Designation: D4483 – 20

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 Rubber-like Materials and is the direct responsibility of Subcommittee D11.16 on Application of Statistical Methods.

Current edition approved June 1, 2020. Published June 2020. Originally approved in 1985. Last previous edition approved in 2018 as D4483 – 18. DOI: 10.1520/D4483-20.

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, health, and environmental practices and determine the applicability of regulatory limitations prior to use.*

1.7 *This international standard was developed in accordance with internationally recognized principles on standardization established in the Decision on Principles for the Development of International Standards, Guides and Recommendations issued by the World Trade Organization Technical Barriers to Trade (TBT) Committee.*

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.

[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.

3.2.3 *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.

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

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

composite material, and on samples of this, testing is conducted and the property response of the target material is evaluated.

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

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

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