Standard Practice for
Rubber—Preparation, Testing, Acceptance, Documentation, and Use of Reference Materials[1]

1. Scope

1.1 This practice covers materials used on an industry-wide basis as reference materials, which are vitally important to conduct product, specification, and development testing in the rubber industry. This practice describes the steps necessary to ensure that any candidate material, that has a perceived need, can become a Reference Material. The practice sets forth the recommendations on the preparation steps for these materials, on the testing that shall be conducted to permit acceptance of any candidate material, and on how the documentation needed for the acceptance shall be recorded for future use and review.

1.2 This practice shall be administered by ASTM Committee D11.

1.2.1 Important sections of this practice are as follows:

	Section
Significance and Use	3
Preparation of Industry Reference Materials	4
Overview of Industry Reference Material Testing	5
Chemical and Physical Specifications for IRM	6
Reference Material Documentation	7
Typical Reference Material Use	8
Recommended Package Size for IRM	Annex A1
Recommended Sampling Plans for Homogeneity Testing of an IRM	Annex A2
Test Plan and Analysis for Homogeneity of an IRM	Annex A3
Test Plan and Analysis to Evaluate an Accepted Reference Value	Annex A4
Statistical Model(s) for IRM Testing	Annex A5
Example of Annex Calculations for a Typical IRM	Appendix X1
Two-Way Analysis of Variance for Calculating Sr	Appendix X2

1.3 *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]

D4483 Practice for Evaluating Precision for Test Method Standards in the Rubber and Carbon Black Manufacturing Industries

D5900 Specification for Physical and Chemical Properties of Industry Reference Materials (IRM)

E122 Practice for Calculating Sample Size to Estimate, With Specified Precision, the Average for a Characteristic of a Lot or Process

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

E826 Practice for Testing Homogeneity of a Metal Lot or Batch in Solid Form by Spark Atomic Emission Spectrometry

3. Significance and Use

3.1 Reference materials are vitally important in product and specification testing, in research and development work, in technical service work, and in quality control operations in the rubber industry. They are especially valuable for referee purposes.

3.2 *Categories, Classes, and Types of Reference Materials (RM):*

3.2.1 Reference materials are divided into two categories:

3.2.1.1 *Industry Reference Materials (IRM)*—Materials that have been prepared according to a specified production process to generate a uniform lot; the parameters that define the quality of the lot are evaluated by a specified measurement program.

3.2.1.2 *Common-Source Reference Materials (CRM)*—Materials that have been prepared to be as uniform as possible but do not have established property (parameter) values; the knowledge of a common or single source is sufficient for certain less critical applications.

3.2.2 Industry reference materials (IRMs) are divided into additional classes and types according to the method of evaluating the lot parameters and according to the production process for generating the lot material. These are explained more fully (refer to Annex A3 and Annex A4 for more details on the discussion in Section 3).

3.2.3 The following lot parameters are important for reference material use:

3.2.3.1 *Accepted Reference Value (AR Value)*—An average IRM property or parameter value established by way of a specified test program.

3.2.3.2 *Test Lot Limits (TL Limits)*—These are limits defined as ±3 times the standard deviation of individual IRM test results across the entire lot for the property or parameter(s) that defines lot quality; the measurements are conducted in the laboratory of the organization producing the IRM.

3.2.3.3 Although the limits as defined in 3.2.3.2 are given in terms of ±3 times the standard deviation, the rejection of individual portions of the lot as being outlier or non-typical portions in assessing the homogeneity of the lot is done on the basis of ±2 times the appropriate standard deviation, that is, on the basis of a 95 % confidence interval. See Annex A3 and Annex A4 for more information and the evaluation procedures.

3.2.4 All IRMs have an AR value and TL limits; however the AR value may be obtained in one of two ways to produce one of two classes of AR values:

3.2.4.1 *Global AR Value*—This AR value is obtained from an interlaboratory test program where the word "global" indicates an average value across many laboratories.

3.2.4.2 *Local AR Value*—This is an AR value obtained in one laboratory or at one location, usually the laboratory responsible for preparation of the homogeneous lot.

3.2.5 An additional parameter is of importance for IRMs that have a global AR value:

3.2.5.1 *Between-Laboratory Limits (BL)*—The group of laboratories that conduct interlaboratory testing to establish an AR-value are not equivalent to a system or population typical of industrial production operations that use the usual ±3 standard deviation limits. Such production operations are systems that have been purged of all assignable causes of variation and are in a state of 'statistical control' with only random variations that cannot be removed. Thus, the recommended limits on all IRMs are the ±2 standard deviation limits that pertain to a 95 % confidence level. If for serious reasons that can be totally justified, ±3 standard deviation limits are required, these may be used provided that full and complete documentation is supplied to justify the limits.

3.2.6 The homogeneity or uniformity of the lot, which determines the magnitude of the TL limits, may be designated as one of two different levels of uniformity. The key factor that determines the level of uniformity is the capability of blending the IRM portions or parts that constitute the lot, to ensure a high degree of uniformity from the blending process. IRMs that cannot be blended will have an extra residual amount of variation (portion to portion) that lowers the level of uniformity.

3.2.6.1 *Uniformity Level 1 (UL-1)*—This is the most uniform or highest level of homogeneity that can be attained by

the use of *a specified test* for measuring the parameter that defines lot quality; it is obtained by the use of a blended material and is referred to as a Type *B* (*B* = blended) IRM.

3.2.6.2 *Uniformity Level 2 (UL-2)*—This is the lesser degree of uniformity that is attained by the use of *a specified test* for measuring the parameter that defines lot quality; it is normally obtained for non-blended materials and is referred to as a Type *NB* (not blended) IRM.

3.3 IRMs have a number of use applications in the technical areas, as cited in 3.1.

3.3.1 *Single Laboratory Self Evaluation*—The IRM may be used in a given laboratory (or with a given test system) to compare the test results within the laboratory to the accepted reference value for the IRM. An IRM can also be used for internal statistical quality control (SQC) operations.

3.3.2 *Multi-Laboratory Evaluation*—The IRM may be used between two or more laboratories to determine if the test systems in the laboratories are operating within selected control limits.

3.3.3 One or more IRMs may be used in the preparation of compounds to be used for evaluating non-reference materials in compound testing and performance.

3.3.4 Reference liquid IRMs may be used for immersion testing of various candidate or other reference compounds. Such immersion testing is important due to the deleterious influences of immersion liquids on rubber compounds.

3.3.5 IRMs may also be used to eliminate interlaboratory testing variation known as "test bias:" a difference between two (or more) laboratories that is essentially constant between the laboratories for a given test property level, irrespective of the time of the test comparisons. In such applications a differential test measurement value, (IRM − experimental material), becomes a corrected test result; this corrected value is used as the measure of performance rather than the "as-measured" test value on the experimental material of interest.

3.4 Average values play an important role in various operations and decisions in this practice. For this practice, "average" is defined as the arithmetic mean.

3.5 The various characteristics of IRMs and CRMs (categories, classes, types) are listed in summary form in Table 1.

3.6 This practice and the IRM program it describes was developed to replace a standardization program conducted by

TABLE 1 Categories of Reference Materials[A]

	IRM				CRM
AR Value	Global		Local		None
Homogeneity (TL Limits)	Type *B* (UL-1)	Type *NB* (UL-2) or (UL-1)	Type *B* (UL-1)	Type *NB* (UL-2) or (UL-1)	Single Source Material

[A] AR value = accepted reference value.
  TL limits = test lot limits.
  Global = AR value obtained from an interlaboratory test program.
  Local = AR value obtained from one laboratory.
  Type-B = IRM that has been blended to ensure high uniformity.
  Type-NB = IRM that cannot be blended.
  UL-1 and UL-2 = levels of uniformity in the IRM lot; UL-1 is higher uniformity than UL-2.
  See Annex A3 and Annex A4 for more information.

the National Institute of Standards and Technology (NIST) that began in 1948 and has been phased out.

3.7 It is not feasible to write into this practice all the necessary specifications, modes of preparation, sampling, and testing protocols, for the wide variety of materials that will eventually become IRM. Therefore this practice is published to give general guidelines for IRMs.

3.8 A permanent IRM Steering Committee within Subcommittee D11.20 shall be constituted by Subcommittee D11.90 to assist in the utilization of this practice and to make technical and, where required, policy decisions regarding the preparation and administration of IRM.

4. Preparation of Industry Reference Materials

4.1 *Basic Preparation Steps:*

4.1.1 An IRM should be prepared in a way that ensures that the entire quantity or lot of the material is as homogeneous, in composition and vital performance properties, as is possible.

4.1.2 For particulate and liquid materials this implies a thorough physical blending operation during or after the manufacturing steps, or both.

4.1.3 For materials not easily blended after manufacture, two options to ensure homogeneity are recommended:

4.1.3.1 Use highly homogeneous components or other materials that are required in the manufacturing steps or conduct certain blending operations at intermediate manufacturing steps to ensure maximum homogeneity.

4.1.3.2 Use intensive statistical quality control procedures to ensure a specified degree of homogeneity among the packets, bales, or other discrete units of the material.

4.1.4 Examples, as cited in 4.1.3.1, are such materials as accelerators, antioxidants, sulfur, and reference test (liquid) fuels.

4.1.5 Examples, as cited in 4.1.3.2, are various synthetic rubbers.

4.2 *Packaging of Industry Reference Materials:*

4.2.1 Industry reference materials should be packaged preferably in small quantities or packages. The packages shall be consecutively numbered as they are filled. Nominally the size should be the smallest amount that the average user of the material would require for normal volume testing. High volume users could therefore order multiple package lots. The use of such minimum volume (mass) packages will of course vary, but Annex A1 gives recommended masses or volumes.

4.2.2 Industry reference materials shall be suitably packaged to prevent or retard the change of IRM values with the passage of time or inadvertent exposure to heat, light, moisture, or combinations thereof, in normal storage. The stringency of this requirement varies with the type of IRM. All precautions shall be taken to make IRMs as stable as possible.

4.2.3 Packages shall be dispensed by the manufacturing or distribution organization with a document that shall furnish the following general information:

4.2.3.1 Name and number of the IRM,

4.2.3.2 Name of the manufacturer,

4.2.3.3 Date of manufacture or preparation,

4.2.3.4 Storage conditions, and

4.2.3.5 Reference to ASTM research report for documentation of testing.

4.2.4 For each test property measured to assess lot quality report the following:

4.2.4.1 Accepted reference value,

4.2.4.2 Test lot limits, and

4.2.4.3 Between-laboratory limits.

4.3 *Packaging of Common–Source Reference Materials:*

4.3.1 CRMs shall be packaged and dispensed in the same manner as for IRMs. Each CRM package shall be furnished with a documentation sheet with the following information:

4.3.1.1 Name and number of the CRM,

4.3.1.2 Name of manufacturer,

4.3.1.3 Date of manufacture or preparation,

4.3.1.4 Storage conditions, and

4.3.1.5 Reference to ASTM research report.

5. Overview of Industry Reference Material Testing

5.1 Testing is conducted to *(1)* demonstrate the uniformity of the IRM lot to some selected limits and evaluate the test lot limits, and *(2)* to establish an accepted reference value for the lot and as a secondary goal to evaluate the between-laboratory limits for interlaboratory testing of the IRM where this is applicable.

5.2 *Testing for Homogeneity:*

5.2.1 Homogeneity testing is ideally conducted in one highly qualified laboratory, which is usually the laboratory of the organization that produces the IRM. The lot size is determined and samples are drawn from the lot. Guidance for the size and number of samples is given in Annex A2. The samples taken from the lot are tested according to the instructions given in Annex A3. This latter annex also addresses the concept of different uniformity levels for an IRM and the importance of this in IRM development and use.

5.2.2 It is important that each sample represents a fraction or portion of the total lot that can be physically separated from the remainder of the lot, in the event that the portion represented by the sample is judged to be significantly different from the remainder of the lot and is therefore rejected.

5.2.3 Those portions of the lot that are shown to be significantly different from the remainder or bulk of the lot shall be rejected.

5.2.4 If, in the statistical analysis of Annex A3, a substantial fraction (25 to 30 %) of the lot is declared to be not acceptable for lack of homogeneity, retesting may be permitted. This retesting shall include all suspected portions and a number of accepted homogeneous portions or parts equal in number to the suspect portions. The retest shall be conducted according to Annex A3.

5.2.5 If on retesting and analysis of the newly generated data these same portions are again found to be unequal in property value to the accepted portions by standard statistical tests, they shall be rejected. If the suspected portions are found to be equal to the accepted portions in property values, they may be accepted as part of the lot.

5.3 *Testing for an Accepted Reference Value*—Testing for an accepted reference value may be undertaken once a homogeneous lot has been achieved. The detailed instructions for conducting the interlaboratory program and analyzing the data of the program for an accepted reference value are given in Annex A4. This annex also gives instructions for evaluating the between-laboratory test limits where this is applicable.

5.4 *Additional Testing Background Information:*

5.4.1 To provide some theoretical background for the analyses conducted in Annex A3 and Annex A4, a discussion on statistical model development is given in Annex A5. This permits a more comprehensive understanding of the rationale for the analysis of the IRM test data and for the use of IRMs in various laboratory applications. See also Section 8 for a detailed discussion of how IRMs may be used in laboratory applications.

5.4.2 Appendix X1 gives an example of a complete set of calculations for homogeneity and accepted reference value testing according to the instructions of Annex A2 – Annex A4.

6. Chemical and Physical Specifications for IRM

6.1 Since the chemical and physical specifications for each IRM will vary in kind and degree among the various candidate materials, the details on such information are to be referred to the IRM Steering Committee. As experience is gained this practice may be amended to include more specific guidelines and test protocols. See Specification D5900 for information on all the current IRMs and their specifications.

7. Documentation for Reference Materials

7.1 *Industry Reference Materials (IRM):*

7.1.1 A full report shall be given for each IRM. This shall contain the following information:

7.1.1.1 Name of the material,

7.1.1.2 IRM number,

7.1.1.3 Organization preparing the IRM,

7.1.1.4 Date of preparation or manufacture and testing,

7.1.1.5 Any special preparation or processing steps for the IRM,

7.1.1.6 Raw data and results of the homogeneity and accepted reference value testing,

7.1.1.7 Date of adoption of the IRM,

7.1.1.8 Names of all laboratories in the AR value program,

7.1.1.9 Specific conditions under which the IRM is to be stored while awaiting distribution to laboratories purchasing the IRM, and

7.1.1.10 Any other information of a special nature needed to document special issues not covered in the above list.

7.1.2 All of the information as called for in 7.1.1 shall be prepared in a special report that can be easily interpreted and sent to ASTM International Headquarters. This shall be given a special research report number and kept on file at ASTM.

7.2 *Common-Source Reference Materials (CRM):*

7.2.1 A report shall be prepared for all CRMs with the following information:

7.2.1.1 Name of the CRM,

7.2.1.2 Number of the CRM,

7.2.1.3 Name of organization preparing the CRM,

7.2.1.4 Date of manufacture or preparation,

7.2.1.5 Storage conditions for the material while awaiting shipment, and

7.2.1.6 Any other special information pertinent to the use of the CRM.

7.2.2 All of the information in 7.2.1 shall be provided in a report sent to ASTM and kept on file as a research report.

8. Typical Reference Material Use (Global AR Value)

8.1 *IRM Application—Single Laboratory Self Evaluation:*

8.1.1 A single laboratory can use an IRM to determine how the test or measurement system in the laboratory is performing in relation to the AR value and the limits associated with the AR value. This self-evaluation of a laboratory can be most effectively conducted by setting up a statistical model. Refer to Annex A5 for background and details.

8.1.2 In A5.3.6 of Annex A5, the model for the testing for an AR value is given in Eq A5.9 and reproduced here as Eq 1, with one new added term, $B(g)$. The term y represents the measured test result.

$$y = \mu(0) + B_m + B_L + B(g) + e(g) + e_b(s) + e_w(s) \qquad (1)$$

The new term is needed because the entire lot may be comprised of a number of portions or units that have (average) test values that span the (maximum to minimum) range of the lot. This new term, $B(g)$, is the bias component related to the particular portion or unit (of the entire lot) purchased and tested by the user or single laboratory.

8.1.3 In the model represented by Eq A5.9, the B_m and B_L terms were variable, because the system of measurement was the collection of laboratories participating in the ITP to evaluate the AR value. In the single laboratory self-evaluation model of Eq 1, the terms B_m and B_L are fixed; they represent values unique to the single laboratory. Thus there are four fixed or constant terms in the Eq 1 model: $\mu(0)$, B_m, B_L, and $B(g)$. The sum of these four terms represents the overall test measurement bias (potential or actual) for the single laboratory.

8.1.4 To determine if the single laboratory measurement system test values agree with the AR value, it is necessary to greatly reduce or eliminate the contribution of the random deviations or (e) terms to the y-value measurement. This is done by making a number of (y-value) measurements over a selected (short-term) period and taking an average of these. As outlined in Annex A5, the random deviations average out to zero in the long run, and thus do not contribute to the measured average y-value. The number of recommended measurements for this purpose is twelve, perhaps one or two per day, for six or twelve consecutive days. On this basis, $e(g) + e_b(s) + e_w(s) \cong 0$. This recommended action demonstrates the power-of-averaging rule. Once the average of twelve is calculated it can be compared to the AR value. Several outcomes are possible for this comparison.

8.1.5 *Potential Outcome 1*—The degree of agreement can be expressed by the difference between the twelve-test average, $y_{(12)}$, and the AR value. If this difference is expressed by Eq 2, where both sides represent absolute values,

$$|y_{(12)} - \text{AR value}| < \text{or} = |\,TL\ limits\,| \qquad (2)$$

then there is good agreement since $y_{(12)}$ falls within the nominal range: AR value ± TL limits. The single laboratory may be said to be operating on target and the sum of all four biases approaches zero. Note that the individual biases may not be zero; their sum is zero.

8.1.6 *Potential Outcome 2*—If the difference between $y_{(12)}$ and the AR value is expressed by Eq 3,

$$\left| y_{(12)} - \text{AR value} \right| > = \left| TL \ limits \right| \qquad (3)$$

then the single laboratory is not operating on target: the sum of the four biases is not zero. If the difference ($y_{(12)}$ − AR value) is negative, the laboratory has a negative total bias; if the difference is positive, the total bias is positive.

8.1.7 *Potential Outcome 3*—If the outcome of the comparison of $y_{(12)}$ versus the AR value is given by Eq 3, the next step is to decide if the laboratory is operating within the between-laboratory limits, which may be considered as current inter-laboratory nominal testing variation (NTV). If the difference between $y_{(12)}$ and the AR value is expressed by Eq 4,

$$\left| y_{(12)} - \text{AR value} \right| < \text{or} = \left| between - laboratory \ limits \right| \qquad (4)$$

then the single laboratory is operating the test system within the NTV limits of typical laboratories in the industry.

8.1.8 *Potential Outcome 4*—If the difference between $y_{(12)}$ and the AR value is expressed by Eq 5,

$$\left| y_{(12)} - \text{AR value} \right| > = \left| between - laboratory \ limits \right| \qquad (5)$$

then the single laboratory is not operating within the NTV limits of typical laboratories in the industry.

8.2 *IRM Application—Multi-Laboratory Evaluation:*

8.2.1 One of the most important uses of an IRM is its application to resolve questions and disputes over poor agreement for producer-consumer testing. As demonstrated above, any numerical deviation between two laboratories (when both laboratories have measured the same material) has two types of components: *Type 1*, a combined deviation component due to random measurement variations in both laboratories, and *Type 2*, a deviation component due to inter-laboratory bias. As previously demonstrated, sufficient replicate testing in each of the laboratories will reduce the random component to zero, but will not influence the inter-laboratory bias.

8.2.2 An estimate of the bias in each of the laboratories may be determined by use of an appropriate IRM. One laboratory should supply an IRM sample, taken from one portion or unit of the IRM lot, to both laboratories. Each sample should be large enough to perform at least twelve tests in each laboratory. Each laboratory performs a selected number of tests, from six to twelve, depending upon the importance of the testing dispute, over a selected short-term time period of several days. The average of these tests is defined as y_{avg}.

8.2.3 The overall bias for each laboratory is estimated by means of Eq 6.

$$\text{Estimated Overall Bias (Laboratory } i) = (y_{avg} - \text{AR value}) \qquad (6)$$

The overall IRM bias values for each laboratory can be compared and used to make decisions about resolving any potential testing problem.

8.2.4 The algebraic difference between the two laboratory overall biases is the direct bias between the two laboratories.

$$\text{Direct Inter} - \text{laboratory Bias} = \qquad (7)$$

$$\Delta \ (\text{Laboratory 2 Bias} - \text{Laboratory 1 Bias})$$

Such information can be potentially used for corrections in test data. The use of such information for correcting interlaboratory test data should be done only on the basis of a mutual agreement between the participating laboratories.

8.2.5 The procedure as outlined in 8.2.2 and 8.2.3, can be extended to any number of laboratories by modifying the procedural steps in an appropriate manner. This operation, as stated, should only be done on the basis of mutual agreement.

9. Keywords

9.1 common source reference material (CRM); industry reference material (IRM); reference material

ANNEXES

(Mandatory Information)

A1. RECOMMENDED PACKAGE SIZES FOR IRM

A1.1 *Lot Size*—A lot size of 250 to 1000 packages is recommended depending upon the anticipated usage rate.

A1.2 For IRM rubber chemicals with a bulk density similar to accelerators, antioxidants, and sulfur, a container of about 2 dm^3 (litres) is recommended, with mass adjusted to the nearest 100 g required to fill the container.

A1.3 For IRM such as carbon black fillers or other materials used in relatively high proportions in rubber, one of two containers is recommended:

A1.3.1 A 20-dm^3 (litre) container (approximately 5 gal) with mass or volume equivalent adjusted to nearest 0.5 kg.

A1.3.2 A 25-kg bag, if the material is particulate. Materials that may potentially absorb moisture or other gases (CO_2) shall be placed inside another outer container to prevent such absorption.

A1.4 For IRM such as rubbers, a package or bale of 34 kg mass (1/30 metric ton) is recommended.

A2. RECOMMENDED SAMPLING PLANS FOR HOMOGENEITY TESTING OF AN IRM

A2.1 *Introduction*—Sampling plans are required to measure the particular property of the lot that has been selected to assess the quality of the lot. Two sampling plans are given. Plan 1 gives instructions to calculate the sample size such that the maximum deviation, *E*, between the measured property average (the estimate of the lot true value) and the actual lot true value (that is, the measured average of all lot packages, items or portions), may be calculated. An advanced knowledge of the measured property standard deviation is required for this plan. Plan 2 is a less rigorous approach that may be used when it is possible to blend the lot material and achieve greater homogeneity prior to sampling. Practice E122 is used as a reference for this annex.

A2.2 *Sampling Plans for the IRM*—One of two sampling plans shall be used to sample from the lot. Sampling Plan 1 is preferred.

A2.2.1 *Sampling Plan 1*—Draw from the lot a number of samples, *n*, to satisfy the selected or desired maximum deviation, *E*, between *Xn*, the lot average using all *n* samples and *X̄a* the lot true value as defined in A2.1. The calculation is performed by using Eq A2.1. Use either the advanced estimate of the property standard deviation, *Se*, or a well-established standard deviation, *S*.

$$n = (3\ Se/E)^2 \qquad (A2.1)$$

where:

E = $\bar{X}a - \bar{X}n$, and
Se = advanced estimate of the standard deviation of the measured property that defines lot quality; this may be obtained from measurements on the lot after its manufacture or from process control data during actual production of the lot.

A2.2.2 The quantity *E* is a 3 (standard deviation) limit on *X̄n*. Thus the range, $\bar{X}n \pm E$, will contain the true value of the lot property with a 99.7 % confidence level. If there is no advanced knowledge of *Se*, it shall be estimated from the lot with a minimum of twelve samples equally spaced throughout the lot. The test results from this preliminary estimate of *Se* may be incorporated into the test results from additional lot sampling and testing, should the value of *n* exceed twelve.

A2.2.3 The deviation, *E*, may be expressed in terms of the lot standard deviation *Se* (or *S* if it is known). Table A2.1 gives a series of values of *n* that correspond to values of *E* expressed as a fraction of *Se* (or *S*).

TABLE A2.1 *E* Values for Selected Values of *n*

n	E
12	0.87 Se
16	0.75 Se
20	0.67 Se
24	0.61 Se
30	0.55 Se
36	0.50 Se
56	0.40 Se
100	0.30 Se

A2.2.4 The value of *E* shall be selected based upon general experience of those familiar with the specific testing in consultation with the permanent D11 IRM Steering Committee. The samples shall be taken during production at intervals so as to sample the entire lot in a uniform process (equally distributed sample selection). Two options are offered: sample during the package filling process or sample from finished packages.

A2.2.5 *Size of Samples*—The physical volume or mass taken for a sample will depend on the material being sampled. For rubbers, the sample size shall be 3 to 4 kg. For rubber chemicals, liquids, carbon black, or fillers, an appropriate amount shall be taken to allow for several test portions per sample. This is important for retest operations.

A2.3 *Sampling Plan 2*—If, for certain justified reasons, a sampling plan as described in Sampling Plan 1 cannot be carried out, Sampling Plan 2 shall be conducted. This second sampling plan requires fewer samples and it may be carried out when extensive blending of the IRM has been conducted during the manufacturing or production process or subsequent to its production but prior to the sampling operation. This blending ensures that a high degree of homogeneity exists and decreases the reliance on an extended sampling operation.

A2.3.1 Select at least twelve samples from the lot on a basis that ensures that all portions of the lot from beginning to end of the production process are represented equally among the twelve or more samples. If the samples are selected from a large container, ensure that all zones or locations in the container are equally represented in the samples.

A2.3.2 The sample size for Plan 2 shall be as specified in A2.2.5, with a sufficient amount in each sample for several tests for whatever property is being measured.

A3. TEST PLAN AND ANALYSIS FOR HOMOGENEITY OF AN IRM

A3.1 *Introduction:*

A3.1.1 This annex gives the instructions for evaluating the homogeneity for any candidate IRM. The testing (*1*) establishes the degree of uniformity for the measured properties that define the quality of the lot, and (*2*) generates data to establish the test lot limits (TL limits).

A3.1.2 The homogeneity is established on the basis of a 95 % confidence interval, that is, portions of the lot are rejected as outlier portions based on their measured values exceeding ±2 standard deviation limits. Once a homogeneous lot has been accepted on this basis, the TL limits are defined as the ±3 standard deviation limits of the measured individual test results in the IRM production laboratory. Practice E826 is used as a reference for this annex.

A3.2 *Brief Theory of Homogeneity Analysis:*

A3.2.1 The homogeneity analysis concepts are developed by considering a perfectly homogenous lot of material. If n samples of this lot are taken and k-replicate tests are conducted on each sample, the homogeneity of the lot is demonstrated if the pooled standard deviation or variance among the n sets of k-replicates is statistically equal to the standard deviation or variance (adjusted for the value of k), among the n samples drawn from the lot, that is, the observed variation among the lot samples is not significantly greater than that contributed by the testing itself!

A3.2.2 If it is not known whether the lot is truly homogeneous, and the sample variation is shown to be statistically greater than the pooled k-replicate variation, it is then presumed that the lot is not homogeneous. The next problem is to decide which part or parts of the lot depart from the bulk (or some remainder) of the lot.

A3.2.3 Decisions about homogeneity are made on the basis of ranges. Tests are conducted (k-replicates) on each of the n samples; averages (X), of the k values are calculated and the range among the averages is evaluated. This observed range among all lot samples, w(obs), which is equal to X(max) − X(min), is compared to a critical range, w(crit), that is evaluated on the basis of the expected range, if the only source of variation among the n samples is that due to the k-replicates. If w(obs) is greater than w(crit), then some part or parts of the lot deviate substantially from the bulk or remainder and are the cause of the non-homogeneity.

A3.2.4 If the n sample averages are sorted from low (minimum) to high (maximum), the deviating part or parts are easily identified and can be discarded from the lot. This part by part sequence is repeated until w(obs) is less than w(crit) and a homogeneous lot is obtained.

A3.2.5 Once a homogeneous lot is obtained, the TL limits are calculated from the standard deviation of the remaining packages, items, or portions of the lot (or the entire lot if no portions are rejected).

A3.3 *Conducting the Homogeneity Analysis:*

A3.3.1 *Correcting for Test Machine Drift:*

A3.3.1.1 The first step is to conduct the testing on the n samples drawn from the lot. Each sample is to be tested k number of times. If a large number of samples has been drawn or a large number taken during a production process and the time span to conduct all $n \times k$ tests is more than one day, an evaluation for measurement system drift shall be made. This evaluation is conducted by testing a control material according to a specific plan, depending on the number of samples taken from the lot.

A3.3.1.2 The control material shall be of the same type and have approximately the same property level as the IRM and be as uniform as possible. If any doubts exist on uniformity, blending shall be done if this is possible. The testing is conducted according to a specified plan with terms defined as follows:

NOTE A3.1—The control material may be a part or small fraction of the lot of the IRM that is sampled.

A3.3.1.3 *Test Number*—The samples are to be numbered in consecutive order as they are drawn from the lot during the sampling process. A random testing order for these consecutive samples is recommended. If this is not possible, a notation should be made in the report.

A3.3.1.4 *Test Replication*—The test sequence involving all n samples is conducted k times (k replicates) in the random sample order as indicated under "Test Number." The value for k is ideally 4. If this imposes a burden on the testing program or if the number of samples is large, a value of 3 or, at the very least, 2, may be selected for k.

A3.3.1.5 *Control Testing Frequency*—The control material, C, shall be tested at a frequency that is dictated by the size of the lot (number of samples drawn). Table A3.1 gives the testing frequency. The frequency is defined as the number of IRM samples, NI_i, tested between successive control samples. The control material is always tested first.

A3.3.2 *Analysis for Drift:*

A3.3.2.1 Tabulate the control test values in order of testing; C_1, C_2, C_3, ... C_n, and calculate the differences Δ between immediate successive values of C as follows:

$$\Delta_1 = C_1 - C_2 \qquad (A3.1)$$

$$\Delta_2 = C_2 - C_3 \qquad (A3.2)$$

$$\Delta_3 = C_3 - C_4, \text{ etc.} \qquad (A3.3)$$

TABLE A3.1 Control Material Testing Frequency

Number of samples, *m*	Testing Order and Frequency Sequence for Control Material[A]
5 or less	Frequency = 2, that is, use [C, NI_1, NI_2, C, NI_3, NI_4, C, NI_5]
5 to 10	Frequency = 3, that is, use [C, NI_1, NI_2, NI_3, C, NI_4, NI_5, NI_6, C, etc.]
11 to 20	Frequency = 4
21 to 50	Frequency = 5

[A] C = control material; NI_i = test frequency = number of IRM samples tested between successive control samples.

A3.3.2.2 Calculate the variance of the C values, called S_1^2, based on successive differences as follows:

$$S_1^2 = \sum \Delta^2 / 2(m - 1) \qquad (A3.4)$$

where:

Δ = difference in immediate successive values of C, and
m = total number of control material samples tested.

A3.3.2.3 Calculate a second estimate of the variance among the C values S_2^2 as follows:

$$S_2^2 = \sum d^2 / (m - 1) \qquad (A3.5)$$

where:

$d = (C_i - \bar{C})$ = difference of each measured C_i value from the average of all C values, designated $\bar{C}$.

A3.3.2.4 A decision on the occurrence of drift is made as follows. Calculate the ratio of S_1^2 to S_2^2. If the ratio S_1^2/S_2^2 obtained from the C value measurements is larger than the critical ratio values listed in Table A3.2 for the specified value of m, which is the total number of C values measured, then a statement can be made that there is no drift. The confidence level for this statement is 95 %.

A3.3.2.5 If the ratio S_1^2/S_2^2 is less than the critical tabulated value, then a statement can be made that drift has occurred. The confidence level for this statement is also 95 %.

A3.3.3 *Correction for Drift:*

A3.3.3.1 If analysis shows that drift is absent, the measured values of the IRM samples should be used. If this is the case, proceed to the next section. If drift is shown to be present, make a correction of the IRM sample values for the drift.

A3.3.3.2 A correction for drift is made on the basis of a linear drift behavior. The "drift" is corrected by the use of drift correction factors, F_i.

A3.3.3.3 Arrange the data values obtained for the control material in chronological order (C_1, C_2, ... C_n). Compute the drift factors, F_i, as follows:

$$F_1 = (C_1 + C_2)/2C_1 \qquad (A3.6)$$
$$F_2 = (C_2 + C_3)/2C_1 \qquad (A3.7)$$

TABLE A3.2 Critical Values of (S_1^2/S_2^2) Ratio
at 95 % Confidence Level[A]

Number of Control Samples Measured, m	(S_1^2/S_2^2) Ratio
4	0.39
5	0.42
6	0.44
8	0.49
10	0.53
12	0.56
15	0.60
20	0.65
25	0.68
30	0.72
35	0.74
40	0.76
45	0.78
50	0.80

[A] See NBS Special Publication N-63-2, U.S. Government Printing Office, Washington, DC 19630; see also C. A. Bennett, *Industrial and Engineering Chemistry*, 43, 2063 (1951). Ratio values for m = 30 to 50 calculated from extrapolation equation; Ratio (S_1^2/S_2^2) = 0.146 + 0.386 × $\log_{10}$ (m).

$$F_3 = (C_3 + C_4)/2C_1 \qquad (A3.8)$$
$$F_4 = (C_4 + C_5)/2C_1 \qquad (A3.9)$$
$$F_5 = (C_5 + C_6)/2C_1 \qquad (A3.10)$$
$$F_6 = (C_6 + C_7)/2C_1 \qquad (A3.11)$$
$$F_7 = (C_7 + C_8)/2C_1 \qquad (A3.12)$$
$$F_8 = (C_8 + C_9)/2C_1 \qquad (A3.13)$$

A3.3.3.4 Divide the measured IRM values by the appropriate drift factor to obtain the corrected value for the IRM samples. The appropriate drift factor for the IRM values is that factor, calculated from the control or C values that brackets the measured materials (IRM samples) within the time or measurement span for the two C values. Apply the factor F_1 to the IRM samples between $\bar{C}_1$ and C_2; apply F_2 to the IRM samples between C_2 and C_3, etc. Therefore, for a Frequency 3 Program:

$$\text{(Corrected) IRM Sample 1, 2, or 3 value} \qquad (A3.14)$$

$$= \frac{\text{Measured IRM Sample 1, 2, or 3 Value}}{F_1}$$

$$\text{(Corrected) IRM Sample 4, 5, or 6 value} \qquad (A3.15)$$

$$= \frac{\text{Measured IRM Sample 4, 5, 6 Value}}{F_2}, \text{ etc.}$$

A3.3.3.5 Tabulate the drift corrected IRM values for subsequent analysis and review.

A3.3.4 *IRM Lot Characteristics:*

A3.3.4.1 An IRM may be one of two different types: (*1*) a particulate or liquid material that can be blended to improve uniformity, or (*2*) a material produced in a form that cannot be blended. Thus there are two types of IRM:

(*1*) a Type *B* IRM, a material that may be blended, and
(*2*) a Type *NB* IRM, a material that cannot be blended.

A3.3.4.2 The procedure to demonstrate what degree homogeneity exists in the lot, differs depending on what type of IRM is being evaluated. The important issue is the "variation metric" or standard deviation that is used to calculate an expected range for the measured average values of the lot samples. This standard deviation depends on the history of the lot at the time of sampling and on the type of IRM, Type *B* or Type *NB*.

A3.3.5 *Basic Concepts for Evaluating Homogeneity:*

A3.3.5.1 The homogeneity analysis is based on the distribution of the q-statistic, defined by Eq A3.16.

$$q = w/[Sr/(k)^{0.5}] \qquad (A3.16)$$

where:

w = range (that is, maximum-minimum) of n average values,
Sr = residual standard deviation (pooled) of individual test values, and
k = number of individual values used for each average.

A3.3.5.2 Table A3.3 gives the 95 % confidence level (or $p = 0.05$) critical q-values for this distribution, with DF = the degrees of freedom for the pooled Sr and n = the number of sample averages in the calculated range, w. In applying the basic statistical expression of Eq A3.16, the equation is

TABLE A3.3 95 % Significance Level Critical Values for *q* for Combinations of: Number of Lot Samples, *n*, and Degrees of Freedom, *DF*, for Residual Standard Deviation, *Sr*

DF ↓ *n*→	2	3	4	5	6	7	8	9	10
1	17.97	26.98	32.82	37.08	40.41	43.12	45.40	47.36	49.07
2	6.08	8.33	9.80	10.88	11.74	12.44	13.03	13.54	13.99
3	4.50	5.91	6.82	7.50	8.04	8.48	8.85	9.18	9.46
4	3.93	5.04	5.76	6.29	6.71	7.05	7.35	7.60	7.83
5	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99
6	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49
7	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92
9	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74
10	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60
11	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39
13	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20
16	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15
17	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11
18	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07
19	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04
20	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92
30	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.72	4.82
40	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.73
60	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65
120	2.80	3.36	3.68	3.92	4.10	4.24	4.36	4.47	4.56
∞	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47

DF ↓ *n*→	11	12	13	14	15	16	17	18	19	20
1	50.60	52.00	53.20	54.30	55.40	56.30	57.20	58.00	58.80	59.60
2	14.40	14.70	15.10	15.40	15.70	15.90	16.10	16.40	16.60	16.80
3	9.72	9.95	10.20	10.40	10.50	10.70	10.80	11.00	11.10	11.20
4	8.03	8.21	8.37	8.52	8.66	8.79	8.91	9.03	9.13	9.23
5	7.17	7.32	7.47	7.60	7.72	7.83	7.93	8.03	8.12	8.21
6	6.65	6.79	6.92	7.03	7.14	7.24	7.34	7.43	7.51	7.59
7	6.30	6.43	6.55	6.66	6.76	6.85	6.94	7.02	7.09	7.17
8	6.05	6.18	6.29	6.39	6.48	6.57	6.65	6.73	6.80	6.87
9	5.87	5.98	6.09	6.19	6.28	6.36	6.44	6.51	6.58	6.64
10	5.72	5.83	5.93	6.03	6.11	6.20	6.27	6.34	6.40	6.47
11	5.61	5.71	5.81	5.90	5.99	6.06	6.14	6.20	6.26	6.33
12	5.51	5.62	5.71	5.80	5.88	5.95	6.03	6.09	6.15	6.21
13	5.43	4.53	5.63	5.71	5.79	5.86	5.93	6.00	6.05	6.11
14	5.36	5.46	5.55	5.64	5.72	5.79	5.85	5.92	5.97	6.03
15	5.31	5.40	5.49	5.58	5.65	5.72	5.79	5.85	5.90	5.96
16	5.26	5.35	5.44	5.52	5.59	5.66	5.72	5.79	5.84	5.90
17	5.21	5.31	5.39	5.47	5.55	5.61	5.68	5.74	5.79	5.84
18	5.17	5.27	5.35	5.43	5.50	5.57	5.63	5.69	5.74	5.79
19	5.14	5.23	5.32	5.39	5.46	5.53	5.59	5.65	5.70	5.75
20	5.11	5.20	5.28	5.36	5.43	5.49	5.55	5.61	5.66	5.71
24	5.01	5.10	5.18	5.25	5.32	5.38	5.44	5.50	5.54	5.59
30	4.92	5.00	5.08	5.15	5.21	5.27	5.33	5.38	5.43	5.48
40	4.82	4.91	4.98	5.05	5.11	5.16	5.22	5.27	5.31	5.36
60	4.73	4.81	4.88	4.94	5.00	5.06	5.11	5.16	5.20	5.24
120	4.64	4.72	4.78	4.84	4.90	4.95	5.00	5.05	5.09	5.13
∞	4.55	4.62	4.68	4.74	4.80	4.85	4.89	4.93	4.97	5.01

rearranged and a critical range, w(crit), is calculated according to Eq A3.17 with the critical q-value obtained from Table A3.3:

$$w(\text{crit}) = q \times Sr/(k)^{0.5} \qquad \text{(A3.17)}$$

With the knowledge of Sr (and n, DF and k also), a critical range may be calculated. If this critical range is compared to the observed range for the entire lot, defined as w(obs), a decision may be made about the homogeneity of the lot. Thus if w(obs) $>$ w(crit), then some parts or portions of the lot deviate from the underlying uniformity defined by Sr and must be eliminated from the lot. If, w(obs) $\leq$ w(crit), then the lot variation among the n samples is consistent with the uniformity defined by Sr and the lot is homogeneous.

A3.3.5.3 The residual standard deviation, Sr, represents a different system-of-causes for residual variation depending on the type of IRM. In general, the underlying system-of-causes variation expressed by Sr is defined by Eq A3.18 (given in terms of additive variances) as follows:

$$Sr^2 = Srt^2 + Srp^2 \qquad \text{(A3.18)}$$

where:

Srt^2 = residual variance due to test measurement (error) variation, and

Srp^2 = residual variance due to production process variation.

All IRMs at some early point in their production contain both components of variation. If the IRM can be blended (and is blended), the production process variation can be eliminated if the blending is sufficient. If blending is not possible, then the value for the residual standard deviation, Sr, used to calculate the expected range must include the production process component of variation.

A3.3.5.4 If the number of samples, n, in the lot exceeds 20, a problem is encountered with Table A3.3, that is, there are no listed values for the critical q values beyond this value. The solution to this problem is to split the lot into as many groups of 20 samples (or less) as needed. Calculate a w(crit) for each of these groups and compare this to the w(obs) of each of the groups. If all groups are homogeneous on an individual basis, the lot (all groups collectively) is homogeneous. If any group of 20 is non-homogeneous, then those portions that need to be eliminated shall be eliminated. All groups or fractions of a group that are homogeneous shall be combined into one homogeneous lot.

A3.3.6 *Uniformity Levels for IRM:*

A3.3.6.1 Based on the IRM lot homogeneity evaluation concepts developed in A3.3.5, it is appropriate to define two uniformity levels for IRMs. These uniformity levels are based on the inherent or residual variation used to evaluate the homogeneity.

(1) Uniformity Level-1, (UL-1), an IRM that has a residual standard deviation, Sr, which contains only the Srt component of lot variation.

(2) Uniformity Level-2, (UL-2), an IRM that has a residual standard deviation, Sr, that has both Srt and Srp components of lot variation.

A3.3.6.2 It is possible to prepare a Type NB, IRM lot that is a UL-1 material. This normally requires that a substantial portion of the lot (approximately 50 % or more) be eliminated. The rejection of this substantial portion reduces w(obs). By selecting this more uniform fraction of the lot, the value of w(obs) can be made to be less than w(crit), which is calculated based on the value of Srt alone. This approach to IRM technology may be of special importance if it is desired to prepare a super-homogeneous Type NB IRM for certain critical IRM applications.

A3.3.7 *Evaluating Homogeneity for Type-B (UL-1) IRM:*

A3.3.7.1 To evaluate the homogeneity level for any Type B (UL-1) IRM, the residual standard deviation, Sr, is obtained from a typical two-way analysis of variance (ANOVA). The two factors or categories in this analysis are samples and replicates. This type of analysis may be conducted most easily by employing computer statistical software that contains a two-way ANOVA option. The tabular organization of the basic test data depends on the particular software program employed. If such a program is available, organize the data as required and conduct the analysis. The analysis will list a residual variance;

the square root of this is the residual standard deviation Sr. Use this Sr for the range calculation as outlined in A3.3.7.4.

A3.3.7.2 If a statistical software program that has a two-way ANOVA option is not available, a two-way ANOVA may be conducted with typical spreadsheet programs. For this analysis organize the data as indicated in Table A3.4, where CTi is any column total, RTj is any row total, $\bar{X}i$ is any row (sample) average, and GT is the grand total of all measured test values. Each cell of the table contains an individual test value designated by the symbol Yij.

A3.3.7.3 Calculate the specified parameters of the table (CT, RT, $\bar{X}$, and GT). Appendix X2 gives the calculation algorithms for the two-way ANOVA. Perform the analysis and determine the residual standard deviation Sr.

A3.3.7.4 Calculate the 95 % confidence level critical range, w(crit), according to Eq A3.17 in A3.3.5.2, using Sr, total DF, n, and k.

A3.3.7.5 Using a spreadsheet program, sort the sample averages from low to high values. In this sort operation maintain the sample identification number in the database to be sorted so that the sample number accompanies (is linked to) the sample value in the sorting operation. Each sample number shall represent a particular identifiable portion of the lot that may be separated from the bulk of the lot if needed. From the sorted database, evaluate the observed range, w(obs).

A3.3.7.6 Compare the value of w(crit) to w(obs) as follows: If w(crit) |Ls w(obs), then the lot of the IRM has a high level of homogeneity; any variation within the lot is equal to or less than the test variation. The value of Sr may be used to calculate the test lot limits for individual test values, according to Eq A3.19.

$$\text{test lot limits} = \pm 3 \ (Sr) \qquad (A3.19)$$

If w(obs) > w(crit), then some portion or portions of the lot depart sufficiently from the bulk of the lot and thereby introduce a level of non-homogeneity into the lot.

A3.3.7.7 If w(obs) > w(crit), the next step is to identify the portion or portions of the lot that contribute to the observed state of non-homogeneity. This is accomplished by reviewing the sorted sample average values. Although this review may be conducted on the tabulated data, it is usually instructive to generate a *sample average profile*, a plot of the sample average (on the y-axis) versus the sample number (on the x-axis) with the sample numbers arranged in ascending order of sample average. This type of plot easily identifies outlier sample averages at either end of the distribution.

A3.3.7.8 The lot may be trimmed or reduced in size to reject the needed portion or portions. If there are a number of portions that contribute to the non-homogeneity, three options

TABLE A3.4 Tabulation of (Type B) IRM Test Data for Homogeneity Analysis

Sample No.	Replicate(s)			Row Totals	Average
	1	2	k		
1	...	...	...	RT1	$\bar{X}1$
2	...	y_{22}	...	RT2	$\bar{X}2$
3	...	...	...	RT3	$\bar{X}3$
n	...	...	...	RT n	$\bar{X}n$
Column Totals	CT1	CT2	CTk	(GT)	

are possible to trim the lot: (*1*) reject samples (and their corresponding portions) at the low end of the range of sample average values, (*2*) reject samples at the high end, or (*3*) reject samples at both ends. The choice depends on the degree of departure of the sample averages of the offending portions. The goal is to make the test lot limits as small as possible.

A3.3.7.9 Once *w*(obs) of the new trimmed lot has been reduced to be equal to or less than *w*(crit), a value for the test lot limits may be calculated by using Eq A3.19. The value to be used for the residual standard deviation (*Sr*), is obtained from a (new) two-way analysis of variance of the data of the trimmed lot.

A3.3.7.10 Calculate the grand average, $\bar{X}n$, of all sample averages of the accepted lot, that is, the entire homogeneous lot if no portion rejection was necessary or the trimmed lot if certain portions were rejected.

A3.3.7.11 The accepted homogeneous lot is characterized by two parameters: (*1*) the grand average of the lot, or test lot average, $\bar{X}n$, and (*2*) the test lot limits, that is, ±3 (*Sr*).

A3.3.8 *Evaluating Homogeneity for Type NB (UL-2) IRM:*

A3.3.8.1 A Type *NB* IRM has a residual standard deviation, *Sr*, that has two components of variation, *Srt* and *Srp*. Type *NB* materials are ordinarily generated by a production process and they are materials that cannot be blended (or have not been blended). The test data for a Type *NB* material is in general identical in format to data for a Type *B* material, that is, both consist of a series of *n* samples, each with *k* replicates. However with a Type *NB* material the value of *Sr* obtained as a residual in a two-way ANOVA cannot be used to evaluate homogeneity, since *Sr* does not include the *Srp* variation.

A3.3.8.2 The required *Sr* for a Type *NB* material must be evaluated from a secondary sampling operation of the IRM production process when this process is in a state of statistical control. The secondary sampling operation shall consist of taking 20 to 30 samples from the process during the period of documented statistical control. This period of statistical control sampling should be concentrated over some reasonable fraction (0.1 to 0.15) of the total run time for the production of the IRM. The appropriate test properties of these secondary samples are measured and the (combined) standard deviation, *Sr*, is evaluated from this sampling data and used to calculate *w*(crit) as specified by Eq A3.17 in A3.3.5.2.

A3.3.8.3 To evaluate the homogeneity for a Type *NB* material as a candidate for an IRM, follow the instructions given in A3.3.7.4 – A3.3.7.11 as given. For all calculations, the value of *Sr* shall be that obtained from the statistical control sampling operation as specified in A3.3.8.2.

A3.3.9 *Evaluating Homogeneity for Type-NB (UL-1) IRM:*

A3.3.9.1 If a lot of Type *NB* material is desired that has a Uniformity Level-1 magnitude for the residual standard deviation and the critical range *w*(crit), this candidate IRM may be prepared according to the instructions outlined in A3.3.7. The residual standard deviation, *Sr*, is evaluated in the same manner as for a Type *B* (UL-1) IRM, that is, it contains only the *Srt* component. Follow the instructions as specified in A3.3.7.4 – A3.3.7.11. To attain the desired level of homogeneity to become a UL-1 material, the rejection of a substantial fraction (one-half or more) of the lot may be required. Reject portions of the lot with the goal of minimizing the *Sr* value.

A3.3.9.2 When the candidate lot of material has been trimmed or reduced to the size that *w*(obs) is equal to or less than *w*(crit), with *w*(crit) calculated on the basis of *Srt*, the candidate lot of material may be accepted as a Type *NB*, UL-1 IRM.

A3.3.10 *Evaluating Homogeneity via Comparison with a Similar Reference Material:*

A3.3.10.1 Special circumstances may arise in the preparation of a candidate IRM that prevent the evaluation of a residual standard deviation to be used to calculate the critical range, *w*(crit). In such a situation it may be possible to make a decision to determine if the observed range, *w*(obs), of the candidate IRM is small enough for the material to serve as an IRM with an acceptable level of homogeneity, by comparing the range, *w*(obs), to a similar range of a similar material previously prepared by some accredited organization known to produce accepted homogeneous reference materials.

A3.3.10.2 The range or the standard deviation (among all tested portions or packages) of the known reference material may be compared to the range or the standard deviation (among all portions or packages) of the candidate IRM. If the range or standard deviation of the candidate IRM is equal to or less than the accepted reference material, the candidate IRM may then be declared as acceptable for homogeneity.

A3.3.11 *Establishing Homogeneity by Alternative Documentation:*

A3.3.11.1 For IRM candidate materials such as oils or liquids that may be and have been thoroughly and extensively blended, the homogeneity testing may be waived. The producer shall furnish documentation on the blending operation and this shall be included in the report on the IRM.

A3.3.11.2 For global AR value evaluation for this alternative homogeneity process, follow the instructions of Annex A4.

A3.3.11.3 For local AR value evaluation accepted on the basis of this alternative homogeneity process, a suitable testing program shall be conducted instead of homogeneity testing. Sampling Plan 2 of Annex A2 shall be followed; this calls for twelve samples to be drawn from the lot. To establish the local AR value, test each sample two times (two replicates), on a Day 1–Day 2 schedule as in accordance with A4.4.2. The AR value is the grand average of all tests. The TL limits are evaluated from the standard deviation of the replicate test results, *Sr*. See A3.3.7.6.

A4. TEST PLAN AND ANALYSIS TO EVALUATE AN ACCEPTED REFERENCE VALUE

A4.1 *Introduction:*

A4.1.1 This annex gives the instructions to evaluate an accepted reference value (AR value) and where applicable the between-laboratory limits. There are two types of AR values: (*1*) a global AR value obtained from the results of an interlaboratory test program (ITP), which may include laboratories on a world-wide basis, and (*2*) a local AR value obtained from one laboratory (or location). A decision, as to which category of AR value shall be evaluated for a particular IRM, shall be made by the IRM Steering Committee or a task group acting in the same capacity.

A4.1.2 For a global AR value the number of the laboratories in the ITP is usually made as large as possible. This produces a more realistic value for the between-laboratory limits and a more robust (more stable average) AR value. Although not of direct interest for interlaboratory comparisons, the within-laboratory variation or repeatability (collective value for all laboratories) may be calculated and may be reported as part of the documentation for the IRM.

Part 1: Global AR Value Evaluation

A4.2 *Organizing the Inter-Laboratory Test Program:*

A4.2.1 The type(s) of test(s) to be performed in each laboratory are selected. The specific details or test conditions, or both, are clearly described. Normally the tests shall be conducted via an ASTM D11 standard test method.

A4.2.2 Each test method produces a test result, which is defined as the average or median of a specified number of determinations (individual measurements) of the property being evaluated. Each test result is defined as a replicate and the number of test replicates in each laboratory shall be at least two, with each replicate test conducted on a separate day. The two or more days for the replicate testing should ideally be one week apart.

A4.2.3 The test dates for ITP testing are selected and this information conveyed to all laboratories. A coordinator and analyst to receive all test results is selected.

A4.3 *Allocation of Test Portions to the Laboratories:*

A4.3.1 Portions or packages of the IRM are allocated to the participating laboratories depending on the nature of the IRM, either Type *B* or Type *NB*. See A3.3.4.

A4.3.1.1 *Type B Procedure*—Take aliquot parts of portions of each of the *n* samples as selected for the homogeneity testing. Blend these aliquot parts or portions (again) to ensure that a sufficient quantity is blended for all participating laboratories. Prepare test packages of this re-blended material to be sent to each of the laboratories of the program.

A4.3.1.2 *Type NB Procedure*—A single package is selected from the lot; this shall be one that is as close as possible to the lot average as measured in the IRM production organization or laboratory. To ensure that the average value for this package is well documented, six additional measurements shall be made

on this package by the production laboratory at the same time as it conducts the tests for homogeneity. Portions of this package are distributed to all laboratories. A procedure will be described in A4.4.3 for verifying the closeness of the property value for this selected package to the production laboratory property lot average and a correction procedure will be outlined for any unintended deviation.

A4.4 *Analysis of Test Data from Inter-Laboratory Program:*

A4.4.1 The data generated by the ITP may be analyzed by means of a typical computer spreadsheet program or by the use of Practices D4483 or E691. The new revised (2004 version) Practice D4483 has a special two-step procedure that may be used to identify outlier values. These procedures may be used in place of the Tietjen-Moore test for outliers as described below in A4.4.3.2.

A4.4.2 An accepted statistical procedure shall be used to reject outliers. In the spreadsheet analysis, the Tietjen-Moore outlier rejection technique may be used. In the Practice E691 approach, the *h*-value analysis is used for outlier rejection. After an outlier analysis is completed, the AR value is determined.

A4.4.3 *Spreadsheet—Outlier Analysis:*

A4.4.3.1 *Tabulating the Data*—Enter the ITP test result data in a computer spreadsheet program in the format of Table A4.1. If more than one type of test is conducted as part of the ITP (for example, measurement at more than one temperature), arrange each type of test in this format. The column symbols are defined as follows; $R1$ = replicate 1, $R2$ = replicate 2, etc., R_{avg} = average of replicates. Each cell in the table has a test result entry defined as xij. The average and standard deviation of each column are indicated at the bottom of the table. The overall average (for all laboratories, all replicates) in the table is indicated by $\bar{X}_N$.

NOTE A4.1—Some spreadsheet computer programs use *n*, rather than *n* − 1, as the divisor for the sum of squares in the calculation of standard deviation. The divisor should be *n* − 1. If *n* is used, multiply the calculated standard deviation by $[n/(n − 1)]^{1/2}$.

A4.4.3.2 *Rejecting Outlier Values: Tietjen-Moore Test*—If one or more outliers are present in a set of data, the Tietjen-Moore test can be used. The test deals with outliers at either end of the distribution simultaneously when the mean and

TABLE A4.1 Recommended Data Format for Accepted Reference Value Analysis

Laboratory	Test (Type) 1			
	$R1$	$R2$	Ri	R_{avg}
1	$x11$	$x12$	$x1j$	$x1$
2	$x21$	$x22$	$x2j$	$x2$
3	. . .	. . .	. . .	. . .
.	. . .	. . .	. . .	. . .
N	$xi1$	$xl2$	xij	$\bar{x}$
Average	$\bar{X}_{R1}$	$\bar{X}_{R2}$	$\bar{X}_{Ri}$	$\bar{X}_N$
Standard Deviation	sd_{R1}	sd_{R2}	sd_{Ri}	sd_N
				($\bar{X}_N$ = average of all labs)

variance are estimated from the data. The test statistic $E(k)$ is defined by Eq A4.1. Critical values of $E(k)$(crit), at the 95 % confidence level ($p = 0.05$) are given in Table A4.2 for sample size $n = 3$ to 30.

$$E(k) = \sum_{i}^{(n-k)} \left(\bar{X}i - \bar{X}k \right)^2 / \sum_{i}^{n} \left(\bar{X}i - \bar{X} \right)^2 \qquad (A4.1)$$

where:

k = number of suspected values in the distribution sample of size n,
$\bar{X}i$ = test result for Laboratory i (average),
$\bar{X}$ = average of all values in the distribution sample, and
$\bar{X}k$ = average of the reduced sample with k suspected values deleted.

A4.4.3.3 Calculate (a column of) absolute deviations, d, for each of the values in the distribution sample: $d1 = \left| \bar{X}1 - \bar{X} \right|$, $d2 = \left| \bar{X}2 - \bar{X} \right|$, and so forth. Calculate (a column of) the absolute deviations squared and calculate the sum of the squared deviations.

A4.4.3.4 Sort the deviations from low to high. During this database sort operation maintain a linkage between the deviation and the laboratory (number) for that deviation. It is helpful for the next step to plot the deviations versus the laboratory number in ascending order of the value of the deviations. The plot enables the outlying laboratories (values) to be easily identified. The suspected values (laboratories) are eliminated and the sample average of the reduced sample is calculated. Repeat the calculations described in A4.4.3.3 for this reduced sample.

A4.4.3.5 Determine the value of $E(k)$ (calc) that is, the ratio of the sum of squared deviations of the reduced sample to the sum of squared deviations of the original (complete) sample. All suspected deleted values are declared to be outliers if $E(k)$ (calc) is less than $E(k)$ (crit) at 95 % confidence level as

TABLE A4.2 Critical Values of $E(k)$ at 95 % Confidence Level—Tietjen–Moore Test[A]

↓n[B] k[C] →	1	2	3	4	5
3	0.001				
4	0.025	0.001			
5	0.081	0.010			
6	0.146	0.034	0.004		
7	0.208	0.065	0.016		
8	0.265	0.099	0.034	0.010	
9	0.314	0.137	0.057	0.021	
10	0.356	0.172	0.083	0.037	0.014
11	0.386	0.204	0.107	0.055	0.026
12	0.424	0.234	0.133	0.073	0.039
13	0.455	0.262	0.156	0.092	0.053
14	0.484	0.293	0.179	0.112	0.068
15	0.509	0.317	0.206	0.134	0.084
16	0.526	0.340	0.227	0.153	0.102
17	0.544	0.362	0.248	0.170	0.116
18	0.562	0.382	0.267	0.187	0.132
19	0.581	0.398	0.287	0.203	0.146
20	0.597	0.416	0.302	0.221	0.163
25	0.652	0.493	0.381	0.298	0.236
30	0.698	0.549	0.443	0.364	0.298

[A] The above table is reproduced from Tietjen, G. L., and Moore, R. H., "Some Grubbs Type Statistics for Detection of Several Outliers," *Technometrics,* Vol 14, 1977, pp. 583–597.
[B] n = sample size.
[C] k = number of suspected values.

specified in Table A4.2. Repeat this procedure until all potentially outlying values have been deleted as indicated by $E(k)$ (calc) > $E(k)$ (crit).

A4.4.4 *Practice E691 Analysis for Outlier Rejection:*

A4.4.4.1 The procedure to reject outlier values in the ITP is based on the use of the h-value criterion as outlined in Practice E691. The value for h is calculated with Eq A4.2.

$$h = d/sd_N \qquad (A4.2)$$

where:

d = deviation between the R_{avg} or cell average, $\bar{x}_N$ for Laboratory i and the grand average of all laboratory or cell averages, and
sd_N = standard deviation of the cell averages $\bar{x}_N$, or R_{avg} values (among all laboratories).

The Practice E691 analysis output will contain a table of calculated h-values for each material in Practice E691 terminology or, for purposes of this analysis, for each type of test conducted in the ITP.

A4.4.4.2 Outliers shall be rejected based on the criterion that the calculated h-value for any laboratory (for any type test) is greater than the critical h-value at the 95 % confidence level (or $p = 0.05$). The critical h-value, h(crit) is calculated based on Eq A4.3.

$$h(\text{crit}) = (p - 1)t/[p(t^2 + p - 2)]^{1/2} \qquad (A4.3)$$

where:

p = number of laboratories in the ITP, and
t = tabulated 95 % confidence level student's t-value at DF = $(p - 2)$ (a two-tailed t-value).

A4.4.4.3 Using the spreadsheet format as defined in Table A4.1, it is informative to perform a database low-to-high sort operation on the $\bar{x}_N$ values. When the sort is performed, maintain a linkage between the $\bar{x}_N$ values and the laboratory number. Plot the sorted $\bar{x}_N$ values versus the laboratory number (in ascending order of $\bar{x}_N$ values).

A4.4.4.4 Calculate h(crit) by means of Eq A4.3. Using this h(crit), examine the calculated h-values in the output table of the Practice E691 analysis. Any test measurement values that have a significant h-value, that is, h(calc) > h(crit), shall be rejected. Apply this criterion to all tests (types) as conducted in the ITP. Any rejected values should be at the extremes of the low-to-high sort plot as outlined above.

A4.4.5 *Calculating the AR Value*—The calculation of the AR value is dependent on the Type of IRM evaluated; Type B or Type NB. In either case, the AR value is based on recalculations made after outliers are rejected.

A4.4.5.1 *Type B IRM*—If any test values were rejected by any of the procedures as described above, recalculate the value of $\bar{X}_N$. The AR value is equal to the recalculated (or original average if there are no outliers) average, $\bar{X}_N$, given by Eq A4.4.

$$(\textit{Type B}) \qquad \text{AR value} = \bar{X}_N \qquad (A4.4)$$

A4.4.5.2 *Type NB IRM*—If any test values were rejected by either of the procedures as described above, recalculate the value of $\bar{X}_N$. The measured AR-value is equal to the recalculated (or original) $\bar{X}_N$, given by Eq A4.5.

($Type\ NB$)   measured AR value $= \bar{X}_N$ (A4.5)

The measured AR value (average of laboratories in ITP) is a provisional AR value that is subject to correction depending upon the closeness of the selected package property level (used for all ITP laboratory samples), to the production laboratory property lot average, see A4.3.1.2. Calculate a correction, dc, as follows according to Eq A4.6.

$$dc = [\text{Production Laboratory lot average} \qquad (A4.6)$$
$$- \text{ITP package} \quad (6\ m) \quad \text{average}]$$

where:

Production Laboratory lot average = average of all homogeneity samples in the production laboratory, after any outlier rejection, and

ITP package (6 m) average = average of the six extra measurements on the package used to supply all the laboratories in the ITP, measured in the production laboratory.

The corrected AR value for an IRM lot where dc is non-zero is given by Eq A4.7.

$$\text{corrected AR value} = \text{measured AR value} + dc \qquad (A4.7)$$

The corrected AR value shall be used for preparation of the documentation sheet for the IRM.

A4.4.6 *Calculating the Between-Laboratory Limits*—The calculation of the between-laboratory limits is performed after all outliers have been rejected. These limits may be calculated via the spreadsheet technique or the Practice E691 approach.

A4.4.6.1 *Spreadsheet Calculations*—This procedure may be followed for either type of IRM. If any outliers were rejected, recalculate the values for sd_{R1}, sd_{R2}, and so forth as given at the bottom of Table A4.1. Calculate a pooled standard deviation for individual test results, sd_{Ri}, where Ri indicates a pooling operation across i replicates as used in the program, that is, (i) = number of replicates, by means of Eq A4.8. Normally, (i) is two.

$$\text{Pooled } sd_{Ri} = [\{(sd_{R1})^2 + (sd_{R2})^2 + \dots (sd_{Ri})^2\}/(i)]^{1/2} \quad (A4.8)$$

The between-laboratory limits are given by Eq A4.9.

$$\text{Between} - \text{laboratory limits} = \pm 2 \ (\text{pooled } sd_R) \qquad (A4.9)$$

These limits apply to individual (that is, single) test results for any laboratory. These limits shall be used in the preparation of the documentation sheet.

A4.4.6.2 *Practice E691 Calculations*—Revise the original Practice E691 database as analyzed for the h-values by eliminating the outlier values. Conduct another analysis on this revised database by means of the Practice E691 computer program. From the Practice E691 worksheets, one for each material or in terms of this IRM program each type of test conducted for the IRM, list the following worksheet values: (*1*) average of cell averages ($\bar{X}_N$), and (*2*) reproducibility standard deviation, S_R.

A4.4.6.3 The AR value for the Practice E691 output is given by the use of Eq A4.4 for a Type *B* IRM. For a Type *NB* IRM the AR value is given by Eq A4.5-A4.7.

A4.4.6.4 The between-laboratory limits are given by Eq A4.10.

$$\text{Between} - \text{laboratory limits} = \pm 2 \ (S_R) \qquad (A4.10)$$

A4.4.7 *Within-Laboratory Variation*—The pooled within-laboratory or repeatability standard deviation, Sr, may be obtained by means of the spreadsheet technique or by way of Practice E691. It is frequently informative to compare the test result standard deviation in any particular laboratory to the pooled or overall standard deviation for all laboratories in any ITP.

A4.4.7.1 *Spreadsheet Technique for Sr*—For each test performed, calculate (a column of) variances, one for each row (laboratory) of data in Table A4.1 format, that is, Day 1, Day 2, Day (i), by means of the special spreadsheet variance function algorithm. Sum the (column of) variances and divide by the number of values summed to give the pooled variance. The square root of this pooled variance is the pooled standard deviation. Repeat this sequence for each type of test performed.

A4.4.7.2 The next step is to calculate for each laboratory the k-value of Practice E691. This statistical parameter indicates how the within-laboratory day-to-day variation of each laboratory compares to the overall within-laboratory variation. The k-value is defined by means of Eq A4.11.

$$k = Sr(i)/Sr(\text{pooled}) \qquad (A4.11)$$

where:

$Sr(i)$ = cell standard deviation of Laboratory i, and
$Sr(\text{pooled})$ = pooled cell standard deviation (over all laboratories).

A4.4.7.3 Calculate (a column of) standard deviations, one for each row of data in Table A4.1 format. Using the pooled standard deviation obtained in A4.4.6.1, calculate (a column of) the ratio of each individual laboratory (row or cell) standard deviation to the pooled standard deviation. Perform a database low-to-high sort operation on the calculated k-values or ratios, with the laboratory number linked to the k-values in the sort procedure. Plot the k-value versus the laboratory number in ascending order of k-value.

A4.4.7.4 Calculate a 95 % confidence level, critical k-value, k(crit), as defined by Eq A4.12.

$$k(\text{crit}) = \{p/[1 + (p - 1)/F]\}^{1/2} \qquad (A4.12)$$

where:

p = number of laboratories in the ITP, and
F = tabulated F-value at 95 % confidence level ($p = 0.05$), for numerator, $DF = (n - 1)$, with n = number of replicates (days) and for denominator, $DF = (p - 1)(n - 1)$.

Compare the sorted k(calc) values to the k(crit) value as calculated by means of Eq A4.12 and reject any cell standard deviations that exceed k(crit). After all cell standard deviation ratios that exceed k(crit) have been rejected, recalculate the row or cell variances as in A4.4.6.1 omitting the cell variances that have been rejected and obtain the pooled variance and

pooled standard deviation of this reduced database. Use this recalculated *Sr* for each test performed.

A4.4.7.5 *Practice E691 Calculation Technique for Sr*— From the Practice E691 worksheet output, review the values of *k*(calc) for each test performed in the ITP evaluation (each test will be labeled as a "material" in the Practice E691 output terminology). Use A4.4.6.4 to calculate *k*(crit). Delete any cells in the Practice E691 *k*-value worksheet, that exceed the *k*(crit) value. Revise the Practice E691 database table by removing those cells (laboratories) that have *k*(calc) > *k*(crit).

A4.4.7.6 Recalculate the precision via the Practice E691 computer program using the revised database and record the values of *Sr* for each test performed as obtained on the Practice E691 output worksheets.

Part 2: Local AR Value Evaluation

A4.5 The local AR value for an IRM is evaluated from the database generated in the homogeneity testing as conducted in Annex A3. A3.3.7.10 and A3.3.7.11 call for the calculation of the grand average of all sample averages (or values) of the accepted homogeneous lot. This is defined as $\bar{X}n$. Eq A4.13 defines the local AR value.

$$\text{Local AR value} = \bar{X}n \qquad (A4.13)$$

A5. STATISTICAL MODEL FOR IRM TESTING

A5.1 *Introduction:*

A5.1.1 This annex presents a generic or basic statistical model that applies to any testing or measurement system. This basic model is then used to develop individual models that address the various types of IRM testing for homogeneity and accepted reference value evaluation. The models give a logical background that permits a more complete understanding of the sources of variation and the way in which each source contributes to the overall variation. The model also gives the rationale for the analysis procedures in Annex A3 and Annex A4.

A5.2 *Basic Statistical Model:*

A5.2.1 For any established measurement system, each measurement, *y*, can be visualized as the sum of a constant and a second complex term as indicated by Eq A5.1.

$$y = \mu + \sum d(j) \qquad (A5.1)$$

where:

μ = the true value, (a constant), obtained when all deviations, $d(j)$, are zero, that is, the ideal outcome of a measurement, and

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

A5.2.2 The term μ may also be represented as the ideal reference value, which stands in contrast to the empirically determined accepted reference value as discussed elsewhere in this standard. A more useful format is obtained when Eq A5.1 is expressed as an expanded model, where $\sum d(j)$ is replaced by a series of terms appropriate to interlaboratory testing, as given by Eq A5.2.

$$y = \mu + b_i + B_m + B_L + B_g + e_b(l) \qquad (A5.2)$$
$$+ e_b(s) + e_w(l) + e_w(s) + e(g)$$

where:

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

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

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

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

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

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

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

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

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

A5.2.3 In a perfect measurement world all biases and random deviations of Eq A5.2 would be zero. In the real world of measurement, these terms take on certain values and the sum of their collective values acts as a perturbation of the true value, μ, for each measurement. Both the actual value and the variance of each of these terms are important when considering testing and application programs. Tests to determine the significance of these individual terms usually involve a statistical comparison of the variances attributed to the terms.

A5.2.4 The value of the (*B*) terms is dependent on the measurement system or the system-of-causes, for the generation of the biases. The (*B*) terms in the model may be either fixed or variable as well as plus or minus, depending on the

measurement system under consideration. For any system, the variable (*B*) terms are typically a non-random finite distribution and therefore the values for a particular bias term will not of necessity sum to zero over the population constituting the system. Bias terms that are fixed under one system-of-causes may be variable under another different system-of-causes and vice-versa.

A5.2.5 The inherent bias, B_i, is characteristic of the overall design of the machine or apparatus. This type of bias is frequently of importance in chemical tests for certain constituents whose theoretical content can be calculated, for example, percent chlorine in sodium chloride. A given test device may always be low or high due to unique design features.

A5.2.6 One or more generic or general bias terms, B_g, may be included in the model to allow for any (non-inherent) unique systematic deviation not attributable to test machines or laboratories.

A5.2.7 The bias terms B_m and B_L are more appropriate for physical testing. As an example, for a particular laboratory (with one test machine) both of these bias terms would be constant or fixed. For a number of test machines, all of the same design in a given laboratory, B_L would be fixed but B_m would be variable, each machine having a unique value. For a measurement system consisting of a number of typical laboratories, both B_m and B_L would be variable for the multilaboratory measurement system but of course both B_m and B_L would be constant for each laboratory.

A5.2.8 The (*e*) terms represent random deviations that take on plus or minus values and that have an expected mean of zero (over the long run) and a variance equal to var(*e*). The (random) value of each of the (*e*) terms influences the measured *y*-value on an individual measurement basis. However in the long run when *y*-values are averaged over a number of measurements, the influence of the (*e*) terms is greatly diminished or eliminated since the terms average out to zero and the *y*-value is perturbed by the (*B*) terms only. This "long run zero average" character stands in contrast to the behavior of the fixed (*B*) terms where increased measurement increases the knowledge (accuracy) of the actual (*B*) value.

A5.2.9 To make the model building as accurate as possible as in the case of the bias terms, one or more generic or general random deviation terms, *e(g),* may be included in the model to account for any potential source of special random deviations not attributable to the within- or between-laboratory categories.

A5.2.10 The variation evaluated as the normal Day 1 versus Day 2 (one week apart) within-laboratory precision or repeatability, is represented by $e_w(s)$. The variation evaluated as the normal Day 1 versus Day 2 (one week apart) between-laboratory precision or reproducibility, is represented by $e_b(s)$.

A5.3 *Using the Statistical Model for the Testing of IRMs:*

A5.3.1 In this section the basic statistical model is used to build special models that apply to the specific test or measurement systems of IRM evaluation. Since not all of the terms in the general model are required for every type of testing, some of the unused terms are combined with the true value, μ, to reduce and simplify the presentation of the model. The symbol for the true value, μ, may be modified as follows:

$$\mu(0) = \mu + B_i; \text{ this ignores the } B_i \text{ term by combining it with } \mu$$
(A5.3)

$$\mu(1) = \mu + B_i + B_L; \text{ this ignores the } B_L \text{ term by combining it with } \mu(0)$$
(A5.4)

$$\mu(2) = \mu + B_i + B_L + B_m; \text{ this ignores } B_m \text{ by combining it with } \mu(1)$$
(A5.5)

The B_i term can generally be ignored for physical testing. For testing conducted within a given laboratory where a number of the same instruments are employed, B_L can be ignored. For testing conducted within a laboratory on one test device, the term B_m can be ignored (in addition to the B_L term). Under each of these three conditions the modified true value acts as the general constant of the model.

A5.3.2 The use of a particular statistical model will be reviewed for a number of testing situations. For IRM evaluation, models for the homogeneity analysis as outlined in Annex A3 will be presented as well as the model for the accepted reference value analysis as outlined in the calculations of Annex A4.

A5.3.3 *Case 1: IRM (Type B), Homogeneity Analysis*—This testing or measurement system example assumes one machine and short-term tests. The following terms are equal to zero; $e_w(l)$, $e_b(l)$ and $e_b(s)$. This produces a three-term model given by Eq A5.6.

$$y = \mu(2) + E(g) + e_w(s)$$
(A5.6)

where:

Var *[e(g)]* = unanticipated variance in the lot after the blending, and
Var*[e_w(s)]* = test measurement variance.

The combined Var[*e(g)*] + Var[$e_w(s)$] is evaluated from the *n* sample measurements (*y*-values) across the lot. The Var[$e_w(s)$] is evaluated from the *k*-replicate tests on each sample. The analysis of Annex A3 evaluates the ratio of Var[*e(g)*] + Var[$e_w(s)$] to Var[$e_w(s)$] by way of a range or *w*(crit) test (which is equivalent to a rearranged ANOVA *F*-Test). If after the blending for a Type *B* IRM, the Var[*e(g)*] is significant, then some level of non-homogeneity exists in the lot. The TL limits are ±3 times the square root of the sum of the two variances; the sum is used whether Var[*e(g)*] is significant or not.

A5.3.4 *Case 2: IRM (Type NB) Homogeneity Analysis*—This testing or measurement system example assumes one machine and short-term tests. The following terms are equal to zero: $e_w(l)$, $e_b(l)$, and $e_b(s)$. This produces a four-term model (two separate *e(g)* terms are required) given by Eq A5.7.

$$y = \mu(2) + e(g)1 + e(g)2 + e_w(s)$$
(A5.7)

where:

Var *[e(g)1]* = variance in the lot,
Var*[e(g)2]* = variance in the production process, and
Var*[e_w(s)]* = test measurement variance.

The variance among the primary lot samples (*y*-values) contains all three actual or potential components of variance.

The sum of $\text{Var}[e(g)2] + \text{Var}[e_w(s)]$ is evaluated from the variation of the 20 to 30 secondary production samples taken during a period of statistical control. The analysis of Annex A3 again uses the critical range approach to assess the significance of the combined variance, $\text{Var}[e(g)1] + \text{Var}[e(g)2] + \text{Var}[e_w(s)]$ when compared to $\text{Var}[e(g)2] + \text{Var}[e_w(s)]$. If $\text{Var}[e(g)1]$ adds a significant component to the combined variance then some non-homogeneity exists. The TL limits are given by ±3 times the square root of the sum of all three variances.

A5.3.5 *Case 3: IRM (Type B) Homogeneity Analysis with Several Machines*—This test system is comprised of more than one machine. Such a situation might exist for a large number of samples taken from the lot and where it is desired to run the tests as quickly as possible. The following terms are equal to zero: $e_w(l)$, $e_b(l)$, and $e_b(s)$. This gives a four-term model expressed by Eq A5.8.

$$y = \mu(1) + B_m + e(g) + e_w(s) \qquad (A5.8)$$

where:

Var $[B_m]$	=	variance among the machines,
Var$[e(g)]$	=	unanticipated variance in the lot after blending, and
Var$[e_w(s)]$	=	test measurement variance.

A5.3.5.1 The overall analysis for homogeneity for this system must begin with a preliminary analysis to determine if there is a significant difference among the test machines. This can be conducted in one of two ways: (*1*) if control samples were tested to assess any potential drift, a simple ANOVA can be conducted on the control sample data, comparing the mean squares for between machines versus within machines, or (*2*) alternatively an ANOVA can be conducted on the lot sample data by a similar comparison of the sum of the three components $\text{Var}[B_m] + \text{Var}[e(g)] + \text{Var}[e_w(s)]$ to the sum of the two components $\text{Var}[e(g)] + \text{Var}[e_w(s)]$.

A5.3.5.2 If significant differences are found among the machines, corrections are required to place all machines on a common basis. This is most easily accomplished when control samples have been tested. Corrections of the sample data are established on the basis of average control sample values for each machine. The TL limits are given by ±3 times the square root of the sum $\text{Var}[e(g)] + \text{Var}[e_w(s)]$.

A5.3.6 *Case 4: IRM (Type B or NB) Accepted Reference Value*—In this system there are *N* participating laboratories and the testing is short term. The following terms are zero: $e_b(l)$, $e_w(l)$. The model is given by Eq A5.9.

$$y = \mu(0) + B_m + B_L + e(g) + e_b(s) + e_w(s) \qquad (A5.9)$$

where:

Var $[B_m]$	=	variance due to machines (one in each laboratory),
Var$[B_L]$	=	variance due to operating conditions in each of the laboratories,
Var$[e(g)]$	=	variance among the portions (samples) sent to each laboratory,
Var$[e_b(s)]$	=	variance (of random nature) among the laboratories, and
Var$[e_w(s)]$	=	test measurement variance.

A5.3.6.1 The between-laboratory variation (among the individual laboratory-measured *y*-values) is the sum total of all five variances. The major source of differences among the laboratories is the variance of the two biases, $B_m + B_L$, with the variance of the random term $e_b(s)$ contributing a lesser amount under ordinary testing conditions for experienced laboratories. The $\text{Var}[e(g)]$ is usually small, since steps are taken to make the material sent out as uniform as possible.

A5.3.6.2 The major purpose of accepted reference value testing is the calculation of an average value, among all the laboratories participating in the interlaboratory program, as described in Annex A4. The second purpose is the establishment of the between-laboratory limits; these limits are equal to ±3 times the square root of the sum of all five variances. As outlined in Annex A4, calculations for both accepted reference value and the limits are performed only after the elimination of any outliers in the interlaboratory data.

APPENDIXES

(Nonmandatory Information)

X1. AN EXAMPLE: EVALUATING AN IRM USING THE CALCULATION PROCEDURES OF Annex A2, Annex A3, AND Annex A4

X1.1 *Introduction*—This appendix gives all the calculations involved in evaluating a Type *NB* IRM. It illustrates how certain optional decisions may be made about some of the steps of the overall analysis. A number of figures (plots) are presented that assist in the understanding of the various operations and the outcome of the analysis.

NOTE X1.1—The AR-value analysis in this example was conducted prior to the adoption of ±2 standard deviation between-laboratory limits for the AR-value (in 2003); it uses ±3 standard deviation limits.

X1.2 *Evaluating the IRM: Background Information*—A new rubber, designated as XPR, was proposed as an IRM. It was produced in a typical production run of 200 bales during the month of September 1992 by the X. P. Menter Chemical Co. Since this is a material that cannot be blended, it is classified as a Type *NB* IRM. The quality of the lot of 200 bales was determined by measurement of the Mooney viscosity, ML1 + 4 at 100°C.

X1.2.1 The homogeneity of the lot was evaluated by the procedures outlined in Annex A3, with primary production lot sampling conducted according to Annex A2. Annex A3 specifies that a secondary sampling operation be conducted for a Type NB IRM to ensure that the residual standard deviation, Sr, contains both test measurement variation and production variation. This was conducted by way of a 20-sample program during a period of "in control" production of the rubber. From the (final) homogeneous lot, the test lot limits and the test lot average were evaluated.

X1.2.2 The accepted reference value and the between-laboratory limits were evaluated from an interlaboratory test program with 24 participating laboratories that typically use the rubber and the Mooney test procedure.

X1.3 *Recommended Sampling Plan*—Annex A2 recommends that one of two sampling plans be adopted. In this case, a variation on the Plan 1 was selected. Rather than calculate the number of samples, *n*, to draw from the lot on the basis of a known lot standard deviation, *S*, or estimated lot standard deviation, *Se*, the number *n* was selected on the basis of *E* (the maximum deviation of the lot average from the true value) being equal to one half of *Se* (or *S*). This relation may be expressed as given by Eq X1.1. See also Eq A2.1.

$$n = (3\ Se/0.50\ Se)^2 = 36 \qquad (X1.1)$$

For purposes of this evaluation the value of 36 was rounded to 40, and this is the number of primary samples drawn from the lot. This more convenient number (40) is equal to a sample for every fifth bale. The lot standard deviation *S* or *Se* is equivalent to *Sr* as defined in A3.3.5 in Annex A3.

X1.4 *Homogeneity Analysis: Step 1 Analysis for Drift*—Table X1.1 lists the Mooney viscosity (ML1 + 4 at 100°C) production run data as generated from the sampling plan outlined above. Since this volume of testing cannot be conducted in a brief one-day period, a bale of the production run was selected as a control and tests were conducted on it as listed in Table X1.1. The frequency for control testing as specified in Table A3.1 of Annex A3 is given as five production samples between each successive pair of control samples, and this is the frequency that was used. The bottom of Table X1.1 gives the average, variance, and standard deviation of the columns in the table.

X1.4.1 Table X1.2 contains the control viscosity data and the drift analysis. The presence of drift is demonstrated if the variance $(S1)^2$, evaluated from successive squared differences as specified in A3.3.2 in Annex A3, is significantly less than the normal variance $(S2)^2$. The table gives all the intermediate calculated values needed to calculate the final parameter, the ratio of $(S1)^2$ to $(S2)^2$, which is equal to 1.03. The critical ratio is 0.53, which is obtained from the tabulated value for *m* = 10 (the nearest value to the actual *m* = 9) in Table A3.2 in Annex A3. Since the calculated ratio is substantially larger than the critical ratio, there is no evidence of drift over the period of the 40 sample production measurement period. Both evaluations for variance give values that are within 3 % of each other.

TABLE X1.1 Mooney Viscosity Data: Production Bale and Control Values

Bale Sample	MV1[A]	MV2[B]	AVGMV[C]	CMMV1[D]	CMMV2[E]	Average CMMV
				49.2	49.3	49.25
1	50.5	50.8	50.63			
2	50.0	50.4	50.18			
3	50.0	50.5	50.23			
4	50.3	50.0	50.13			
5	50.3	50.6	50.43.	49.2	49.3	49.25
6	50.2	50.6	50.40			
7	50.2	50.3	50.25			
8	50.0	50.1	50.05			
9	50.0	50.1	50.05			
10	50.0	49.9	49.93	49.1	49.1	49.10
11	50.6	50.6	50.58			
12	50.3	50.5	50.38			
13	50.0	49.9	49.93			
14	50.0	50.4	50.18			
15	50.2	50.6	50.38	49.2	49.5	49.35
16	50.0	50.0	49.98			
17	50.0	50.1	50.03			
18	50.3	50.5	50.40			
19	50.1	50.3	50.20			
20	50.2	50.1	50.15	49.4	49.2	49.30
21	50.1	50.2	50.15			
22	50.0	49.9	49.95			
23	49.9	49.9	49.90			
24	49.8	49.9	49.85			
25	50.0	50.0	50.00	49.3	49.3	49.30
26	50.0	49.8	49.88			
27	49.8	49.8	49.78			
28	49.9	49.8	49.83			
29	49.9	49.8	49.83			
30	50.0	50.0	49.98	49.1	49.2	49.15
31	50.0	50.1	50.03			
32	50.5	50.4	50.43			
33	50.2	50.3	50.23			
34	50.4	50.3	50.33			
35	50.6	50.1	50.33	49.2	49	49.10
36	50.7	50.6	50.63			
37	50.7	50.6	50.63			
38	50.5	50.0	50.25			
39	51.1	51.1	51.10			
40	51.2	51.2	51.20	49.6	49.4	49.50
Average	50.21	50.23	50.22	49.26	49.26	49.26
Variance	0.1081	0.1250	0.1029	0.0253	0.0228	0.0165
Standard Deviation	0.329	0.354	0.321	0.159	0.151	0.129

[A] MV1 = Replicate 1 Mooney ML1 + 4.
[B] MV2 = Replicate 2 Mooney ML1 + 4.
[C] AVGMV = Average Both Replicates.
[D] CMMV1 = No. 1-ML1 + 4 Value on Control Machine.
[E] CMMV2 = No. 2-ML1 + 4 Value on Control Machine.

X1.5 *Homogeneity Analysis: Step 2 Evaluating the Range, w(obs)*—Table X1.3 gives the Mooney viscosity for the 20 secondary samples taken during a period of "in control" production. The parameter of importance here is the standard deviation for the 20 samples, which is equal to 0.256. This standard deviation is equal to *Sr*, the residual standard deviation as initially defined in A3.3.5 and discussed more fully in A3.3.5.3 of Annex A3. The standard deviation value, 0.256, contains both components of variation, that is, the measurement as well as the inherent production variation. This value of *Sr* is used in the calculations outlined in the next section.

X1.5.1 Table X1.4 gives the IRM lot (production) Mooney viscosity in two groups of 20 samples each. This separation into groups of 20 is required for any homogeneity analysis

TABLE X1.2 Analysis for Drift—Control Machine Mooney Viscosity

CMMV No.	CMMV1	CMMV2	Average CMMV	Delta Average CMMV	Delta SQ
1	49.2	49.3	49.25	...	...
2	49.2	49.3	49.25	0.00	0.0000
3	49.1	49.1	49.10	0.15	0.0225
4	49.2	49.5	49.35	−0.25	0.0625
5	49.4	49.2	49.30	0.05	0.0025
6	49.3	49.3	49.30	0.00	0.0000
7	49.1	49.2	49.15	0.15	0.0225
8	49.2	49.0	49.10	0.05	0.0025
9	49.6	49.4	49.50	−0.40	0.1600
Average	49.26	49.26	49.26		
SUM					0.2725
Variance	0.02528	0.02278	0.01653		
S1(SQ) = 0.2725 ∕ 2(9-1) =					0.0170
S2(SQ) = VAR Average CMMV =					0.0165
Ratio S1(SQ)/S2(SQ) =					1.03
Critical Ratio S1(SQ)/S2(SQ) =					0.53

NOTE 1—
CMMV = Control machine Mooney viscosity.
CMMV1 or 2 = Replicate measurement 1 or 2 of CMMV.
Average CMMV = Average of two replicates, CMMV.
Delta Average CMMV = Immediate successive diff $(C_i – C_i + 1)$, where C_i = Average CMMV.
Delta SQ = (Delta average CMMV) squared.
S1(SQ) = Variance estimate from immediate successive differences.
S2(SQ) = Normal variance estimate.

TABLE X1.3 Secondary Sample: 20 Individual Mooney Values Taken During In Control Production Run

1	50.4	11	50.6
2	50.3	12	50.2
3	50.7	13	50.3
4	50.4	14	49.9
5	50.4	15	50.3
6	50.6	16	50.4
7	50.8	17	50.3
8	50.9	18	50.5
9	50.8	19	50.0
10	50.6	20	50.4

For all 20:
Average	50.44
Variation	0.0657
Standard Deviation	0.256

TABLE X1.4 Production Mooney Viscosity AVG MV in Sorted Increasing Order

Bale Sample	MV1	MV2	AVG MV
Group 1[A]			
10	50.0	49.9	49.93
13	50.0	49.9	49.93
16	50.0	50.0	49.98
17	50.0	50.1	50.03
8	50.0	50.1	50.05
9	50.0	50.1	50.05
4	50.3	50.0	50.13
20	50.2	50.1	50.15
14	50.0	50.4	50.18
2	50.0	50.4	50.18
19	50.1	50.3	50.20
3	50.0	50.5	50.23
7	50.2	50.3	50.25
15	50.2	50.6	50.38
12	50.3	50.5	50.38
18	50.3	50.5	50.40
6	50.2	50.6	50.40
5	50.3	50.6	50.43
11	50.6	50.6	50.58
1	50.5	50.8	50.63
Average	50.15	50.29	50.22
Standard Deviation	0.189	0.270	0.204
Group 2[B]			
27	49.8	49.8	49.78
29	49.9	49.8	49.83
28	49.9	49.8	49.83
24	49.8	49.9	49.85
26	50.0	49.8	49.88
23	49.9	49.9	49.90
22	50.0	49.9	49.95
30	50.0	50.0	49.98
25	50.0	50.0	50.00
31	50.0	50.1	50.03
21	50.1	50.2	50.15
33	50.2	50.3	50.23
38	50.5	50.0	50.25
35	50.6	50.1	50.33
34	50.4	50.3	50.33
32	50.5	50.4	50.43
37	50.7	50.6	50.63
36	50.7	50.6	50.63
39	51.1	51.5	51.10
40	51.2	51.2	51.20
Average	50.27	50.16	50.21
Standard Deviation	0.423	0.418	0.412

[A] For first group of 20: W(obs) = (50.63 − 49.93) = 0.70
Sr = 0.256
q(crit) = 5.75
k = 2
$w(crit) = q \times Sr/\sqrt{k}$
w(crit) = 5.75 × 0.256 ∕ 1.41 = 1.04
w(obs) < w(crit)

[B] For second group of 20:
w(obs) = (51.2 − 49.78) = 1.42
w(crit) = 5.75 × 0.256 ∕ 1.41 = 1.04
w(obs) > w(crit)

where there are more than 20 samples taken from the IRM lot. For each group the average viscosity values have been sorted from low to high. The analysis for each group is conducted by calculating w(obs) and comparing it to w(crit) as described in A3.3.5 of Annex A3. The calculations for each group of 20 are listed in Table X1.4. The critical value of the q-statistic is obtained from Table A3.3 as follows:

$$DF \, (\text{for } Sr) = (20 − 1) = 19 \qquad (X1.2)$$

(this is obtained from the secondary sampling operation)

$$n = 20 \qquad (X1.3)$$

(this is the size of each of the groups of 20)

Entering Table A3.3 at DF = 19 and n = 20, the critical value of q is 5.75. The value of w(crit) is obtained from Eq A3.17 of Annex A3. The listed calculations for the first group show that w(obs) is less than w(crit) and thus the first group of 20 is

homogeneous. For Group 2, w(obs) is greater than w(crit), which demonstrates that this group is not homogeneous.

X1.5.2 Figs. X1.1 and X1.2 give a good graphic illustration of the production of this IRM lot. Fig. X1.1, the average sample viscosity versus the sample number (in order of production and testing), shows the variation in the production-testing operation. Fig. X1.2, a plot of average sample viscosity versus the sample number in order of ascending viscosity magnitude for all 40 samples, clearly illustrates that the final
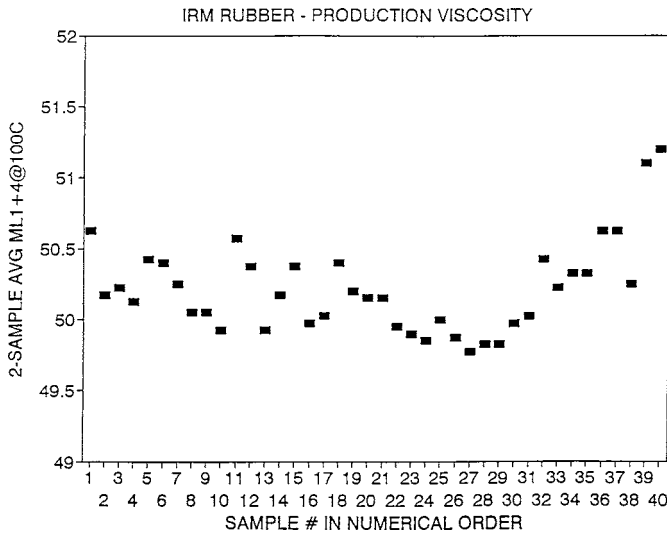
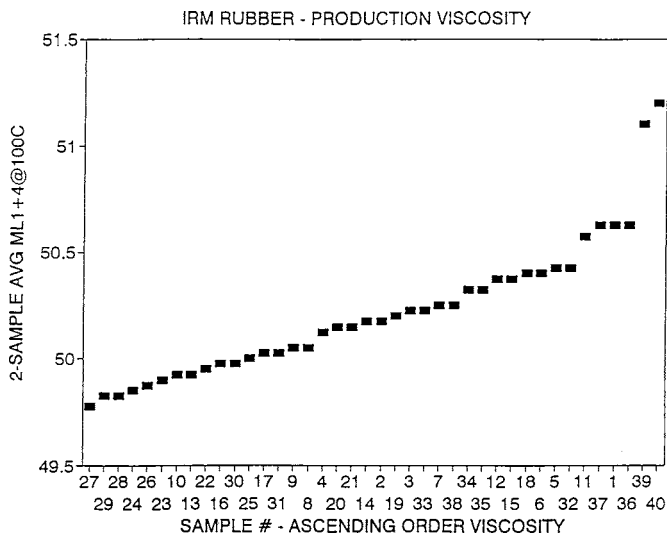FIG. X1.1 Production Mooney Viscosity—Testing and Production Order

FIG. X1.2 Production Mooney Viscosity—Samples in Ascending Mooney Order

TABLE X1.5 Trimmed Group 2—Lot Portions for Samples 39 and 40 Eliminated

Bale Sample[A]	MV1	MV2	AVG MV
27	49.8	49.8	49.78
29	49.9	49.8	49.83
28	49.9	49.8	49.83
24	49.8	49.9	49.85
26	50.0	49.8	49.88
23	49.9	49.9	49.90
22	50.0	49.9	49.95
30	50.0	50.0	49.98
25	50.0	50.0	50.00
31	50.0	50.1	50.03
21	50.1	50.2	50.15
33	50.2	50.3	50.23
38	50.5	50.0	50.25
35	50.6	50.1	50.33
34	50.4	50.3	50.33
32	50.5	50.4	50.43
37	50.7	50.6	50.63
36	50.7	50.6	50.63
Average	50.17	50.05	50.11
Standard Deviation	0.312	0.258	0.272

[A] For group of 18:
w(obs) = (50.63 – 49.78) = 0.85
w(crit) = 5.65 × 0.256 / 1.41 = 1.03
w(obs) < w(crit)

grand average ($\bar{X}n$) of the 38 samples.

X1.6.1 *Test Lot Limits*—The standard deviations for individual or single measurements of viscosity from Group 1 and (trimmed) Group 2 in Table X1.4 and Table X1.5 are 0.189, 0.270, 0.312, and 0.258, respectively, for the MV1 and MV2 columns of the two tables. The pooled variance for these four is 0.0681 and the square root of this is 0.261. This value of *Sr* is quite close to the *Sr* value of 0.256 from the secondary sampling operation. A pooled value for these two is 0.259. Therefore the test lot limits are given by Eq X1.4.

$$\text{test lot limits} = \pm3(0.259) = 0.776 = 0.78 \qquad (X1.4)$$

X1.6.2 *Test Lot Average*—The average of all four columns as described in X1.6.1 for MV1 and M2 in Table X1.4 and Table X1.5, gives the value

$$\text{Test Lot Average} = \bar{X}n = 50.16 \qquad (X1.5)$$

In X1.3 the number of lot samples was selected to be 40 on the basis that the maximum deviation between the test lot average and the true value of the lot would be E, where $E = 0.5$ *Sr*. Thus, $E = 0.5 \times 0.259 = 0.130$.

X1.7 *Homogeneity Data: Analysis of Variance* —Although not strictly needed for the evaluation of this Type *NB* IRM, the two-way analysis of variance (ANOVA) is presented, for the 38 samples of the homogeneous lot in Table X1.6, to illustrate how the calculations are performed for a Type *B* IRM. This analysis would be required in place of the analysis in X1.4 – X1.6 if this IRM were a Type *B* and if a blending operating had been conducted. The lot data of Table X1.6 are organized in the format as specified in Table A3.4 of Annex A3.

X1.7.1 The analysis is conducted by way of a spreadsheet procedure. Columns 2 and 3 of Table X1.6 list the replicate viscosities; columns 4 and 5 (although not strictly required for

production stages of the lot represented by samples 39 and 40 are much higher in viscosity compared to the remainder of the lot. Group 2 must be trimmed (reduced in size) to approach a homogeneous state. The 10 bales represented by these two samples must be eliminated.

X1.5.3 Table X1.5 lists the sorted average sample viscosity values for the remaining 18 samples after 39 and 40 are eliminated. With these two eliminated, w(obs) is now less than w(crit), and the remaining 18 samples represent a homogeneous portion or segment of the lot. Both groups together, samples 1 to 38, constitute a homogeneous lot.

X1.6 *Homogeneity Analysis: Step 3 Test Lot Limits and Test Lot Average*—The test lot limits are calculated from *Sr*. We have two sources to estimate *Sr*: (*1*) the value from the secondary sampling operation, 0.256, used for w(crit) calculations, and (*2*) the pooled standard deviation value from the 38 sample lot. The test lot average is obtained from the

TABLE X1.6 Analysis of Variance—IRM Homogeneity Mooney Data[A]

Sample No.	MV1	MV2	REP STD	REP VAR	RT = Sum MV1 + 2	(RT)^2	(MV1)^2	(MVS)^2
1	50.5	50.8	0.177	0.0313	101.25	10251.563	2550.2505	2575.563
2	50.0	50.4	0.247	0.0613	100.35	10070.123	2500.000	2535.123
3	50.0	50.5	0.318	0.1013	100.45	10090.203	2500.000	2545.203
4	50.3	50.0	0.247	0.0612	100.25	10050.063	2530.090	2495.003
5	50.3	50.6	0.177	0.0313	100.85	10170.722	2530.090	2555.303
6	50.2	50.6	0.283	0.0800	100.80	10160.640	2520.040	2560.360
7	50.2	50.3	0.071	0.0050	100.50	10100.250	2520.040	2530.090
8	50.0	50.1	0.071	0.0050	100.10	10020.010	2500.000	2510.010
9	50.0	50.1	0.071	0.0050	100.10	10020.010	2500.000	2510.010
10	50.0	49.9	0.106	0.0112	99.85	9970.023	2500.000	2485.023
11	50.6	50.6	0.035	0.0012	101.15	10231.323	2560.360	2555.303
12	50.3	50.5	0.106	0.0113	100.75	10150.563	2530.090	2545.203
13	50.0	49.9	0.106	0.0112	99.85	9970.023	2500.000	2485.023
14	50.0	50.4	0.318	0.1012	100.35	10070.123	2495.003	2540.160
15	50.2	50.6	0.318	0.1013	100.75	10150.563	2515.023	2560.360
16	50.0	50.0	0.035	0.0012	99.95	9990.003	2495.003	2500.000
17	50.0	50.1	0.106	0.0112	100.05	10010.003	2495.003	2510.010
18	50.3	50.5	0.141	0.0200	100.80	10160.640	2530.090	2550.250
19	50.1	50.3	0.141	0.0200	100.40	10080.160	2510.010	2530.090
20	50.2	50.1	0.071	0.0050	100.30	10060.090	2520.040	2510.010
21	50.1	50.2	0.071	0.0050	100.30	10060.090	2510.010	2520.040
22	50.0	49.9	0.071	0.0050	99.90	9980.010	2500.000	2490.010
23	49.9	49.9	0.000	0.0000	99.80	9960.040	2490.010	2490.010
24	49.8	49.9	0.071	0.0050	99.70	9940.090	2480.040	2490.010
25	50.0	50.0	0.000	0.0000	100.00	10000.000	2500.000	2500.000
26	50.0	49.8	0.177	0.0313	99.75	9950.063	2500.000	2475.063
27	49.8	49.8	0.035	0.0012	99.55	9910.203	2480.040	2475.063
28	49.9	49.8	0.106	0.0112	99.65	9930.123	2490.010	2475.063
29	49.9	49.8	0.106	0.0112	99.65	9930.123	2490.010	2475.063
30	50.0	50.0	0.035	0.0012	99.95	9990.003	2500.000	2495.003
31	50.0	50.1	0.035	0.0013	100.05	10010.003	2500.000	2505.003
32	50.5	50.4	0.106	0.0112	100.85	10170.722	2550.250	2535.123
33	50.2	50.3	0.035	0.0012	100.45	10090.203	2520.040	2525.063
34	50.4	50.3	0.106	0.0112	100.65	10130.423	2540.160	2525.063
35	50.6	50.1	0.389	0.1512	100.65	10130.423	2560.360	2505.003
36	50.7	50.6	0.106	0.0112	101.25	10251.563	2570.490	2555.303
37	50.7	50.6	0.106	0.0112	101.25	10251.563	2570.490	2555.303
38	50.5	50.0	0.354	0.1250	100.50	10100.250	2550.250	2500.000
CT =	1906.00	1906.75						
AVG (samples)	50.16	50.18						
VAR (samples)	0.0633	0.0831						
STD (samples)	0.252	0.288						
REP VAR pooled =				0.0282				
REP STD pooled =			0.168					
Sum (RT)^2 =						382562.9825		
Sum each column (MVi)^2: [where MV i = Yij] =							95603.290	95679.2725
Grand sum (MVi)^2 or (Yij)^2 =								191282.563

Calculations:

	MV1	MV2	
Sum = (CT)	1906.00	1906.75	Sum (Yij)^2 = 191282.5625
(CT)^2	3632836.000	3635695.563	Sum (RT)^2 = 382562.9825
Sum (CT)^2 =		7268531.5625	(Sum (RT)^2)/2 = 191281.4913
Sum (CT)^2/38 =		191277.146382	
GT = 3812.75			(GT)^2 = 14537062.56
(GT)^2/(2 × 38) = 191277.13898			
SSn = 4.35227			(Sr)^2 = ($SStot$ − SSn − SSk)/(38 − 1) (2 − 1)
SSk = 0.00740			(Sr)^2 = (5.42352 − 4.3523 − 0.00740) /37
$SStot$ = 5.42352			(Sr)^2 = 0.0288
			(Sr) = 0.170

[A] Samples 39 and 40 eliminated.
AVG, VAR, STD (samples) = Parameters among sample values each column.
(MV1), (MV2) or (MVi) = Yij of Table A3.4.
(RT)^2 = RT squared.
REP VAR = Pooled variance for replicate 1, 2 values.
REP STD = Standard Deviation for replicate 1, 2 values (pooled).
See Table A3.4 and Appendix X2 for additional details.

the ANOVA) give the 1 DF standard deviation and variance for the replicate pair for each sample. Column 6 gives the row totals (RT) or the sum of MV1 and MV2. Column 7 is RT^2; Columns 8 and 9 give the squares of each measured viscosity,

that is, MV1 and MV2 column values. Various values are given at the bottom of the table; the totals of Columns 2 and 3 designated as CT1 and CT2; the average, variance, and standard deviation of all individual sample values in Columns

2 and 3; the pooled replicate (Rep) standard deviation and variance; the sum of RT^2 and the sum of Columns 8 and 9 as well as the grand sum of all sample values $(Yij)^2$. Note in the table, the retention of at least three decimal places in the large numbers resulting from the calculations. As illustrated in the "Calculations" part of the table, these are required to obtain a sufficient number of significant figures when subtracting the various sums-of-squares.

X1.7.2 Table X1.6 gives the calculations needed to evaluate the residual standard deviation, Sr, which is equal to Srt, the test measurement variation, since for a Type B IRM there is no production process variation, Srp. The calculations are carried out according to the procedure outlined in Appendix X2. Three sum-of-squares (SS) are calculated; $SStot$, the total sum-of-squares, SSn due to samples and SSk due to replicates. Subtracting SSn and SSk from $SStot$ gives the residual sum-of-squares and this divided by the DF associated with the residual sum-of-squares gives the mean-square or variance, Sr^2. The square root of this is Sr, which has the value 0.170. Again this is the standard deviation for test measurements, that is, it is Srt.

X1.7.3 Although it is well to conduct the classical or normative ANOVA for the evaluation of a Type B IRM residual standard deviation, Sr (Srt in this case), the Column 4 calculations of Table X1.6 illustrate that a quick evaluation of Sr (or Srt) may be made. The pooled replicate standard deviation (Rep Std) given at the bottom of the table is 0.168. This is equal to the ANOVA Sr (Srt) value of 0.170 to within rounding error. This serves to better illustrate what is really evaluated by the residual standard deviation of the classical ANOVA.

X1.8 *Homogeneity Data: Evaluating the Production Process Variation* —Although again this is not strictly required, it is instructive to evaluate the variation due to the production process. For this Type NB IRM, a knowledge of Srt (as evaluated in the ANOVA) and a knowledge of Sr (the total variation due to testing and production process variation evaluated by the secondary sampling and testing operation) permits a calculation of Srp, the standard deviation due exclusively to the production process. Thus

$$SR \text{ (total)} = 0.259; \ Sr^2 \text{ (total)} = 0.0671 \qquad (X1.6)$$

$$Srt = 0.170; \ Srt^2 = 0.0289$$

Using the sum of variances relationship

$$Sr^2(\text{total}) = Srt^2 + Srp^2 \qquad (X1.7)$$

substituting

$$0.0671 = 0.0289 + Srp^2 \qquad (X1.8)$$

rearranging

$$Srp^2 = 0.0671 - 0.0289 = 0.0382 \qquad (X1.9)$$

$$Srp = 0.195$$

This calculation shows that the production process variation is slightly greater than the test measurement variation.

X1.9 *Accepted Reference Value and Between-Laboratory Limits*—Table X1.7 contains the interlaboratory test program (ITP) data. For the ITP, a single bale of rubber was selected from the center of the lot to be as close to the test lot average value of 50.16 as possible. Samples of this bale were sent out to all the participating laboratories. Mooney viscosity tests (ML1 + 4 at 100°C; single measurement) were conducted on two separate days, one week apart.

X1.9.1 Table X1.7 lists Day 1–Day 2 data in Columns 2 and 3; the average viscosity for both days in Column 4; the 1 DF variance and standard deviation for the Day 1–Day 2 data in Columns 5 and 6 and the Practice E691 h-value and k-value parameters in the last two columns. The bottom of the table lists the average, standard deviation, and variance calculated over all the laboratories for Columns 2, 3, and 4. Pooled values are given next for the individual Day 1–Day 2 variance and standard deviation and for the variance and standard deviation over all laboratories for individual viscosity values.

X1.9.2 An outlier analysis on the Mooney data was conducted via the Practice E691 precision program software and from the output, the h-values and k-values were listed as given in the last two columns. Critical values for h and k at the 95 % confidence level ($p = 0.05$) were calculated by way of A4.4.4.1 and A4.4.6.4, respectively, in Annex A4. These are listed at the bottom of the table as 1.90 and 1.94, respectively.

X1.9.2.1 *Mooney Viscosity Outliers*—Laboratory 14 has a calculated h-value of −2.59, which exceeds the critical h-value and is declared as a low Mooney value outlier. This is marked with an asterisk in Column 7. Fig. X1.3 is a plot of average viscosity versus the laboratory number in ascending viscosity value, generated from a spreadsheet sort operation. This clearly shows the departure of Laboratory 14 from the remaining laboratories.

X1.9.2.2 *Cell Standard Deviation Outliers*—Laboratory 5 has a calculated k-value of 3.21, which exceeds the critical k-value and is declared as an outlier. This is marked with an asterisk in Column 8.

X1.9.3 The lower part of the table lists the average and pooled values excluding the atypical or outlying Laboratories 5 and 14. Thus the AR value, as given by Eq X1.10, is equal to the 23-laboratory average viscosity.

$$\text{Accepted Reference Value} = 50.14 \qquad (X1.10)$$

Note that this AR value is equal to the test lot average of 50.16 to within 0.02 Mooney unit. Thus the selection of the center of the lot sampling bale as discussed in X1.9 was successful in being a good representation of the lot and a correction calculation using the dc correction factor as discussed in A4.4.5 is not required.

X1.9.4 The individual measurement standard deviation among all laboratories (excluding Laboratory 14), which represents typical or nominal testing variation (NTV) operation (see 8.1.7), is 0.744. The between-laboratory limits are given in terms of this standard deviation, which is designated as S_R.

$$\text{between} - \text{laboratory limits} = \pm 3(S_R) = \pm 3 \times 0.744 = 2.23$$
$$(X1.11)$$

By combining the AR value and the between-laboratory limits a statement may be made that describes the nominal testing variation (NTV) range. NTV conditions exist for any

TABLE X1.7 Interlaboratory Test Program Data to Establish Accepted Reference Value

Laboratory No.	ML1 + 4 at 100°C		AVG MV D1,D2	VAR D1D2	STD D1D2	Practice E691 Parameters	
	Day 1	Day 2				h-value	k-value
1	49.4	49.4	49.40	0.0000	0.000	−0.82	0.00
2	50.4	50.5	50.45	0.0050	0.071	0.51	0.16
3	48.7	49.6	49.15	0.4050	0.636	−1.14	1.44
4	49.5	50.0	49.75	0.1250	0.354	−0.38	0.80
5	50.0	52.0	51.00	2.0000	1.414	1.20	3.21[A]
6	50.9	51.2	51.05	0.0450	0.212	1.27	0.48
7	49.4	49.8	49.60	0.0800	0.283	−0.57	0.64
8	50.4	50.3	50.35	0.0050	0.071	0.38	0.16
9	49.6	49.8	49.70	0.0200	0.141	−0.44	0.32
10	51.6	50.9	51.25	0.2450	0.495	1.52	1.12
11	50.0	49.3	49.65	0.2450	0.495	−0.50	1.12
12	50.0	50.2	50.10	0.0200	0.141	0.07	0.32
13	50.1	49.8	49.95	0.0450	0.212	−0.12	0.48
14	47.9	48.1	48.00	0.2000	0.141	−2.59[B]	0.32
15	50.9	51.2	51.05	0.0450	0.212	1.27	0.48
16	48.8	49.9	49.35	0.6050	0.778	−0.88	1.76
17	50.0	50.0	50.00	0.0000	0.000	−0.06	0.00
18	49.2	49.5	49.35	0.0450	0.212	−0.88	0.48
19	51.0	50.8	50.90	0.0200	0.141	1.08	0.32
20	50.4	50.0	50.20	0.0800	0.283	0.19	0.64
21	50.1	50.7	50.40	0.1800	0.424	0.45	0.96
22	50.8	50.6	50.70	0.0200	0.141	0.82	0.32
23	51.2	50.3	50.75	0.4050	0.636	0.89	1.44
24	49.0	49.1	49.05	0.0050	0.071	−1.26	0.16
Average (Laboratories)	49.97	50.13	50.05	95 % Confidence Level Crit =		1.90	1.94
Standard Deviation (Laboratories)	0.886	0.810	0.791				
Variance (Laboratories)	0.7848	0.6559	0.6251				
Pooled Variance (Day 1–Day 2) =				0.1944			
Pooled Standard Deviation (Day 1–Day 2) =				0.441			
Pooled Variance (laboratories) for individual values =				0.7203			
Pooled Standard Deviation (laboratories) for individual values =				0.849			
Average (laboratories) excluding Laboratory 14 =				50.14			
Pooled Variance (laboratories) excluding Laboratory 14 =				0.5541			
Pooled Standard Deviation (laboratories) excluding Laboratory 14 =				0.744			
Pooled Variance (Day 1–Day 2) excluding Laboratory 5 =				0.1159			
Pooled Standard Deviation (Day 1–Day 2) excluding Laboratory 5 =				0.340			

[A] Outlier.
[B] Low Mooney value outlier.

group of laboratories if the measured individual IRM Mooney ML1 + 4 at 100°C values fall in the range (expressed to the nearest 0.1 Mooney unit) given by Eq X1.12.

$$\text{NTV Range} = 50.1 \pm 2.2 = 47.9 \text{ to } 52.3 \qquad (X1.12)$$

X1.9.5 The (pooled) within-laboratory (Day 1–Day 2) standard deviation (excluding Laboratory 5) is given at the bottom of the table as 0.340. This pooled or overall laboratory value includes good, intermediate, and poor laboratories from the standpoint of within-laboratory testing skill and control. This may be compared to the analogous within-laboratory (Rep 1—Rep 2) standard deviation of the organization (laboratory) producing the IRM, which is evaluated in Table X1.6 as 0.168 (or 0.170). The ratio of the two is 0.340/0.168 = 2.02.
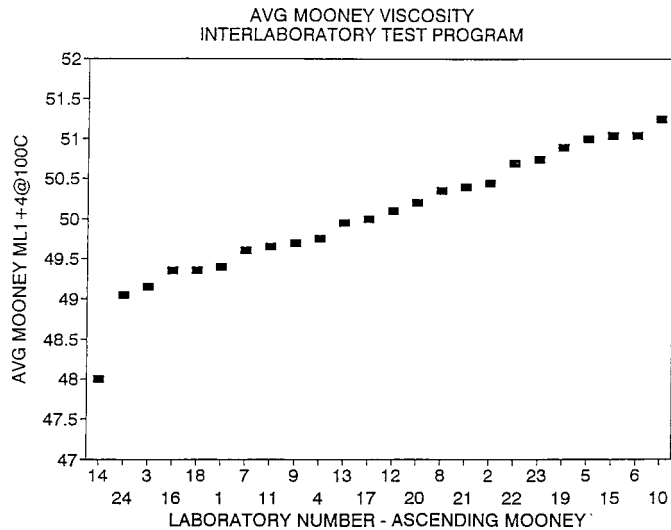
FIG. X1.3 Interlaboratory Mooney Viscosity—Laboratories in Ascending Mooney Order

X2. TWO-WAY ANALYSIS OF VARIANCE FOR CALCULATING *Sr*

X2.1 This appendix gives the calculation algorithms for calculating *Sr* for a blended material or Type *B* IRM, where the only component of residual variation is *Srt*, the testing variation.

X2.2 To calculate the residual standard deviation, *Sr*, and evaluate *w*(crit) for any lot of an IRM, arrange the data in the form of Table A3.4 in Annex A3. The columns of the table contain data for each replicate, there are a total of *k* replicates. Each row of the table represents one of the total of *n* samples. Column totals, *CTi*, and row totals, *RTj*, are taken as well as the grand total of all $k \times n$ values; this is defined as *GT*. The square of this is GT^2. Calculate also the average for each sample, from *X1* to *Xn*.

X2.3 Calculate the (corrected) sum of squares due to samples, *SSn*, according to Eq X2.1.

$$SSn = [(RT1^2 + RT2^2 + \dots RTj^2)/k] - (GT^2/k \times n) \quad (X2.1)$$

X2.4 Calculate the (corrected) sum of squares due to replicates, *SSk*, according to Eq X2.2.

$$SSk = [(CT\,1^2 + CT\,2^2 + \dots CTi^2 / n)] - (GT^2/k \times n) \quad (X2.2)$$

X2.5 Calculate the (corrected) total sum of squares, due to all individual data values, each represented by *Yij*, according to Eq X2.3.

$$SStot = [(\sum Yij^2) - (GT^2/k \times n)] \quad (X2.3)$$

X2.6 The residual variance, Sr^2, eliminating the effects of samples and replicates is calculated according to Eq X2.4.

$$Sr^2 = (SStot - SSn - SSk)/(n-1)(k-1) \quad (X2.4)$$

and

$$Sr = (Sr^2)^{1/2} \quad (X2.5)$$

X2.7 The total degrees of freedom, *DF*, is given by Eq X2.6.

$$DF(\text{total}) = (n-1)(k-1) \quad (X2.6)$$