE 1—With linear trend line and PRV indicated.

FIG. A6.4 AOT Plots—Cell Ranges for Materials 3 and 4

of 0.76 (24 % less for *R*) using the *R1* database generated by the outlier replacement procedure.

A6.3.4 *Step 2 Analysis: Detection and Replacement of 2 % Significance Outliers*—Once DRVs for the 5 % outliers are entered into the Table A6.8 (1-R1-OR), the calculation operations for all subsequent tables follow automatically. Critical values for *h* and *k* at the 2 % significance level are obtained from Table A3.1. Table A6.10 (3-R1-OR) shows a cell average outlier for Material 4 in Laboratory 8. The calculated *h*-value of 2.07 exceeds the critical *h*-value of 2.00. Table A6.13 (5-R1-OR) indicates that the cell range for Material 1 in Laboratory 1, is an outlier with a calculated *k*-value of 2.15 exceeding the 2 % critical value 2.09.

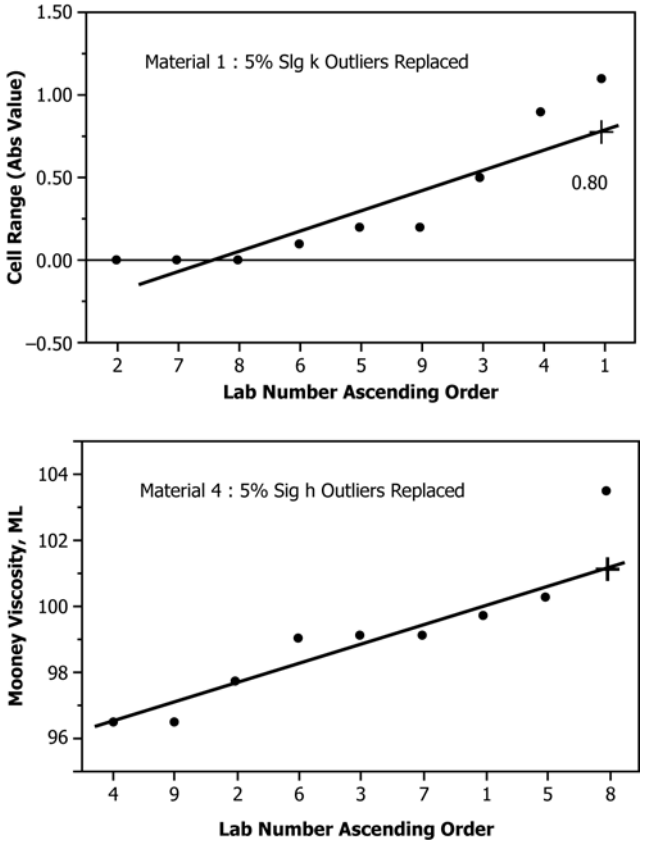
A6.3.4.1 The final action required for a Step 2 analysis is the replacement of the data values found to be outliers at the 2 % significance level. Fig. A6.5 illustrates AOT plots for Material 1 with the range value of 0.80 indicated as the replacement of outlier value 1.10 for Laboratory 1. Also shown is the plot for Material 4 with the cell average replacement value of 101.2 for the outlier 103.5 for Laboratory 8. The two PRVs, 0.80 and 101.2, need to be converted into DRVs. The cell range PRV of 0.80 is converted to DRVs using A5.4.3 and the cell average PRV of 101.2 is converted to DRVs using A5.4.1, as described in A6.3.2.1 and A6.3.2.2. These replacement values are shown in Table A6.10 (3-R1-OR) in bold italic font.

A6.4 Part 1: Outlier Replacement—Analysis Step 3

A6.4.1 When the DRVs for the two 2 % significance outlier values in the Step 2 analysis are inserted into Table A6.8 (1-R1-OR), a new Table A6.15 (1-R2-OR) is generated, an *R2* database. Refer to the sequence, Table A6.15 (1-R2-OR) to Table A6.21 (6-R2-OR); the last table gives the *R2* and final Option 2 precision for repeatability and reproducibility. Comments on the improved precision or reduction in *r* and *R* will be postponed until the Option 1 analysis is conducted in Part 2.

A6.5 Part 2: General Precision Analysis—Option 1: Outlier Deletion

A6.5.1 A substantial portion of the work for Part 2–Option 1 has already been done in Part 1. Tables A6.1 (1)-A6.6 (5) and Table A6.36, and the two sub-tables at the bottom of Table A6.21 (6-R2-OR) all indicate the values that have been declared as *h* and *k* outliers in the Part 1 analysis. If Option 1, outlier deletion, had been an initial analysis decision or a decision after Step 1, the preliminary review of section A6.2.1



NOTE 1—With linear trend line and PRV indicated.

FIG. A6.5 AOT Plots—Revised (R1) Database for Materials 1 and 4

and the precision calculations and outlier review of the original database as described in section A6.2.2 would be the first operation for a Part 2 analysis. These constitute Part 2–Step 1 and do not need to be repeated here.

A6.6 Part 2: Outlier Deletion—Analysis Step 2

A6.6.1 *Deletion of 5 % Significance Outliers*—Since all outliers have been detected in Part 1, the deletion process is all that is required for this Part 2 analysis. However in the ordinary analysis of an ITP, if Option 1 is chosen as an initial decision, the outlier detection steps for both the 5 % and 2 % significance outliers would be required prior to the action now described.

A6.6.1.1 Table A6.22 (1-R1-OD) shows the results of the deletion process on the original database Table A6.1 (1), to generate the *R1* database. The tabulated values that have been declared significant, at the 5 % level for *h* and *k* outliers, have been deleted. Table A6.23 (2-R1-OD) to A6.28 (6-R1-OD) are also shown with the blank cells at the locations indicated by the deleted 5 % outliers. In the spreadsheet analysis, all of the blank cells in this series of tables will initially have an *ERROR* indication. As explained in Annex A5, each cell *ERROR* value must be deleted to produce a blank cell. The final precision results are given in Table A6.28 (6-R1-OD). Comparing the results of the outlier replacement Option 2 with the outlier deletion Option 1, indicates that Option 1 in general gives smaller values for both *r* and *R*. A more detailed discussion of the two options will be conducted in section A6.8.

A6.6.2 *Deletion of 2 % Significance Outliers*—The next operation is the deletion of cell values that have been declared as outliers at the 2 % significance level. Note at the bottom of Table A6.25 (6-R1-OD) that two values are indicated; the cell average for Material 4 for Laboratory 8 and cell range (or standard deviation) for Material 1 for Laboratory 1. The case of Material 1–Laboratory 1 requires some special consideration by the analyst. Refer to A6.25 (4R-R1-OD). If the Laboratory 1 range of 1.10 is deleted, we are left with six range values much smaller than 1.10, three of which are zero.

A6.6.2.1 Although it is possible to get perfect agreement for two Mooney viscosity measurements one week apart in three of the laboratories, this occurrence must be viewed with some caution. Most technicians know when a special test or ITP is being conducted and they know that good agreement is the goal. A temptation exists to make the results look good. The analyst's judgment in this instance is that the pooled standard deviation (pooled range) would be unrealistically low if the Laboratory 1 value of 1.10 were to be deleted. Therefore, a decision is made to override the objective analysis outcome and not delete the 1.10.

A6.6.2.2 In the Part 1 analysis, the Laboratory 1 range of 1.10 for Material 1 was removed, but it was replaced by a value of 0.80. This is different than an outright deletion that removes a laboratory from the list of participants for any material. The deletion of only the Material 4 Laboratory 8 value from the *R1* database, yields Table A6.29 (1-R2-OD). This table represents the *R2* database.

A6.6.3 *Alternative Option for Special Case Outlier Treatment*—The decision to retain the Material 1–Laboratory 1 range of 1.10, brings up a possibility for consideration; the combined use of Option 1 and Option 2 for outlier treatment. In the case of the Part 2 Step 2 analysis, it is possible for the analyst to use the Option 2 AOT replacement of 0.80 for this Laboratory range value, rather than deleting it. This is an alternative option that may be used. It is a judgment call by the analyst.

A6.7 Part 2: Outlier Deletion—Analysis Step 3

A6.7.1 The final precision results for Part 2–Option 1 are given in Table A6.35 (6-R2-OD). Comparing the results of the outlier replacement Option 2 with the outlier deletion Option 1, Table A6.21 (6-R2-OR) versus Table A6.35 (6-R2-OD), indicates that Option 1 in general gives smaller values for both *r* and *R*.

A6.8 Discussion of Precision Results: Option 1 versus Option 2

A6.8.1 *Option 1 (Deletion) versus Option 2 (AOT Replacement)*—The comparison of the two options is illustrated in Table A6.37, and in Table A6.38 reduction factors for *r* and *R* are given. Both tables may be summarized as follows.

A6.8.1.1 For repeatability, the two Options are essentially equal for Materials 1 and 2. However, for Material 3 and especially Material 4, the Option 1 outlier deletion procedure gives increased reductions or substantially improved repeatability. The pooled values give a reduction factor of 0.65 for Option 1 deletion versus a reduction factor of 0.78 for Option 2 replacement; an overall 20 % advantage for Option 1.

A6.8.1.2 For reproducibility, the two Options are essentially equal for Material 1 and 3, but the Option 1 (deletion) gives improvement for Material 2 and substantial improvement for Material 4. The pooled values give a reduction factor of 0.64 for Option 1 deletion versus a reduction factor of 0.70 for Option 2 replacement; an overall 9 % advantage for Option 1.

A6.8.2 *Precision versus the Four Materials*—The precision performance among the four materials for the Option 1 (deletion) procedure is indicated in Table A6.37. These results have been inserted into the standard Table 6 format summary of precision as described in Section 12. The precision in this format for the Mooney viscosity example is given in Table A6.39 that lists all the precision parameters and also the final number of laboratories in the ITP database after deletion of all outliers.

A6.8.2.1 Materials 1, 2, and 4 give repeatability values, *r*, that are roughly equal, 0.92, 0.76, and 1.03, respectively. These three *r* values differ substantially as a group, from those obtained for the original database: 1.29, 3.43, and 2.54, respectively, for Materials 1, 2, and 4. The outlier removal operation has reduced the *r* parameter and gives an indication that all three are very nearly equal. In a technical sense this is not surprising since Materials 1, 2, and 4 are non-pigmented or clear rubbers, and they should respond to the measurement process in a similar manner within the confines of a single laboratory.

A6.8.2.2 Material 3 is an SBR black masterbatch (SBR-BMB) with 65 phr of N339 carbon black. Note that the repeatability for Material 3 is substantially poorer (higher *r*) compared to the other three by a factor of 2.7 on an overall basis. Reasons for this lack of precision are discussed in A6.8.3.

A6.8.2.3 The Option 1 (deletion) reproducibility, *R*, for Materials 1 and 4 is essentially equal (2.71 and 2.50) while Material 2 has the lowest *R* at 1.49. Again Material 3 is very high, $R = 10.84$; roughly by a factor of 5 compared to the other three materials on a overall basis. This is about twice the repeatability comparative precision factor of 2.7. For Materials 1 to 4, the Option 1 reproducibility is substantially improved (lower *R*) compared to the original database *R* values of 3.37, 1.97, 15.15, and 8.84 respectively. Note the considerable differences for the original database *R* values among Materials 1, 2, and 4 compared to the much more nearly equal values (Materials 1, 2, 4) as previously noted.

A6.8.2.4 The roughly equal reproducibility, *R*, for Materials 1 and 4 (SBR and NR) is again a reasonably expected outcome; similar test response in a between laboratory sense for these two un-pigmented rubbers. Material 2 (butyl, reference rubber) is produced to have high uniformity (good homogeneity bale to bale); it is used as a reference rubber to check the operation of Mooney viscometers. This uniformity undoubtedly accounts for part of its good performance. Also this rubber was not subjected to the mill-massing operation.

A6.8.3 *SBR-BMB Precision*—The very poor performance for Material 3, the SBR-BMB, was the subject of further investigation when this ITP was conducted. Subsequent laboratory work showed that the problem was attributed to the procedure used to mill-mass the rubber prior to conducting the Mooney test. In the mill-massing procedure, the mill temperature, the mill nip (opening) and the time on the mill were not sufficiently well-specified and controlled. Both factors were found to play a very important role in the amount of rubber breakdown. Variation in this prior mill massing operation was the source of the poor precision; variable breakdown leads to variable viscosity.

A6.8.3.1 The breakdown for the SBR-BMB was a combination of (*1*) rupture of rubber-carbon black intermolecular bonding and (*2*) ordinary chain rupture. The clear mill-massed rubbers, SBR 1712 and NR, also suffered some chain rupture, but the existence of the additional greater magnitude breakdown mechanism for the SBR-BMB made it much more susceptible to mill-massing variations and produced the poor precision. Test Methods D1646 and ISO 289 were subsequently revised to eliminate the mill massing operation for BMB rubbers.

A6.8.3.2 Due to the poor precision (high *r* and *R*) for the SBR-BMB, this material was not included in the pooled value calculations in Table A6.39. Pooling is recommended only when the precision values are reasonably close or vary in some known way for all materials in any ITP.

A6.8.4 *Final Observations*—The 3 Step Analysis outlier removal operation using the *h* and *k* statistics, Step 1 at the 5 % significance level and Step 2 at the 2 % significance level on the revised database, has given improved repeatability and

reproducibility, compared to the original database. Option 1 (deletion) yields nearly equal *r* and *R* parameters for all three un-pigmented rubbers. A good analysis outcome can be obtained using either Option 1 or Option 2, but Option 1 involves less computation and it yields better precision. Option 1 is the preferred choice when there are nine or more laboratories in any ITP.

A6.8.4.1 The 3 Step Option 1 Analysis has in essence isolated a *core group* of laboratories that have good control of Mooney viscosity testing. Table A6.29 (1-R2-OD) indicates that Laboratories 4 and 8 each had three outliers deleted. These two laboratories have poor control over testing and are in need of improvement. Laboratory 1 also is in need of some remedial efforts, it had two outliers, one of which was not deleted in Option 1 as previously cited. Laboratory 8 had one outlier, and it may need some attention to testing procedures. The *core group* of five laboratories (2, 3, 5, 6, and 7) had good control over their testing domain. For Materials 1, 2, and 4, the relative repeatability (*r*) was 1.8, 1.1, and 1.0 % and the relative reproducibility (*R*) was 5.4, 2.2, and 2.5 % respectively. The precision attained by this *core group* should be the benchmark for Mooney viscosity testing in the rubber manufacturing industry.

TABLE A6.1 (1): Mooney Viscosity: Original Basic Data from the ITP

	Material 1		Material 2		Material 3		Material 4	
Lab #	Day1	Day2	Day1	Day2	Day1	Day2	Day1	Day2
1	48.8	49.9	70.0	70.3	72.3	74.2	100.0	99.5
2	51.0	51.0	68.0	68.5	69.0	70.0	97.5	98.0
3	50.4	49.9	68.1	68.6	72.6	73.6	98.7	99.6
4	51.0	49.5	68.0	68.0	79.0	75.5	98.0	95.0
5	50.3	50.1	68.5	68.5	76.0	77.1	100.2	100.4
6	52.4	52.3	69.5	69.0	80.4	82.3	99.0	99.1
7	50.8	50.8	69.5	69.4	71.8	72.4	98.9	99.4
8	51.0	51.0	69.0	68.5	76.0	76.0	104.0	103.0
9	48.1	48.3	69.0	68.6	63.6	61.6	93.0	91.2
Day Avg	50.42	50.31	68.84	68.82	73.41	73.63	98.81	98.36
2 Day Avg		50.37		68.83		73.52		98.58
Bet Lab SDev	1.28	1.13	0.74	0.67	5.17	5.66	2.88	3.41
Pooled Bet Lab SDev		1.21		0.71		5.42		3.16

TABLE A6.2 (2): Cell Average and Averages Squared: Original Data

	Cell Averages					Cell Averages Squared				
	Lab #	Matl 1	Matl 2	Matl 3	Matl 4	Lab #	Matl 1	Matl 2	Matl 3	Matl 4
	1	49.35	70.15	73.25	99.75	1	2435.42	4921.02	5365.56	9950.06
	2	51.00	68.25	69.50	97.75	2	2601.00	4658.06	4830.25	9555.06
	3	50.15	68.35	73.10	99.15	3	2515.02	4671.72	5343.61	9830.72
	4	50.25	68.00	77.25	96.50	4	2525.06	4624.00	5967.56	9312.25
	5	50.20	68.50	76.55	100.30	5	2520.04	4692.25	5859.90	10060.09
	6	52.35	69.25	81.35	99.05	6	2740.52	4795.56	6617.82	9810.90
	7	50.80	69.45	72.10	99.15	7	2580.64	4823.30	5198.41	9830.72
	8	51.00	68.75	76.00	103.50	8	2601.00	4726.56	5776.00	10712.25
	9	48.20	68.80	62.60	92.10	9	2323.24	4733.44	3918.76	8482.41
	T1 =	453.300	619.500	661.700	887.250	**T2 =**	22841.950	42645.925	48877.880	87544.473
	Cell Avg	50.37	68.83	73.52	98.58					
Var Cell Avg		1.3425	0.4594	28.5282	9.5513					
SDev Cell Avg		1.159	0.678	5.341	3.091					

Note: variance cell avg = S^2(Yav)

TABLE A6.3 (3): Cell Average Deviations, d and h-values: Original Data

	Cell Deviations, d					Cell h-values				
	Lab #	Matl 1	Matl 2	Matl 3	Matl 4	Lab #	Matl 1	Matl 2	Matl 3	Matl 4
	1	−1.02	1.32	−0.27	1.17	1	−0.88	**1.94**	−0.05	0.38
	2	0.63	−0.58	−4.02	−0.83	2	0.55	−0.86	−0.75	−0.27
	3	−0.22	−0.48	−0.42	0.57	3	−0.19	−0.71	−0.08	0.18
	4	−0.12	−0.83	3.73	−2.08	4	−0.10	−1.23	0.70	−0.67
	5	−0.17	−0.33	3.03	1.72	5	−0.14	−0.49	0.57	0.56
	6	1.98	0.42	7.83	0.47	6	1.71	0.61	1.47	0.15
	7	0.43	0.62	−1.42	0.57	7	0.37	0.91	−0.27	0.18
	8	0.63	−0.08	2.48	4.92	8	0.55	−0.12	0.46	1.59
	9	−2.17	−0.03	−10.92	−6.48	9	***−1.87***	−0.05	***−2.04***	***−2.10***
All Lab Cell Avg		50.37	68.83	73.52	98.58	**h(crit) 5%Sig Level at indicated p**				
SDev Cell Avgs		1.159	0.678	5.341	3.091	**p =**	9	9	9	9
						h(crit)	1.78	1.78	1.78	1.78
						Lab#>h(crit)	9	1	9	9

Bold and Italic = significant values
h = d / S (Yav); where d = avg Cell i – avg All Cells, S (Yav) = std dev of Cell avgs.

TABLE A6.4 (4R): Cell Ranges and Ranges Squared: Original Data

	Cell Ranges					Cell Ranges Squared			
Lab #	Matl 1	Matl 2	Matl 3	Matl 4	Lab #	Matl 1	Matl 2	Matl 3	Matl 4
1	1.100	0.300	1.900	0.500	1	1.210	0.090	3.610	0.250
2	0.000	0.500	1.000	0.500	2	0.000	0.250	1.000	0.250
3	0.500	0.500	1.000	0.900	3	0.250	0.250	1.000	0.810
4	1.500	0.000	3.500	3.000	4	2.250	0.000	12.250	9.000
5	0.200	0.000	1.100	0.200	5	0.040	0.000	1.210	0.040
6	0.100	0.500	1.900	0.100	6	0.010	0.250	3.610	0.010
7	0.000	0.100	0.600	0.500	7	0.000	0.010	0.360	0.250
8	0.000	0.500	0.000	1.000	8	0.000	0.250	0.000	1.000
9	0.200	0.400	2.000	1.800	9	0.040	0.160	4.000	3.240
Avg Range	0.400	0.311	1.444	0.944	**.T3 =**	3.8000	1.2600	27.0400	14.8500

T3 = Sum Cell 'Ranges Squared'
Calculation algorithm for any ITP cell Range, with duplicates in cells, cxx and dxx;
@IF ((cxx – dxx) < 0, (cxx – dxx)* – 1, (cxx – dxx))

TABLE A6.5 (4s): Cell Standard Deviations and Variances: Original Data

	Cell Std Deviations					Cell Variances			
Lab #	Matl 1	Matl 2	Matl 3	Matl 4	Lab #	Matl 1	Matl 2	Matl 3	Matl 4
1	0.778	0.212	1.344	0.354	1	0.6050	0.0450	1.8050	0.1250
2	0.000	0.354	0.707	0.354	2	0.0000	0.1250	0.5000	0.1250
3	0.354	0.354	0.707	0.636	3	0.1250	0.1250	0.5000	0.4050
4	1.061	0.000	2.475	2.121	4	1.1250	0.0000	6.1250	4.5000
5	0.141	0.000	0.778	0.141	5	0.0200	0.0000	0.6050	0.0200
6	0.071	0.354	1.344	0.071	6	0.0050	0.1250	1.8050	0.0050
7	0.000	0.071	0.424	0.354	7	0.0000	0.0050	0.1800	0.1250
8	0.000	0.354	0.000	0.707	8	0.0000	0.1250	0.0000	0.5000
9	0.141	0.283	1.414	1.273	9	0.0200	0.0800	2.0000	1.6200
Pooled SDev	0.459	0.265	1.226	0.908	**T4 =**	1.90000	0.63000	13.52000	7.42500
					Pooled Variance	0.2111	0.0700	1.5022	0.8250

TABLE A6.6 (5): Cell k-values: Original Data

Lab #	Matl 1	Matl 2	Matl 3	Matl 4
1	1.69	0.80	1.10	0.39
2	0.00	1.34	0.58	0.39
3	0.77	1.34	0.58	0.70
4	***2.31***	0.00	***2.02***	***2.34***
5	0.31	0.00	0.63	0.16
6	0.15	1.34	1.10	0.08
7	0.00	0.27	0.35	0.39
8	0.00	1.34	0.00	0.78
9	0.31	1.07	1.15	1.40
Pooled SDev	0.459	0.265	1.226	0.908
k(crit) 5% Sig Level at n=2, indicated p;				
p =	9	9	9	9
k(crit) =	1.90	1.90	1.90	1.90
Lab#>k(crit)	4	none	4	4

Bold and italic = significant values
k = S(i) / Sr; where S(i) = individual cell standard deviation, Sr = pooled all lab standard deviation

TABLE A6.7 (6): Mooney Viscosity: Calculations for Precision–Original Data

	ITP for : n =	2	2	2	2
	p =	9	9	9	9
		Matl 1	Matl 2	Matl 3	Matl 4
	T1 =	453.300	619.500	661.700	887.250
	T2 =	22841.950	42645.925	48877.880	87544.473
	T4 =	1.90000	0.63000	13.52000	7.42500
Calc 1	(Sr)^2 =T4/p =	0.2111	0.0700	1.5022	0.8250
	(SL)^2 = {[p T2 – (T1)^2]/p(p – 1)} – [(Sr)^2/2]				
Calc 2	(SL)^2 =	1.2369	0.4244	27.7771	9.1388
	(SR)^2 = (SL)^2 + (Sr)^2				
Calc 3	(SR)^2 =	1.4481	0.4944	29.2793	9.9638
	r = 2.8 [(Sr)^2]^0.5 = Repeatability				
Calc 4	r =	1.287	0.741	3.432	2.543
	R = 2.8 [(SR)^2]^0.5 = Reproducibility				
Calc 5	R =	3.37	1.97	15.15	8.84

Material Averages	50.37	68.83	73.52	98.58

Standard Deviation, Sr =	0.459	0.265	1.226	0.908
Standard Deviation, SR =	1.203	0.703	5.411	3.157

Relative Precision:	Matl 1	Matl 2	Matl 3	Matl 4
(r) ==>	2.55	1.08	4.67	2.58
(R) ==>	6.69	2.86	20.61	8.97

Step 1: Outliers at 5% Significance Level for Materials 1 to 4					
		Matl 1	Matl 2	Matl 3	Matl 4
For h :	Lab #	9	1	9	9
For k :	Lab #	4	none	4	4

TABLE A6.8 (1-R1-OR): Mooney Viscosity: AOT Replacement Values (Italic) for 5 % Outliers

	Material 1		Material 2		Material 3		Material 4	
Lab #	Day1	Day2	Day1	Day2	Day1	Day2	Day1	Day2
1	48.8	49.9	***69.6***	***70.0***	72.3	74.2	100.0	99.5
2	51.0	51.0	68.0	68.5	69.0	70.0	97.5	98.0
3	50.4	49.9	68.1	68.6	72.6	73.6	98.7	99.6
4	***49.8***	***50.7***	68.0	68.0	***76.2***	***78.4***	***95.9***	***97.1***
5	50.3	50.1	68.5	68.5	76.0	77.1	100.2	100.4
6	52.4	52.3	69.5	69.0	80.4	82.3	99.0	99.1
7	50.8	50.8	69.5	69.4	71.8	72.4	98.9	99.4
8	51.0	51.0	69.0	68.5	76.0	76.0	104.0	103.0
9	***49.3***	***49.5***	69.0	68.6	***68.0***	***70.0***	***95.9***	***97.1***
Day Avg	50.42	50.58	68.80	68.79	73.59	74.89	98.90	99.24
2 Day Avg		50.50		68.79		74.24		99.07
Bet Lab SDev	1.06	0.84	0.67	0.59	3.92	4.02	2.47	1.82
Pooled Bet Lab SDev		0.96		0.63		3.97		2.17

Replaced Values = Bold, Italic

TABLE A6.9 (2-R1-OR): Cell Averages, Average Squared: AOT Replacements for 5 % Outliers

Cell Averages					Cell Averages Squared				
Lab #	Matl 1	Matl 2	Matl 3	Matl 4	Lab #	Matl 1	Matl 2	Matl 3	Matl 4
1	49.35	69.80	73.25	99.75	1	2435.42	4872.04	5365.56	9950.06
2	51.00	68.25	69.50	97.75	2	2601.00	4658.06	4830.25	9555.06
3	50.15	68.35	73.10	99.15	3	2515.02	4671.72	5343.61	9830.72
4	50.25	68.00	77.30	96.50	4	2525.06	4624.00	5975.29	9312.25
5	50.20	68.50	76.55	100.30	5	2520.04	4692.25	5859.90	10060.09
6	52.35	69.25	81.35	99.05	6	2740.52	4795.56	6617.82	9810.90
7	50.80	69.45	72.10	99.15	7	2580.64	4823.30	5198.41	9830.72
8	51.00	68.75	76.00	103.50	8	2601.00	4726.56	5776.00	10712.25
9	49.40	68.80	69.00	96.50	9	2440.36	4733.44	4761.00	9312.25
T1 =	454.500	619.150	668.150	891.650	**T2 =**	22959.070	42596.943	49727.848	88374.313
Cell Avg	50.50	68.79	74.24	99.07					
Var Cell Avg	0.8525	0.3578	15.6417	4.5707					
SDev Cell Avg	0.923	0.598	3.955	2.138					

Note: variance cell avg = S^2(Yav)

TABLE A6.10 (3-R1-OR): Cell Average Deviation d and h-values: AOT Replacement for 5 % Outliers

	Cell Deviations, d					Cell h-values				
	Lab #	Matl 1	Matl 2	Matl 3	Matl 4	Lab #	Matl 1	Matl 2	Matl 3	Matl 4
	1	−1.15	1.01	-0.99	0.68	1	−1.25	1.68	−0.25	0.32
	2	0.50	−0.54	−4.74	−1.32	2	0.54	−0.91	−1.20	−0.62
	3	−0.35	−0.44	−1.14	0.08	3	−0.38	−0.74	−0.29	0.04
	4	−0.25	−0.79	3.06	−2.57	4	−0.27	−1.33	0.77	−1.20
	5	−0.30	−0.29	2.31	1.23	5	−0.32	−0.49	0.58	0.57
	6	1.85	0.46	7.11	−0.02	6	2.00	0.76	1.80	−0.01
	7	0.30	0.66	−2.14	0.08	7	0.32	1.10	−0.54	0.04
	8	0.50	−0.04	1.76	4.43	8	0.54	−0.07	0.45	***2.07***
	9	−1.10	0.01	−5.24	−2.57	9	−1.19	0.01	−1.32	−1.20
						h(crit) 2%Sig Level at indicated p				
All Lab Cell Avg		50.50	68.79	74.24	99.07	p =	9	9	9	9
SDev Cell Avg		0.923	0.598	3.955	2.138	h(crit)	2.00	2.00	2.00	2.00
						Lab#>h(crit)	none	none	none	8

Bold and italic = signficant values
h = d / S (Yav); where d = avg Cell i – avg All Cells, S (Yav) = std dev of Cell avgs

TABLE A6.11 (4R-R1-OR): Cell Range, Range Squared: AOT Replacement for 5 % Outliers

	Cell Ranges					Cell Ranges Squared				
	Lab #	Matl 1	Matl 2	Matl 3	Matl 4	Lab #	Matl 1	Matl 2	Matl 3	Matl 4
	1	1.100	0.400	1.900	0.500	1	1.210	0.160	3.610	0.250
	2	0.000	0.500	1.000	0.500	2	0.000	0.250	1.000	0.250
	3	0.500	0.500	1.000	0.900	3	0.250	0.250	1.000	0.810
	4	0.900	0.000	2.200	1.200	4	0.810	0.000	4.840	1.440
	5	0.200	0.000	1.100	0.200	5	0.040	0.000	1.210	0.040
	6	0.100	0.500	1.900	0.100	6	0.010	0.250	3.610	0.010
	7	0.000	0.100	0.600	0.500	7	0.000	0.010	0.360	0.250
	8	0.000	0.500	0.000	1.000	8	0.000	0.250	0.000	1.000
	9	0.200	0.400	2.000	1.200	9	0.040	0.160	4.000	1.440
Avg Range		0.333	0.322	1.300	0.678	T3 =	2.3600	1.3300	19.6300	5.4900

T3 = Sum Cell 'Ranges Squared'

TABLE A6.12 (4S-R1-OR): Cell Standard Deviations and Variances: AOT Replacement for 5 % Outliers

	Cell Std Deviations					Cell Variances				
	Lab #	Matl 1	Matl 2	Matl 3	Matl 4	Lab #	Matl 1	Matl 2	Matl 3	Matl 4
	1	0.778	0.283	1.344	0.354	1	0.6050	0.0800	1.8050	0.1250
	2	0.000	0.354	0.707	0.354	2	0.0000	0.1250	0.5000	0.1250
	3	0.354	0.354	0.707	0.636	3	0.1250	0.1250	0.5000	0.4050
	4	0.636	0.000	1.556	0.849	4	0.4050	0.0000	2.4200	0.7200
	5	0.141	0.000	0.778	0.141	5	0.0200	0.0000	0.6050	0.0200
	6	0.071	0.354	1.344	0.071	6	0.0050	0.1250	1.8050	0.0050
	7	0.000	0.071	0.424	0.354	7	0.0000	0.0050	0.1800	0.1250
	8	0.000	0.354	0.000	0.707	8	0.0000	0.1250	0.0000	0.5000
	9	0.141	0.283	1.414	0.849	9	0.0200	0.0800	2.0000	0.7200
Pooled SDev		0.362	0.272	1.044	0.552	**T4 =**	1.18000	0.66500	9.81500	2.74500
						Pooled Variance	0.1311	0.0739	1.0906	0.3050

TABLE A6.13 (5-R1-OR): k-values: AOT Replacement for 5 % Outliers

Lab #	Matl 1	Matl 2	Matl 3	Matl 4
1	***2.15***	1.04	1.29	0.64
2	0.00	1.30	0.68	0.64
3	0.98	1.30	0.68	1.15
4	1.76	0.00	1.49	1.54
5	0.39	0.00	0.74	0.26
6	0.20	1.30	1.29	0.13
7	0.00	0.26	0.41	0.64
8	0.00	1.30	0.00	1.28
9	0.39	1.04	1.35	1.54
Pooled SDev	0.362	0.272	1.044	0.552
k(crit) 2% Sig Level at n=2, indicated p;				
p =	9	9	9	9
k(crit) =	2.09	2.09	2.09	2.09
Lab#>k(crit)	1	none	none	none

Bold and italic = significant values
k = S(i) / Sr; where S(i) = indiv cell std dev, Sr = pooled all lab std dev

TABLE A6.14 (6-R1-OR): Mooney Viscosity Calculations for Precision AOT Replacements for 5 % Outliers

	ITP for : n =	2	2	2	2
	p =	9	9	9	9
		Matl 1	Matl 2	Matl 3	Matl 4
	T1 =	454.500	619.150	668.150	891.650
	T2 =	22959.070	42596.943	49727.848	88374.313
	T4 =	1.18000	0.66500	9.81500	2.74500
Calc 1	(Sr)^2 = T4/p =	0.1311	0.0739	1.0906	0.3050
	(SL)^2 = {[p T2 – (T1)^2]/p (p – 1)} – [(Sr)^2/2]				
Calc 2	(SL)^2 =	0.7869	0.3208	15.0965	4.4182
	(SR)^2 = (SL)^2 + (Sr)^2				
Calc 3	(SR)^2 =	0.9181	0.3947	16.1870	4.7232
	r = 2.8 [(Sr)^2]^0.5 = Repeatability				
Calc 4	r =	1.014	0.761	2.924	1.546
	R = 2.8 [(SR)^2]^0.5 = Reproducibility				
Calc 5	R =	2.68	1.76	11.27	6.09
	Material Averages	50.50	68.79	74.24	99.07
Standard Deviation, Sr =		0.362	0.272	1.044	0.552
Standard Deviation, SR =		0.958	0.628	4.023	2.173
	Relative Precision:	Matl 1	Matl 2	Matl 3	Matl 4
	(r) ==>	2.01	1.11	3.94	1.56
	(R) ==>	5.31	2.56	15.17	6.14
	Step 1: Outliers at 5% Significance Level for Materials 1 to 4				
		Matl 1	Matl 2	Matl 3	Matl 4
	For h : Lab #	9	1	9	9
	For k : Lab #	4	none	4	4
	Step 2: Outliers at 2% Significance Level for Materials 1 to 4				
		Matl 1	Matl 2	Matl 3	Matl 4
	For h : Lab #	none	none	none	8
	For k : Lab #	1(a)	none	none	none

Note: Cell values for Lab 1 Material 1 not deleted for 2 % Sig k-value. See Annex A6 for discussion.

TABLE A6.15 (1-R2-OR): Mooney Viscosity: AOT Replacement Values (italic) for 2 % Outliers

	Material 1		Material 2		Material 3		Material 4	
Lab #	Day1	Day2	Day1	Day2	Day1	Day2	Day1	Day2
1	***49.0***	***49.8***	***69.6***	***70.0***	72.3	74.2	100.0	99.5
2	51.0	51.0	68.0	68.5	69.0	70.0	97.5	98.0
3	50.4	49.9	68.1	68.6	72.6	73.6	98.7	99.6
4	***49.8***	***50.7***	68.0	68.0	***76.2***	***78.4***	**95.9**	**97.1**
5	50.3	50.1	68.5	68.5	76.0	77.1	100.2	100.4
6	52.4	52.3	69.5	69.0	80.4	82.3	99.0	99.1
7	50.8	50.8	69.5	69.4	71.8	72.4	98.9	99.4
8	51.0	51.0	69.0	68.5	76.0	76.0	**100.7**	**101.7**
9	***49.3***	***49.5***	69.0	68.6	***68.0***	***70.0***	**95.9**	**97.1**
Day Avg	50.44	50.57	68.80	68.79	73.59	74.89	98.53	99.10
2 Day Avg		50.51		68.79		74.24		98.82
Bet Lab SDev	1.02	0.85	0.67	0.59	3.92	4.02	1.77	1.51
Pooled Bet Lab SDev		0.94		0.63		3.97		1.64

Replaced Values = Bold, Italic

TABLE A6.16 (2-R2-OR): Cell Average, Average Squared: AOT Replacement for 2 % Outliers

	Cell Averages					Cell Averages Squared			
Lab #	Matl 1	Matl 2	Matl 3	Matl 4	Lab #	Matl 1	Matl 2	Matl 3	Matl 4
1	49.40	69.80	73.25	99.75	1	2440.36	4872.04	5365.56	9950.06
2	51.00	68.25	69.50	97.75	2	2601.00	4658.06	4830.25	9555.06
3	50.15	68.35	73.10	99.15	3	2515.02	4671.72	5343.61	9830.72
4	50.25	68.00	77.30	96.50	4	2525.06	4624.00	5975.29	9312.25
5	50.20	68.50	76.55	100.30	5	2520.04	4692.25	5859.90	10060.09
6	52.35	69.25	81.35	99.05	6	2740.52	4795.56	6617.82	9810.90
7	50.80	69.45	72.10	99.15	7	2580.64	4823.30	5198.41	9830.72
8	51.00	68.75	76.00	101.20	8	2601.00	4726.56	5776.00	10241.44
9	49.40	68.80	69.00	96.50	9	2440.36	4733.44	4761.00	9312.25
T1 =	454.550	619.150	668.150	889.350	**T2 =**	22964.008	42596.943	49727.848	87903.503
Cell Avg	50.51	68.79	74.24	98.82					
Var Cell Avg	0.8384	0.3578	15.6417	2.6125					
SDev Cell Avg	0.916	0.598	3.955	1.616					

Note: variance cell avgs = S^2(Yav)

TABLE A6.17 (3-R2-OR): Cell Average Deviation d and h-values: AOT Replacement for 2 % Outliers

	Cell Deviations, d				Cell h-values				
	Lab #	Matl 1	Matl 2	Matl 3	Matl 4	Lab #	Matl 1	Matl 2	Matl 3	Matl 4
	1	−1.11	1.01	−0.99	0.93	1	−1.21	1.68	−0.25	0.58
	2	0.49	−0.54	−4.74	−1.07	2	0.54	−0.91	−1.20	−0.66
	3	−0.36	−0.44	−1.14	0.33	3	−0.39	−0.74	−0.29	0.21
	4	−0.26	−0.79	3.06	−2.32	4	−0.28	−1.33	0.77	−1.43
	5	−0.31	−0.29	2.31	1.48	5	−0.33	−0.49	0.58	0.92
	6	1.84	0.46	7.11	0.23	6	2.01	0.76	1.80	0.14
	7	0.29	0.66	−2.14	0.33	7	0.32	1.10	−0.54	0.21
	8	0.49	−0.04	1.76	2.38	8	0.54	−0.07	0.45	1.47
	9	−1.11	0.01	−5.24	−2.32	9	−1.21	0.01	−1.32	−1.43
						h(crit) 2% Sig Level at indicated p				
All Lab Cell Avg		50.51	68.79	74.24	98.82	p =	9	9	9	9
SDev Cell Avg		0.916	0.598	3.955	1.616	h(crit)	2.00	2.00	2.00	2.00
						Lab#>h(crit)	NA	NA	NA	NA

Bold and italic = significant values
h = d / S (Yav); where d = avg Cell i – avg All Cells, S (Yav) = std dev of Cell avgs

TABLE A6.18 (4R-R2-OR): Cell Range, Range Squared: AOT Replacement for 2 % Outliers

Cell Ranges					Cell Ranges Squared				
Lab #	Matl 1	Matl 2	Matl 3	Matl 4	Lab #	Matl 1	Matl 2	Matl 3	Matl 4
1	0.800	0.400	1.900	0.500	1	0.640	0.160	3.610	0.250
2	0.000	0.500	1.000	0.500	2	0.000	0.250	1.000	0.250
3	0.500	0.500	1.000	0.900	3	0.250	0.250	1.000	0.810
4	0.900	0.000	2.200	1.200	4	0.810	0.000	4.840	1.440
5	0.200	0.000	1.100	0.200	5	0.040	0.000	1.210	0.040
6	0.100	0.500	1.900	0.100	6	0.010	0.250	3.610	0.010
7	0.000	0.100	0.600	0.500	7	0.000	0.010	0.360	0.250
8	0.000	0.500	0.000	1.000	8	0.000	0.250	0.000	1.000
9	0.200	0.400	2.000	1.200	9	0.040	0.160	4.000	1.440
Avg Range	0.300	0.322	1.300	0.678	**T3 =**	1.7900	1.3300	19.6300	5.4900

T3 = Sum Cell 'Ranges Squared'

TABLE A6.19 (4S-R2-OR): Cell Standard Deviation and Variances: AOT Replacement for 2 % Outliers

	Cell Std Deviations					Cell Variances				
	Lab #	Matl 1	Matl 2	Matl 3	Matl 4	Lab #	Matl 1	Matl 2	Matl 3	Matl 4
	1	0.566	0.283	1.344	0.354	1	0.3200	0.0800	1.8050	0.1250
	2	0.000	0.354	0.707	0.354	2	0.0000	0.1250	0.5000	0.1250
	3	0.354	0.354	0.707	0.636	3	0.1250	0.1250	0.5000	0.4050
	4	0.636	0.000	1.556	0.849	4	0.4050	0.0000	2.4200	0.7200
	5	0.141	0.000	0.778	0.141	5	0.0200	0.0000	0.6050	0.0200
	6	0.071	0.354	1.344	0.071	6	0.0050	0.1250	1.8050	0.0050
	7	0.000	0.071	0.424	0.354	7	0.0000	0.0050	0.1800	0.1250
	8	0.000	0.354	0.000	0.707	8	0.0000	0.1250	0.0000	0.5000
	9	0.141	0.283	1.414	0.849	9	0.0200	0.0800	2.0000	0.7200
Pooled SDev		0.315	0.272	1.044	0.552	**T4 =**	0.89500	0.66500	9.81500	2.74500
						Pooled Variance	0.0994	0.0739	1.0906	0.3050

TABLE A6.20 (5-R2-OR): k-values: AOT Replacement for 2 % Outliers

Lab #	Matl 1	Matl 2	Matl 3	Matl 4
1	1.79	1.04	1.29	0.64
2	0.00	1.30	0.68	0.64
3	1.12	1.30	0.68	1.15
4	2.02	0.00	1.49	1.54
5	0.45	0.00	0.74	0.26
6	0.22	1.30	1.29	0.13
7	0.00	0.26	0.41	0.64
8	0.00	1.30	0.00	1.28
9	0.45	1.04	1.35	1.54
Pooled SDev	0.315	0.272	1.044	0.552
k(crit) 2% Sig Level at n = 2, indicated p;				
p =	9	9	9	9
k(crit) =	2.09	2.09	2.09	2.09
Lab#>k(crit)	NA	NA	NA	NA

Bold and italic = significant values
k = S(i) / Sr; where S(i) = individual cell standard deviation, Sr = pooled all lab standard deviation

TABLE A6.21 (6-R2-OR): Mooney Viscosity Calculations for Precision AOT Replacements for 5 % and 2 % Outliers: Final Precision

		ITP for : n =	2	2	2	2
		p =	9	9	9	9
			Matl 1	Matl 2	Matl 3	Matl 4
		T1 =	454.550	619.150	668.150	889.350
		T2 =	22964.008	42596.943	49727.848	87903.503
		T4 =	0.89500	0.66500	9.81500	2.74500
Calc 1	(Sr)^2	= T4/p =	0.0994	0.0739	1.0906	0.3050
	(SL)^2 = {[p T2 – (T1)^2]/p(p – 1)} – [(Sr)^2/2]					
Calc 2	(SL)^2 =		0.7887	0.3208	15.0965	2.4600
	(SR)^2 = (SL)^2 + (Sr)^2					
Calc 3	(SR)^2 =		0.8881	0.3947	16.1870	2.7650
	r = 2.8 [(Sr)^2]^0.5 = Repeatability					
Calc 4	r =		0.883	0.761	2.924	1.546
	R = 2.8 [(SR)^2]^0.5 = Reproducibility					
Calc 5	R =		2.64	1.76	11.27	4.66
Material Averages			50.51	68.79	74.24	98.82
Standard Deviation, Sr =			0.315	0.272	1.044	0.552
Standard Deviation, SR =			0.942	0.628	4.023	1.663
	Relative Precision:		Matl 1	Matl 2	Matl 3	Matl 4
	(r) ==>		1.75	1.11	3.94	1.56
	(R) ==>		5.22	2.56	15.17	4.71
	Step 1: Outliers at 5% Significance Level for Materials 1 to 4					
			Matl 1	Matl 2	Matl 3	Matl 4
	For h :	Lab #	9	1	9	9
	For k :	Lab #	4	none	4	4
	Step 2 : Outliers at 2% Significance Level for Materials 1 to 4					
			Matl 1	Matl 2	Matl 3	Matl 4
	For h :	Lab #	none	none	none	8
	For k :	Lab #	1	none	none	none

TABLE A6.22 (1-R1-OD): Mooney Viscosity–Revised Data: Outliers 5 % Sig Outliers Removed

	Material 1		Material 2		Material 3		Material 4	
Lab #	Day1	Day2	Day1	Day2	Day1	Day2	Day1	Day2
1	48.8	49.9			72.3	74.2	100.0	99.5
2	51.0	51.0	68.0	68.5	69.0	70.0	97.5	98.0
3	50.4	49.9	68.1	68.6	72.6	73.6	98.7	99.6
4			68.0	68.0				
5	50.3	50.1	68.5	68.5	76.0	77.1	100.2	100.4
6	52.4	52.3	69.5	69.0	80.4	82.3	99.0	99.1
7	50.8	50.8	69.5	69.4	71.8	72.4	98.9	99.4
8	51.0	51.0	69.0	68.5	76.0	76.0	104.0	103.0
9			69.0	68.6				
Day Avg	50.67	50.71	68.70	68.64	74.01	75.09	99.76	99.86
2 Day Avg		50.69		68.67		74.55		99.81
Bet Lab SDev	1.08	0.86	0.64	0.41	3.73	3.94	2.07	1.56
Pooled Bet Lab SDev		0.97		0.54		3.84		1.83

TABLE A6.23 (2-R1-OD): Cell Averages and Averages Sqaured: 5 % Outliers Removed

Cell Averages					Cell Averages Squared				
Lab #	Matl 1	Matl 2	Matl 3	Matl 4	Lab #	Matl 1	Matl 2	Matl 3	Matl 4
1	49.35		73.25	99.75	1	2435.42		5365.56	9950.06
2	51.00	68.25	69.50	97.75	2	2601.00	4658.06	4830.25	9555.06
3	50.15	68.35	73.10	99.15	3	2515.02	4671.72	5343.61	9830.72
4		68.00			4		4624.00		
5	50.20	68.50	76.55	100.30	5	2520.04	4692.25	5859.90	10060.09
6	52.35	69.25	81.35	99.05	6	2740.52	4795.56	6617.82	9810.90
7	50.80	69.45	72.10	99.15	7	2580.64	4823.30	5198.41	9830.72
8	51.00	68.75	76.00	103.50	8	2601.00	4726.56	5776.00	10712.25
9		68.80			9		4733.44		
T1 =	354.850	549.350	521.850	698.650	T2 =	17993.648	37724.903	38991.558	69749.813
Cell Avg	50.69	68.67	74.55	99.81					
Var Cell Avg	0.8812	0.2464	14.6067	3.2587					
SDev Cell Avg	0.939	0.496	3.822	1.805					

Note: variance cell avg = S^2 (Yav)

TABLE A6.24 (3-R1-OD): Cell Average Deviation, d and h-values: 5 % Outliers Removed

	Cell Deviations, d					Cell h-values				
	Lab #	Matl 1	Matl 2	Matl 3	Matl 4	Lab #	Matl 1	Matl 2	Matl 3	Matl 4
	1	−1.34		−1.30	−0.06	1	−1.43		−0.34	−0.03
	2	0.31	−0.42	−5.05	−2.06	2	0.33	−0.84	−1.32	−1.14
	3	−0.54	−0.32	−1.45	−0.66	3	−0.58	−0.64	−0.38	−0.36
	4		−0.67			4		−1.35		
	5	−0.49	−0.17	2.00	0.49	5	−0.53	−0.34	0.52	0.27
	6	1.66	0.58	6.80	−0.76	6	1.77	1.17	1.78	−0.42
	7	0.11	0.78	−2.45	−0.66	7	0.11	1.57	−0.64	−0.36
	8	0.31	0.08	1.45	3.69	8	0.33	0.16	0.38	***2.05***
	9		0.13			9		0.26		
						h(crit) 2% Sig Level at indicated p				
All Lab Cell Avg		50.37	68.83	73.52	98.58	p =	7	8	7	7
SDev Cell Avg		0.939	0.496	3.822	1.805	h(crit)	1.89	1.95	1.89	1.89
						Lab#>h (crit)	none	none	none	8

Bold and italic = significant values
h = d/S (Yav); where d = avg Cell i – avg All Cells, S (Yav) = std dev of Cell avgs

TABLE A6.25 (4R-R1-OD): Cell Range, Range Squared: 5 % Outliers Removed

	Cell Ranges					Cell Ranges Squared				
	Lab #	Matl 1	Matl 2	Matl 3	Matl 4	Lab #	Matl 1	Matl 2	Matl 3	Matl 4
	1	1.100		1.900	0.500	1	1.210		3.610	0.250
	2	0.000	0.500	1.000	0.500	2	0.000	0.250	1.000	0.250
	3	0.500	0.500	1.000	0.900	3	0.250	0.250	1.000	0.810
	4		0.000			4	0.000	0.000	0.000	0.000
	5	0.200	0.000	1.100	0.200	5	0.040	0.000	1.210	0.040
	6	0.100	0.500	1.900	0.100	6	0.010	0.250	3.610	0.010
	7	0.000	0.100	0.600	0.500	7	0.000	0.010	0.360	0.250
	8	0.000	0.500	0.000	1.000	8	0.000	0.250	0.000	1.000
	9		0.400			9		0.160		
Avg Range		0.271	0.313	1.071	0.529	**T3 =**	1.5100	1.1700	10.7900	2.6100

T3 = Sum Cell 'Ranges Squared'

TABLE A6.26 (4S-R1-OD): Cell Standard Deviation and Variance: 5 % Sig Outliers Removed

	Cell Std Deviations					Cell Variances				
	Lab #	Matl 1	Matl 2	Matl 3	Matl 4	Lab#	Matl 1	Matl 2	Matl 3	Matl 4
	1	0.778		1.344	0.354	1	0.6050		1.8050	0.1250
	2	0.000	0.354	0.707	0.354	2	0.0000	0.1250	0.5000	0.1250
	3	0.354	0.354	0.707	0.636	3	0.1250	0.1250	0.5000	0.4050
	4		0.000			4		0.0000		
	5	0.141	0.000	0.778	0.141	5	0.0200	0.0000	0.6050	0.0200
	6	0.071	0.354	1.344	0.071	6	0.0050	0.1250	1.8050	0.0050
	7	0.000	0.071	0.424	0.354	7	0.0000	0.0050	0.1800	0.1250
	8	0.000	0.354	0.000	0.707	8	0.0000	0.1250	0.0000	0.5000
	9		0.283			9		0.0800		
Pooled SDev		0.328	0.270	0.878	0.432	**T4 =**	0.75500	0.58500	5.39500	1.30500
					Pooled Variance		0.1079	0.0731	0.7707	0.1864

TABLE A6.27 (5-R1-OD): k-values: 5 % Sig Outliers Removed

Lab #	Matl 1	Matl 2	Matl 3	Matl 4
1	***2.37***		1.53	0.82
2	0.00	1.31	0.81	0.82
3	1.08	1.31	0.81	1.47
4		0.00		
5	0.43	0.00	0.89	0.33
6	0.22	1.31	1.53	0.16
7	0.00	0.26	0.48	0.82
8	0.00	1.31	0.00	1.64
9		1.05		
Pooled SDev	0.328	0.270	0.878	0.432
k(crit) 2% Sig Level at n=2, indicated p ;				
P =	7	8	7	7
k(crit) =	1.90	1.90	1.90	1.90
Lab#>k(crit)	1	none	none	none

Bold and italic = significant values
k = S(i) / Sr; where S(i) = indiv cell std dev, Sr = pooled all lab std dev

TABLE A6.28 (6-R1-OD): Mooney Viscosity Calculations for Precision Outliers 5 % Sig Level Removed

		ITP for n =	2	2	2	2
		p =	7	8	7	7
			Matl 1	Matl 2	Matl 3	Matl 4
		T1 =	354.850	549.350	521.850	698.650
		T2 =	17993.648	37724.903	38991.558	69749.813
		T4 =	0.75500	0.58500	5.39500	1.30500
Calc 1	(Sr)^2	= T4 /p =	0.1079	0.0731	0.7707	0.1864
	(SL)^2	= { [p T2 – (T1)^2] / p (p – 1)} – [(Sr)^2 / 2]				
Calc 2		(SL)^2 =	0.8273	0.2098	14.2213	3.1655
	(SR)^2 = (SL)^2 + (Sr)^2					
Calc 3		(SR)^2 =	0.9351	0.2829	14.9920	3.3519
	r = 2.8 [(Sr)^2]^0.5 = Repeatability					
Calc 4		r =	0.920	0.757	2 .458	1.209
	R = 2.8 [(Sr)^2]^0.5 = Reproducibility					
Calc 5		R =	2.71	1.49	10.84	5.13
Material Averages			50.69	68.67	74.55	99.81
Standard Deviation, Sr =			0.328	0.270	0.878	0.432
Standard Deviation, SR =			0.967	0.532	3.872	1.831
Relative Precision:			Matl 1	Matl 2	Matl 3	Matl 4
		(r) ==>	1.81	1.10	3.30	1.21
		(R) ==>	5.34	2.17	14.54	5.14
Step 1:		Outliers at 5% Significance Level for Materials 1 to 4				
			Matl 1	Matl 2	Matl 3	Matl 4
For h :		Lab #	9	1	9	9
For k :		Lab #	4	none	4	4
Step 2:		Outliers at 2% Significance Level for Materials 1 to 4				
			Matl 1	Matl 2	Matl 3	Matl 4
For h:		Lab #	none	none	none	8
For k:		Lab #	1(a)	none	none	none

(a) Note: Cell values for Lab 1 Material 1 not deleted for 2 % significant k-value. See Annex A6 for discussion.

TABLE A6.29 (1-R2-OD): Mooney Viscosity Revised Data: 2 % Sig Outliers Removed

	Material 1		Material 2		Material 3		Material 4	
Lab #	Day1	Day2	Day1	Day2	Day1	Day2	Day1	Day2
1	48.8	49.9			72.3	74.2	100.0	99.5
2	51.0	51.0	68.0	68.5	69.0	70.0	97.5	98.0
3	50.4	49.9	68.1	68.6	72.6	73.6	98.7	99.6
4			68.0	68.0				
5	50.3	50.1	68.5	68.5	76.0	77.1	100.2	100.4
6	52.4	52.3	69.5	69.0	80.4	82.3	99.0	99.1
7	50.8	50.8	69.5	69.4	71.8	72.4	98.9	99.4
8	51.0	51.0	69.0	68.5	76.0	76.0		
9			69.0	68.6				
Day Avg	50.67	50.71	68.70	68.64	74.01	75.09	99.05	99.33
2 Day Avg		50.69		68.67		74.55		99.19
Bet Lab SDev	1.08	0.86	0.64	0.41	3.73	3.94	0.98	0.78
Pooled Bet Lab SDev		0.97		0.54		3.84		0.89

Note: Lab 1 Material 1, 2 % significant k-value outlier not removed. See Annex A6 for discussion.

TABLE A6.30 (2-R2-OD): Cell Averages and Averages Squared: 2 % Outliers Removed

Cell Averages					Cell Averages Squared				
Lab #	Matl 1	Matl 2	Matl 3	Matl 4	Lab #	Matl 1	Matl 2	Matl 3	Matl 4
1	49.35		73.25	99.75	1	2435.42	0.00	5365.56	9950.06
2	51.00	68.25	69.50	97.75	2	2601.00	4658.06	4830.25	9555.06
3	50.15	68.35	73.10	99.15	3	2515.02	4671.72	5343.61	9830.72
4		68.00			4		4624.00		
5	50.20	68.50	76.55	100.30	5	2520.04	4692.25	5859.90	10060.09
6	52.35	69.25	81.35	99.05	6	2740.52	4795.56	6617.82	9810.90
7	50.80	69.45	72.10	99.15	7	2580.64	4823.30	5198.41	9830.72
8	51.00	68.75	76.00		8	2601.00	4726.56	5776.00	
9		68.80			9		4733.44		
T1 =	354.850	549.350	521.850	595.150	**T2 =**	17993.648	37724.903	38991.558	59037.563
Cell Avg	50.69	68.67	74.55	99.19					
Var Cell Avg	0.8812	0.2464	14.6067	0.7284					
SDev Cell Avg	0.939	0.496	3.822	0.853					

Note: variance cell avg = S^2(Yav)

TABLE A6.31 (3-R2-OD): Cell Average Deviations, d and h-values: 2 % Sig Outliers Removed

	Cell Deviations, d				Cell h-values				
	Lab #	Matl 1	Matl 2	Matl 3	Matl 4	Lab #	Matl 1	Matl 2	Matl 3	Matl 4
	1	−1.34		−1.30	0.56	1	−1.43		−0.34	0.65
	2	0.31	−0.42	−5.05	−1.44	2	0.33	−0.84	−1.32	−1.69
	3	−0.54	−0.32	−1.45	−0.04	3	−0.58	−0.64	−0.38	−0.05
	4		−0.67			4		−1.35		
	5	−0.49	−0.17	2.00	1.11	5	−0.53	−0.34	0.52	1.30
	6	1.66	0.58	6.80	−0.14	6	1.77	1.17	1.78	−0.17
	7	0.11	0.78	−2.45	−0.04	7	0.11	1.57	−0.64	−0.05
	8	0.31	0.08	1.45		8	0.33	0.16	0.38	
	9		0.13			9		0.26		
						h(crit) 2%Sig Level at indicated p				
All Lab										
Cell Avg		50.69	68.67	74.55	99.19	p =	7	8	7	6
SDev Cell Avg		0.939	0.496	3.822	0.853	h(crit)	1.89	1.95	1.89	1.80
						Lab#>h(crit)	NA	NA	NA	NA

Bold and italic = significant values
h = d / S (Yav); where d = avg Cell i – avg All Cells, S (Yav) = std dev of Cell avgs

TABLE A6.32 (4R-R2-OD): Cell Ranges and Ranges Squared: 2 % Outliers Removed

	Cell Ranges				Cell Ranges Squared				
	Lab #	Matl 1	Matl 2	Matl 3	Matl 4	Lab #	Matl 1	Matl 2	Matl 3	Matl 4
	1	1.100		1.900	0.500	1	1.210		3.610	0.250
	2	0.000	0.500	1.000	0.500	2	0.000	0.250	1.000	0.250
	3	0.500	0.500	1.000	0.900	3	0.250	0.250	1.000	0.810
	4		0.000			4		0.000		
	5	0.200	0.000	1.100	0.200	5	0.040	0.000	1.210	0.040
	6	0.100	0.500	1.900	0.100	6	0.010	0.250	3.610	0.010
	7	0.000	0.100	0.600	0.500	7	0.000	0.010	0.360	0.250
	8	0.000	0.500	0.000		8	0.000	0.250	0.000	
	9		0.400			9		0.160		
Avg Range		0.271	0.313	1.071	0.450	T3 =	1.5100	1.1700	10.7900	1.6100

T3 = Sum Cell 'Ranges Squared'

TABLE A6.33 (4S-R2-OD): Cell Standard Deviations and Variances: 2 % Outliers Removed

	Cell Std Deviations					Cell Variances				
	Lab #	Matl 1	Matl 2	Matl 3	Matl 4	Lab #	Matl 1	Matl 2	Matl 3	Matl 4
	1	0.778		1.344	0.354	1	0.6050		1.8050	0.1250
	2	0.000	0.354	0.707	0.354	2	0.0000	0.1250	0.5000	0.1250
	3	0.354	0.354	0.707	0.636	3	0.1250	0.1250	0.5000	0.4050
	4		0.000			4		0.0000		
	5	0.141	0.000	0.778	0.141	5	0.0200	0.0000	0.6050	0.0200
	6	0.071	0.354	1.344	0.071	6	0.0050	0.1250	1.8050	0.0050
	7	0.000	0.071	0.424	0.354	7	0.0000	0.0050	0.1800	0.1250
	8	0.000	0.354	0.000		8	0.0000	0.1250	0.0000	
	9		0.283			9		0.0800		
Pooled SDev		0.328	0.270	0.878	0.366	T4 =	0.75500	0.58500	5.39500	0.80500
				Pooled Variance			0.1079	0.0731	0.7707	0.1342

TABLE A6.34 (5-R2-OD): k-values: 2 % Sig Outliers Removed

Lab #	Matl 1	Matl 2	Matl 3	Matl 4
1	**2.37**		1.53	0.97
2	0.00	1.31	0.81	0.97
3	1.08	1.31	0.81	1.74
4		0.00		
5	0.43	0.00	0.89	0.39
6	0.22	1.31	1.53	0.19
7	0.00	0.26	0.48	0.97
8	0.00	1.31	0.00	
9		1.05		
Pooled SDev	0.328	0.270	0.878	0.366
k(crit) 2% Sig Level at n=2 , indicated p ;				
p =	7	8	7	6
k(crit) =	2.04	2.07	2.04	2.00
Lab#>k(crit)	NA	NA	NA	NA

TABLE A6.35 (6-R2-OD): Mooney Viscosity Calculations for Precision: Final Results

	n =	2	2	2	2
	p =	7	8	7	6
		Matl 1	Matl 2	Matl 3	Matl 4
	T1 =	354.850	549.350	521.850	595.150
	T2 =	17993.648	37724.903	38991.558	59037.563
	T4 =	0.75500	0.58500	5.39500	0.80500
Calc 1	(Sr)^2 = T4 / p =	0.1079	0.0731	0.7707	0.1342
	(SL)^2 = {[p T2 – (T1)^2] / p (p – 1)} – [(Sr)^2 / 2]				
Calc 2	(SL)^2 =	0.8273	0.2098	14.2213	0.6613
	(SR)^2 = (SL)^2 + (Sr)^2				
Calc 3	(SR)^2 =	0.9351	0.2829	14.9920	0.7955
	r = 2.8 [(Sr)^2]^0.5 = Repeatability				
Calc 4	r =	0.920	0.757	2.458	1.026
	R = 2.8 [(SR)^2]^0.5 = Reproducibility				
Calc 5	R =	2.71	1.49	10.84	2.50
Material Averages		50.69	68.67	74.55	99.19
Standard Deviation, Sr =		0.328	0.270	0.878	0.366
Standard Deviation, SR =		0.967	0.532	3.872	0.892
Relative Precision:		Matl 1	Matl 2	Matl 3	Matl 4
(r) ==>		1.81	1.10	3.30	1.03
(R) ==>		5.34	2.17	14.54	2.52
Step 1: Outliers at 5% Significance Level for Materials 1 to 4					
		Matl 1	Matl 2	Matl 3	Matl 4
For h :	Lab #	9	1	9	9
For k :	Lab #	4	none	4	4
Step 2: Outliers at 2% Significance Level for Materials 1 to 4					
		Matl 1	Matl 2	Matl 3	Matl 4
For h :	Lab #	none	none	none	8
For k :	Lab #	1(a)	none	none	none

(a) Note: Cell values for Lab 1 Material 1 not deleted for 2 % sig k-value. See Annex A6 for discussion.

TABLE A6.36 Replacement Values for Outliers at Both 5 % and 2 % Significance level

Part A : AOT Cell - Parameter Replacement Values, PRV				
1. AOT : PRV for Cell Average Outliers				
1		69.7 (0.30)		
8				***101.2 (1.00)***
9	49.4 (0.20)		69.0 (2.00)	96.5 (1.80)
Note: Cell PRV Cell Averages listed with individual Cell Range in ( )				
2. AOT : PRV for Cell Range Outliers				
1	***0.80 (49.35)***			
4	0.85 (50.25)		2.20 (77.25)	1.20 (96.50)
Note: Cell PRV Cell Ranges listed with individual Cell Averages in ( )				

Part B : AOT Cell - Data Replacement Values, DRV				
3. AOT: DRV for Cell Average Outliers				
Lab Number	**Matl 1**	**Matl 2**	**Matl 3**	**Matl 4**
1		69.6, 70.0		
8				***100.7, 101.7***
9	49.3, 49.5		68.0, 70.0	95.6, 97.4
4. AOT : DRV for Cell Range Outliers				
Lab Number	**Matl 1**	**Matl 2**	**Matl 3**	**Matl 4**
1	***49.8, 49.0***			
4	49.8, 50.7		76.2, 78.4	95.9, 97.1

Note: 2% sig level AOT replacements in bold and italic.

TABLE A6.37 Comparison of Outlier Handling Procedures

Part 1 Outlier Procedure	Repeatability, r				Pooled Precision
	Matl 1	Matl 2	Matl 3	Matl 4	
Original Database (No Outliers Deleted)	1.29	0.74	3.43	2.54	2.26
AOT Outlier Replacement, Option 2 (a)	0.88	0.76	2.92	1.55	1.75
Outliers Deleted, Option 1 (a)	0.92	0.76	2.46	1.03	1.46

Part 2 Outlier Procedure	Reproducibility, R				Pooled Precision
	Matl 1	Matl 2	Matl 3	Matl 4	
Original Database (No Outliers Deleted)	3.37	1.97	15.15	8.84	8.98
AOT Outlier Replacement, Option 2 (a)	2.64	1.76	11.27	4.66	6.30
Outliers Deleted, Option 1 (a)	2.71	1.49	10.84	2.50	5.77

(a) Final precision results.
Note: See Table A6.36 for Materials (and Labs) with Outliers.

TABLE A6.38 Relative Reduction Factors: Precision Parameters, r and R

Part 1 Outlier Procedure	Reduction Factor for Repeatability, r				Pooled Precision
	Matl 1	Matl 2	Matl 3	Matl 4	
Original Database (No Outliers Deleted)	1.0	1.0	1.0	1.0	1.0
AOT Outlier Replacement, Option 2 (a)	0.68	1.0	0.85	0.61	0.78
Outliers Deleted, Option 1 (a)	0.71	1.0	0.72	0.41	0.65

Part 2 Outlier Procedure	Reduction Factor for Reproducibility, R				Pooled Precision
	Matl 1	Matl 2	Matl 3	Matl 4	
Original Database (No Outliers Deleted)	1.0	1.0	1.0	1.0	1.0
AOT Outlier Replacement, Option 2 (a)	0.78	0.89	0.74	0.53	0.70
Outliers Deleted, Option 1 (a)	0.80	0.76	0.72	0.28	0.64

(a) Final precision results.
Reduction factor = (Prec Revised Database / Prec Original Database)
Note: See Table A6.36 for Materials (and Labs) with Outliers.

TABLE A6.39 Precision for Mooney Viscosity

		Within Laboratories			Between Laboratories			
Material	**Mean Level**	**Sr**	**r**	**(r)**	**SR**	**R**	**(R)**	**No. Labs (b)**
1 - SBR 1712	50.7	0.328	0.92	1.81	0.967	2.71	5.35	7
2 - IIR (Butyl)	68.7	0.270	0.76	1.10	0.532	1.49	2.17	8
3- SBR-BMB	74.6	0.878	2.46	3.29	3.87	10.84	14.5	7
4- NR	99.2	0.366	1.03	1.03	0.892	2.50	2.52	6
Pooled or Avg Val (a)	72.9	0.328	0.918	1.26	0.819	2.29	3.14	

(a) Pooled values calculated for Materials 1, 2, and 4 only; SBR-BMB omitted. See A6.8.4.1 for details.
(b) Number of labs after outliers deleted, (Option 1); 3 step analysis.
Notation used: Sr = within-laboratory standard deviation (in measurement units)
r = repeatability (in measurement units)
(r) = repeatability (in percent of mean level)
SR = between-laboratory standard deviation (for total between laboratory variation in measurement units)
R = reproducibility (in measurement units)

REFERENCES

(1) Youden, W. J., "Graphical Analysis of Interlaboratory Test Results," *Industrial Quality Control*, 24-8, May 1959 .

(2) Youden, W. J., and Steiner, E. H., *Statistical Manual of the Association of Official Analytical Chemists*, AOAC Washington DC, 1975.

(3) Veith, A. G., "Precision in Polymer Testing: An Important World-Wide Issue," *Polymer Testing*, Vol 7, 1987, pp. 239-267.

(4) Veith, A. G., "A New Approach to Evaluating Inter-Laboratory Testing Precision," *Polymer Testing*, Vol 12, 1993, pp. 113-184.

(5) ASTM Customer Service, 100 Barr Harbor Dr, W. Conshohocken, PA 19428, USA; Phone: 610-832-9585, Fax: 610-832-9555, web site: www.astm.org.