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 Σ 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