



Standard Practice for Determining the Precision of ASTM Methods for Analysis and Testing of Industrial and Specialty Chemicals¹

This standard is issued under the fixed designation E 180; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ϵ) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 This practice establishes uniform standards for expressing the precision and bias of test methods for industrial and specialty chemicals. It includes an abridged procedure for developing this information, based on the simplest elements of statistical analysis. There is no intent to restrict qualified groups in their use of other techniques.

1.2 *This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety and health practices and determine the applicability of regulatory limitations prior to use.*

2. Referenced Documents

2.1 ASTM Standards:

- D 1013 Test Method for Total Nitrogen in Resins and Plastics²
- D 1727 Test Method for Urea Content of Nitrogen Resins³
- E 29 Practice for Using Significant Digits in Test Data to Determine Conformance with Specification⁴
- E 177 Practice for Use of the Terms Precision and Bias in ASTM Test Methods⁴
- E 178 Practice for Dealing with Outlying Observations⁴
- E 456 Terminology Relating to Quality and Statistics⁴
- E 691 Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method⁴
- E 1169 Guide for Conducting Ruggedness Tests⁴

3. Significance and Use

3.1 All test methods require statements of precision and bias. The information for these statements is generated by an interlaboratory study (ILS). This practice provides a specific design and analysis for the study, and specific formats for the precision and bias statements. It is offered primarily for the guidance of task groups having limited statistical experience.

3.2 It is recognized that the use of this simplified procedure will sacrifice considerable information that could be developed through other designs or methods of analyzing the data. For example, this practice does not afford any estimate of error to be expected between analysts within a single laboratory. Statements of precision are restricted to those variables specifically mentioned. Task groups capable of handling the more advanced procedures are referred to the literature (**1, 2, 3, 5, 13**)⁵ and specifically to Practice E 691, the current Committee E-11 practice for interlaboratory studies. The latter includes graphical display and interpretation of ILS data.

3.3 The various parts appear in the following order:

- Part A*—Glossary.
- Part B*—Preliminary Studies.
- Part C*—Planning the Interlaboratory Study.
- Part D*—Testing for Outlying Observations.
- Part E*—Statistical Analysis of Collaborative Data.
- Part F*—Format of Precision Statements.
- Part G*—Bias (Systematic Error).
- Part H*—Presentation of Data.

4. Keywords

4.1 bias; industrial chemicals; interlaboratory study; precision

PART A—GLOSSARY

5. Scope

5.1 The following statistical terms are defined in the sense in which they will be used in presenting precision and bias information. These definitions have been simplified and are not necessarily universally acceptable nor as defined in Terminology E 456 and Practice E 177. For definitions and explanations of other statistical terms used in this practice, refer to Terminology E 456 and Practice E 177.

6. Terminology

6.1 Definitions and Descriptions of Terms:

6.1.1 *accuracy*—the agreement between an experimentally determined value and the accepted reference value. In chemical work, this term is frequently used to express freedom from

⁵ The boldface numbers in parentheses refer to the list of references at the end of this practice.

¹ This practice is under the jurisdiction of ASTM Committee E15 on Industrial and Specialty Chemicals and is the direct responsibility of Subcommittee E15.01 on General Standards.

Current edition approved Sept. 10, 1999. Published December 1999. Originally published as E 180 – 61 T. Last previous edition E 180 – 94.

² *Annual Book of ASTM Standards*, Vol 06.03.

³ *Discontinued*. See 1983 *Annual Book of ASTM Standards*, Vol 06.01.

⁴ *Annual Book of ASTM Standards*, Vol 14.02.

bias, but in other fields it assumes a broader meaning as a joint index of precision and bias (see Practice E 177 and (4)). To avoid confusion, the term “bias” will be used in appraising the systematic error of test methods for industrial chemicals.

6.1.2 *bias*—a constant or systematic error as opposed to a random error. It manifests itself as a persistent positive or negative deviation of the method average from the accepted reference value.

6.1.3 *coefficient of variation*—a measure of relative precision calculated as the standard deviation of a series of values divided by their average. It is often multiplied by 100 and expressed as a percentage.

6.1.4 *duplicates*—two independent determinations performed by one analyst at essentially the same time.

6.1.5 *error*—in a statistical sense, any deviation of an observed value from the true, but generally unknown value. When expressed as a fraction or percentage of the value measured, it is called a relative error. All statements of precision or bias should indicate clearly whether they are expressed in absolute or relative sense.

6.1.6 *95 % limit (difference between two results)*—the maximum absolute difference expected for approximately 95 % of all pairs of results from laboratories similar to those in the interlaboratory study.

6.1.7 *precision*—the degree of agreement of repeated measurements of the same property. Precision statements in ASTM methods for industrial and specialty chemicals will be derived from the estimated standard deviation or coefficient of variation of a series of measurements and will be expressed in terms of the repeatability; the within-laboratory, between days variability; and the reproducibility of a method (see 6.1.14, 6.1.3, 6.1.10, 6.1.16, 6.1.12).

6.1.8 *random error*—the chance variation encountered in all experimental work despite the closest possible control of variables. It is characterized by the random occurrence of both positive and negative deviations from the mean value for the method, the algebraic average of which will approach zero in a long series of measurements.

6.1.9 *range*—the absolute value of the algebraic difference between the highest and the lowest values in a set of data.

6.1.10 *repeatability*—the precision of a method expressed as the agreement attainable between two independent determinations performed at essentially the same time (duplicates) by one analyst using the same apparatus and techniques. (see also 6.1.17.)

6.1.11 *replicates*—two or more repetitions of a test determination.

6.1.12 *reproducibility*—the precision of a method expressed as the agreement attainable between determinations performed in different laboratories (12).

6.1.13 *result*—a value obtained by carrying out the test method. The value can be a single determination, an average of duplicates, or other specified grouping of replicates.

6.1.14 *significance level*—the decimal probability that a result will exceed the critical value. (see 21.3 and 21.4.)

6.1.15 *standard deviation*—a measure of the dispersion of a series of results around their average, expressed as the positive square root of the quantity obtained by summing the squares of

the deviations from the average of the results and dividing by the number of observations minus one. It is also the square root of the variance and can be calculated as follows:

$$s = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n - 1}} \quad (1)$$

where:

s = estimated standard deviation of the series of results,

X_i = each individual value,

\bar{X} = average (arithmetic mean) of all values, and

n = number of values.

The following forms of this equation are more convenient for computation, especially when using a calculator:

$$s = \sqrt{\frac{\sum X^2 - (\sum X)^2/n}{n - 1}} \quad (2)$$

or

$$s = \sqrt{\frac{n\sum X^2 - (\sum X)^2}{n(n - 1)}} \quad (3)$$

where:

s = estimated standard deviation,

$\sum X^2$ = sum of the squares of all of the individual values,

$(\sum X)^2$ = square of the total of the individual values, and

n = number of values.

NOTE 1—Care must be taken in using either of these equations that a sufficient number of decimal places is carried in the sum of the values and in the sum of their squares so that serious rounding errors do not occur. For best results, all rounding should be postponed until after a value has been obtained for s .

In this practice, the standard deviation is obtained from the difference between duplicate determinations and from an analysis of variance of an interlaboratory test program (see Part E).

6.1.16 *variance*—a measure of the dispersion of a series of results around their average. It is the sum of the squares of the individual deviations from the average of the results, divided by the number of results minus one.

6.1.17 *within-laboratory, between-days variability (formerly called repeatability)*—the precision of a method expressed as the agreement attainable between independent determinations (each the average of duplicates) performed by one analyst using the same apparatus and techniques on each of two days. (This term is further defined and limited in 10.1.6, 25.1, and 25.2.9.2) (12).

PART B—PRELIMINARY STUDIES

7. Scope

7.1 This part covers the preliminary work that should be carried out in a few laboratories before undertaking a full interlaboratory evaluation of a method.

8. Discussion

8.1 When a task group is asked to provide a specific test procedure, there may be available one or more methods from the literature or from laboratories already performing such analyses. In such cases, these methods have usually been the subject of considerable research and any additional study of

variables, at this stage, would be wasteful of available task group time. It is recommended that such methods be rewritten in ASTM format, with full descriptions of the equipment and procedure, and be evaluated in a pilot run by a few laboratories on selected materials. Three laboratories and at least three such materials, using one or two analysts performing duplicate determinations on each of two days, by each method, constitutes a practical plan which can be analyzed by the procedures described in Part E—Statistical Analysis of Collaborative Data. Such a pilot study will confirm the adequacy of the methods and supply qualitative indications of relative precision and bias.

8.2 When the method to be evaluated is new, or represents an extensive modification of an available method, it is recommended that a study on variables be carried out by at least one laboratory to establish the parameters and conditions to be used in the description of the method. This should be followed by a three-laboratory pilot study before undertaking a full interlaboratory evaluation.

8.3 Detailed procedures for executing such preliminary studies are not described in this practice but are available in the general statistical literature.⁶ Practice E 691 and Guide E 1169 also provide information on this subject.

PART C—PLANNING THE INTERLABORATORY STUDY

9. Scope

9.1 This part covers some commonsense recommendations for the planning of interlaboratory studies.

10. Variables

10.1 The major variables to be considered are the following: methods, materials or levels, laboratories, apparatus, analysts, days, and runs. These are discussed as follows:

10.1.1 *Methods*—The preliminary studies of Part B should lead to agreement on a single method, which can then be evaluated in a full interlaboratory study. If it is necessary to evaluate two or more methods, the complete program must be carried out on each such method. In either case, it will be assumed that the method variables have been explored and that a well-standardized, fully detailed procedure has been prepared. Nothing short of this will justify the time and expense required for an extensive precision study.

10.1.2 *Materials or Levels*—The number of samples distributed should be held to the minimum needed to evaluate the method adequately. (Increasing the number of samples will not increase significantly the degrees of freedom (see 25.2.8) available for predicting the reproducibility of the method. This can be achieved only by increasing the number of laboratories.) Some interlaboratory studies can be limited to a single sample, as in the case of preparing a specific standard solution. Methods applicable to a single product of high purity can usually be evaluated with one or two samples. When different

concentrations of a constituent or values of a physical property are involved, the samples should represent the approximate lower, middle, and top levels of the expected range. If these vary over a wide range, the number of levels should be increased and spaced to cover the range. If technical grade products are used in a precision study, the bias of the method may be undeterminable unless the accepted reference value and its limits of error are known from other sources. For this reason, it is well to include one or more samples of known purity in the interlaboratory study.

10.1.3 *Laboratories*—To obtain a reliable precision estimate, it is recommended that the interlaboratory study include approximately ten qualified laboratories.⁷ When this number of independent laboratories cannot be recruited, advantage can be taken of a liberalized definition of collaborating laboratories, quoted as follows from the *ASTM Manual for Conducting an Interlaboratory Study of a Test Method (STP 335)*, p. 9 (5):

Here the term “collaborating laboratory” has a more specific meaning than in common usage. For example, a testing process often consists of an integrated sequence of operations using apparatus, reagents, and measuring instruments; and several more or less independent installations may be set up in the same area or “laboratory.” Each such participating installation should be considered as a collaborating laboratory so far as this procedure is concerned. Similarly, sets of test results obtained with different participants or under different conditions of calibration would in general constitute results from different collaborating laboratories even though they were obtained on the same sets of equipment.

This concept makes it possible to increase the available “laboratories” by using two analysts (but not more than two) in as many laboratories as needed to bring the total to the recommended minimum of ten. In such cases the two analysts must evaluate the method independently in the fullest sense of the word, interpreted as using different samples, different reagents, different apparatus where possible, and performing the work on different calendar days. (In the design in Section 16, laboratories using two analysts are designated as A-1, A-2, B-1, B-2, etc.) The most desirable laboratories and analysts are those having previous experience with the proposed method or with similar methods. It is essential that enough experience be acquired to establish confidence in the performance of a laboratory before starting the interlaboratory test series. Such preliminary work must be done with samples other than those to be used in the formal interlaboratory test program.

10.1.4 *Apparatus*—The effect of duplicate setups is not often a critical variable in chemical analysis. In instrumental methods, however, apparatus can become an important factor because the various laboratories may be using different makes or types of equipment, for example, the various colorimeters and spectrophotometers used in photometric methods. In such cases, the effect of apparatus becomes confounded with between-laboratory variability, and special care must be used to avoid misinterpreting the results. Of course, if enough laboratories have instruments of each type, “apparatus” can be made a planned variable in the study.

⁶ Task group chairmen are referred specifically to Youden, W. J. “Experimental Design and ASTM Committees,” *Materials Research & Standards*, MTRSA Vol 1, No. 11, November 1961, p. 862.

⁷ Practice E 691 insists on a minimum of six laboratories, but would prefer more than ten.



10.1.5 *Analysts*—The use of a single analyst in each “laboratory” (as defined in 10.1.3) is adequate to provide the information needed for calculating the within-laboratory, between-days variability and reproducibility of the method as defined in this practice. It is essential that all analysts complete the entire interlaboratory test program. With regard to analyst qualifications, an analyst who is proficient in the method should be selected.

10.1.6 *Days*—As defined in 6.1.17, the within-laboratory, between-days variability of the method shall be evaluated in terms of independent determinations by the same analyst. To achieve this, all scheduled determinations must be performed on each of two days (see Sections 16 and 25).

NOTE 2—As used in this practice, the term “days” represents replication of a set of determinations performed on any day other than that on which the first set was run. It may become a systematic variable to the extent that it is desirable that a given laboratory run the entire set of samples on one day and repeat the entire set on another. Although this may introduce a bias for that laboratory, there appears to be little chance that such a bias would be common to all laboratories. Where preliminary studies suggest that instability may result in an over-all systematic “days” effect, special planning will be required to take care of this problem.

10.1.7 *Runs*—The multiple determinations performed at the same time or within a very short time interval, on each day. In this practice, two runs (that is, duplicate determinations) are performed on each of two days.

11. Number of Determinations

11.1 Each analyst is required to perform duplicate determinations on each sample on each of two days. If one determination of a paired set is accidentally ruined, another pair must be run. An odd or unusual value does not constitute a “ruined” determination. In such cases, an additional set of duplicate determinations should be run and all values reported, with an assignable cause if at all possible.

12. Samples

12.1 One person should be made responsible for accumulating, subdividing, and distributing the materials to be used in the test program. Extra samples should be held in reserve to permit necessary replacement of any that may be lost or damaged in transit. Proper techniques in packaging and sampling should be followed, particularly with corrosive or otherwise hazardous materials. It is recommended that: all liquid samples be tested for closure leakage by laying the bottles on their side for 24 h prior to packaging, sample bottles be packed in boxes with strict attention to right side up labels, sample bottles be enclosed in plastic bags with plastic ties, packing of severely corrosive liquids be supervised by a technically trained person, and that strict attention be paid to DoT regulations. If a collaborating laboratory should receive a sample which shows evidence of leakage, or which is suspect for any other reason, the recipient should not use it but should immediately request a replacement.

12.2 The most important requirement is that the sampling units to be distributed to the participating laboratories be random selections from a reasonably homogeneous quantity (sample) of material. Single-phase liquids usually present no problem unless they are hygroscopic or unstable. Solid mix-

tures, in which the components vary in particle size, should be ground, sieved, and recombined to give a homogeneous product, and then checked (microscopically, or by any other available means) to confirm its homogeneity.

12.3 In the case of stable, homogeneous materials, one sampling unit can be distributed to each collaborating laboratory. If the material is hygroscopic, or otherwise unstable, multiple sampling units should be provided for each day’s run by each analyst.

12.4 Instability of any type may impose other restrictions on the execution of a planned program. It is the responsibility of the task group chairman to include in the plans for the interlaboratory study specific instructions on selecting, preparing, storing, and handling of the standard samples.

12.5 The sampling units distributed for the formal interlaboratory test program should not be used for practice runs. Where “dry-runs” are performed to develop proficiency in an inexperienced analyst or laboratory, this must be done on samples other than these.

13. Scheduling and Timing

13.1 Interlaboratory studies fail occasionally because no timetable had been established to cover the program, particularly in cases where the materials have changed in storage, after opening the container, etc. The instructions to the collaborators should cover such points as the time between receipt of samples and their testing, time elapsing between start and finish of the program, the order of performing the tests, etc., with particular attention to randomizing as a means of avoiding systematic errors.

NOTE 3—A discussion of randomizing is beyond the scope of this practice. Refer to standard textbooks on statistics and specifically to the indicated references (9, 10).

14. Instructions and Preliminary Questionnaire

14.1 Having decided on the variables and levels for each, the task group chairman should distribute to all participants a complete description of the planned collaborative study, emphasizing any special conditions or precautions to be observed. A detailed procedure and description of equipment, prepared in ASTM format, must be included. A questionnaire similar to the one in Table 1 will aid materially in the successful execution of the interlaboratory study.

15. Report Form

15.1 A form for reporting the essential data should be prepared and distributed (in duplicate) to all collaborators, who should be instructed on the number of decimal places to be used. It is recommended that interlaboratory studies be reported to one decimal place beyond that called for in the “Report” instructions of the method under study. Any subsequent rounding should be done by the task group chairman or the data analyst.

16. Design for an Interlaboratory Test Program

16.1 The plan given in Table 2 should cover most cases where laboratories and levels (or materials) are the principal variables. It calls for each analyst to perform two determinations in parallel on each of two days, at each level. Where

TABLE 1 Questionnaire on Interlaboratory Study

Title of Method (attached):

1. Our laboratory wishes to participate in the cooperative testing of this method for precision data. YES... NO...
2. As a participant, we understand that:
 - (a) All essential apparatus, chemicals, and other requirements specified in the method must be available in our laboratory when the program begins,
 - (b) Specified "timing" requirements (such as starting date, order of testing specimens, and finishing date) of the program must be rigidly met,
 - (c) The method must be strictly adhered to,
 - (d) Samples must be handled in accordance with instruction, and
 - (e) A qualified analyst must perform the tests.
 Having studied the method and having made a fair appraisal of our capabilities and facilities, we feel that we will be adequately prepared for cooperative testing of this method.
3. We can supply ___ qualified analysts. YES... NO...
4. Comments:

Signature

Company

TABLE 2 Single Method, Single Analyst, Ten Laboratories, N Levels or Materials Level or Material I

Laboratory		A	B	C	D	E	F	G	H	I	J
or Laboratory		A-1	A-2	B-1	B-2	C	D	E	F	G	H
Day 1	Run a
	Run b
Day 2	Run a
	Run b
Level or Material II											
Laboratory		A	B	C	D	E	F	G	H	I	J
or Laboratory		A-1	A-2	B-1	B-2	C	D	E	F	G	H
Day 1	Run a
	Run b
Day 2	Run a
	Run b
etc. to Level or Material N (N = 3 or Greater)											
Laboratory		A	B	C	D	E	F	G	H	I	J
or Laboratory		A-1	A-2	B-1	B-2	C	D	E	F	G	H
Day 1	Run a
	Run b
Day 2	Run a
	Run b

additional variables must be included, the proposed program should be referred to a statistician, the Subcommittee on Precision and Bias, or to Committee E-11 on Quality and Statistics for a specific recommendation.

PART D—TESTING FOR OUTLYING OBSERVATIONS

17. Scope

17.1 This part covers some elementary recommendations for dealing with outlying observations and rejection of data. Lacking a universally accepted practice for the rigid application of available statistical tests, considerable technical and common sense judgment must be exercised in using them. Accordingly, the following procedures are offered only as guides for the data analyst and all decisions to exclude or to include any suspect data shall be subject to the approval of the task group concerned. Rejection of data as outliers should be done only after attempts have been made to ascertain why the suspect values differ from other values; for example, a calculation error, transposition of digits, misunderstanding of, or failure to follow the test method provisions, etc.

NOTE 4—The test for outlying observations should be applied only once to a set of interlaboratory test data. Although two or more values can

be rejected simultaneously, in no case should the remaining data again be tested for outliers.

18. Principle of Method

18.1 The tests for outliers among the "multiple runs" and "different days" data are based on control chart limits for the range, as described in the *ASTM Manual on Presentation of Data and Control Chart Analysis, MNL 7, (14)*.

18.2 The test for outlying observations among laboratory averages is that described in Practice E 178.

18.3 The choice of significance levels for each of the three tests is based on practical experience gained from a number of interlaboratory studies involving chemical or physical properties.

NOTE 5—In choosing significance levels, there are two alternatives: (1) use of a low-significance level, accepting the divergent data, inflating variances, and perhaps failing to find significant differences, or (2) use of a higher significance level, rejecting the divergent data, deflating variances, and perhaps finding significance where none exists. In the case of multiple runs in an interlaboratory test program, the choice of the 0.001 level is based on the premise that only a high degree of divergence should justify rejection of data from a laboratory for this reason. The 0.01 level for days also reflects this premise. The 0.05 level for laboratories is frequently used and is chosen here because an outlying laboratory average, even at this significance level, may have a pronounced effect on

the claimed reproducibility of the method (see also 23.2).

18.4 The procedures are illustrated by data developed in an interlaboratory study on the determination of hydroxyl number (see Table 3).

19. Outliers Between Runs

19.1 Using the data of Table 3, tabulate the results of the duplicate runs on each of two days, in each of the eleven laboratories. Calculate the individual ranges and the average range as shown in Table 4.

19.2 Multiply the average range by the factor 3.488 to obtain the critical range at a 0.001 significance level. For the four materials in question, these values are:

Material	Average Range	Critical Range
Dodecanol	1.63	5.7
Ethylene glycol	18.69	65.2
Nonylphenol	1.52	5.3
Pentaerythritol	22.21	77.4

NOTE 6—The factor 3.488 is the D_4 value used to calculate the upper control limit for the range and is derived by the equation:

$$D_4 = 1 + td_3/d_2 \quad (4)$$

where $t = 3.291$, the two-tailed value of the “ t ” distribution for $p = 0.001$ and $DF = \infty$, $d_3 = 0.853$, and $d_2 = 1.128$.⁸

The following are the D_4 factors at other significance levels, for values of $n = 2, 3$, and 4:

Significance Level, %	$n = 2$	$n = 3$	$n = 4$
0.001	3.488	2.728	2.405
0.0027	3.267	2.575	2.282
0.01	2.947	2.352	2.100
0.05	2.482	2.029	1.837

19.3 Scan the individual ranges of Table 4 for values exceeding the critical range. For this example, the following occur:

Material	Critical Range	Observed Range	Suspect Laboratory
Dodecanol	5.7	(4.2, max)	none
Ethylene glycol	65.2	92.0	B
Nonylphenol	5.3	(3.0, max)	none
Pentaerythritol	77.4	101.9, 97.0	B, E

The data from the indicated laboratories are suspect as rejectable at a 0.001 significance level.

20. Outliers Between Days

20.1 Calculate the averages (to 0.1 unit) of the duplicate runs performed each day (see Table 3). Tabulate and determine the individual ranges and the average range as in Table 5.

20.2 Multiply the average range by the factor 2.947 (see Note 6) to obtain the critical range at a 0.01 significance level. Scan the individual ranges of Table 5 for values exceeding the critical range. For this example, the values are as follows:

Material	Average Range	Critical Range	Observed Range	Suspect Laboratory
Dodecanol	2.02	6.0	(6.0, max)	none
Ethylene glycol	10.2	30.1	32.3	B
Nonylphenol	2.25	6.6	9.4	C
Pentaerythritol	18.2	53.6	96.1	D

The data from the indicated laboratories are suspect as

TABLE 3 Hydroxyl Number Data—Acetylation Method

Material	Day	Run	Lab A	Lab B	Lab C	Lab D ^A	Lab E	Lab F	Lab G	Lab H	Lab I	Lab J	Lab K
Dodecanol	1	a	292.0	292.1	290.3	297.1	309.0	289.8	295.9	296.2	294.8	291.4	291.2
		b	294.6	288.0	291.1	296.9	311.0	288.7	294.9	296.7	295.8	292.2	289.9
		avg	293.3	290.0 ^B	290.7	297.0	310.0	289.2	295.4	296.4	295.3	291.8	290.6
	2	a	291.2	287.2	291.6	298.6	305.0	289.4	294.2	292.3	296.3	297.6	289.5
		b	293.4	287.2	289.2	301.4	303.0	289.6	293.5	294.8	294.0	293.4	290.6
		avg	292.3	287.2	290.4	300.0	304.0	289.5	293.8	293.6	295.2	295.5	290.0
Ethylene glycol	1	a	1767.0	1767.9	1798.0	1818.1	1783.0	1716.1	1782.0	1782.7	1805.4	1776.2	1778.3
		b	1790.0	1801.5	1809.0	1830.7	1787.0	1717.2	1760.0	1836.5	1789.3	1782.8	1755.8
		avg	1778.5	1784.7	1803.5	1824.4	1785.0	1716.6	1771.0	1809.6	1797.4	1779.5	1767.0
	2	a	1777.2	1706.4	1783.0	1817.4	1785.0	1725.7	1777.0	1801.6	1769.3	1781.7	1743.5
		b	1787.0	1798.4	1786.0	1848.6	1785.0	1721.7	1761.0	1817.6	1784.3	1783.7	1759.4
		avg	1782.1	1752.4	1784.5	1833.0	1785.0	1723.7	1769.0	1809.6	1776.8	1782.7	1751.4
Nonylphenol	1	a	248.8	243.8	261.8	250.1	248.0	245.0	246.7	249.3	246.9	244.3	242.3
		b	250.0	244.7	263.4	252.1	251.0	244.7	248.7	249.6	247.5	247.1	245.0
		avg	249.4	244.2	262.6	251.1	249.5	244.8	247.7	249.4	247.2	245.7	243.6
	2	a	247.2	245.2	273.0	249.7	245.0	245.2	249.7	246.5	247.7	247.8	243.2
		b	248.3	247.7	271.1	250.4	246.0	246.4	247.2	246.8	245.8	245.3	242.8
		avg	247.8	246.4	272.0	250.0	245.5	245.8	248.4	246.6	246.8	246.6	243.0
Pentaerythritol	1	a	1555.0	1551.0	1566.9	1469.5	1553.0	1492.2	1559.0	1611.2	1528.6	1537.1	1579.6
		b	1541.9	1449.1	1561.7	1484.3	1550.0	1492.7	1550.0	1566.6	1533.5	1530.6	1523.5
		avg	1548.4	1500.0	1564.3	1476.9	1551.5	1492.4	1554.5	1588.9	1531.0	1533.8	1551.6
	2	a	1550.8	1468.6	1567.1	1579.8	1531.0	1487.2	1560.0	1548.6	1540.3	1536.9	1565.3
		b	1555.5	1516.0	1558.3	1566.3	1628.0 ^C	1482.5	1560.0	1555.6	1533.7	1533.3	1529.6
		avg	1553.2	1492.3	1562.7	1573.0	1579.5	1484.8	1560.0	1552.1	1537.0	1535.1	1547.4

^ACondensers were rinsed with pyridine and crushed ice was added prior to titration of all samples.

^BAverages in this table are rounded to 0.1 because the method calls for reporting to 0.1 unit. Rounding follows the procedure shown in Section 2.3 of Practice E 29.

^CTemperature may have increased during titration.

TABLE 4 Outliers Between Runs

Laboratory	Day	Dodecanol			Ethylene Glycol			Nonylphenol			Pentaerythritol		
		Run a	Run b	Range	Run a	Run b	Range	Run a	Run b	Range	Run a	Run b	Range
A	1	292.0	294.6	2.6	1767.0	1790.0	23.0	248.8	250.0	1.2	1555.0	1541.9	13.1
	2	291.2	293.4	2.2	1777.2	1787.0	9.8	247.2	248.3	1.1	1550.8	1555.5	4.7
B	1	292.1	288.0	4.1	1767.9	1801.5	33.6	243.8	244.7	0.9	1551.0	1449.1	101.9
	2	287.2	287.2	0.0	1706.4	1798.4	92.0	245.2	247.7	2.5	1468.6	1516.0	47.4
C	1	290.3	291.1	0.8	1798.0	1809.0	11.0	261.8	263.4	1.6	1566.9	1561.7	5.2
	2	291.6	289.2	2.4	1783.0	1786.0	3.0	273.0	271.1	1.9	1567.1	1558.3	8.8
D	1	297.1	296.9	0.2	1818.1	1830.7	12.6	250.1	252.1	2.0	1469.5	1484.3	14.8
	2	298.6	301.4	2.8	1817.4	1848.6	31.2	249.7	250.4	0.7	1579.8	1566.3	13.5
E	1	309.0	311.0	2.0	1783.0	1787.0	4.0	248.0	251.0	3.0	1553.0	1550.0	3.0
	2	305.0	303.0	2.0	1785.0	1785.0	0.0	245.0	246.0	1.0	1531.0	1628.0	97.0
F	1	289.8	288.7	1.1	1716.1	1717.2	1.1	245.0	244.7	0.3	1492.2	1492.7	0.5
	2	289.4	289.6	0.2	1725.7	1721.7	4.0	245.2	246.4	1.2	1487.2	1482.5	4.7
G	1	295.9	294.9	1.0	1782.0	1760.0	22.0	246.7	248.7	2.0	1559.0	1550.0	9.0
	2	294.2	293.5	0.7	1777.0	1761.0	16.0	249.7	247.2	2.5	1560.0	1560.0	0.0
H	1	296.2	296.7	0.5	1782.7	1836.5	53.8	249.3	249.6	0.3	1611.2	1566.6	44.6
	2	292.3	294.8	2.5	1801.6	1817.6	16.0	246.5	246.8	0.3	1548.6	1555.6	7.0
I	1	294.8	295.8	1.0	1805.4	1789.3	16.1	246.9	247.5	0.6	1528.6	1533.5	4.9
	2	296.3	294.0	2.3	1769.3	1784.3	15.0	247.7	245.8	1.9	1540.3	1533.7	6.6
J	1	291.4	292.2	0.8	1776.2	1782.8	6.6	244.3	247.1	2.8	1537.1	1530.6	6.5
	2	297.6	293.4	4.2	1781.7	1783.7	2.0	247.8	245.3	2.5	1536.9	1533.3	3.6
K	1	291.2	289.9	1.3	1778.3	1755.8	22.5	242.3	245.0	2.7	1579.6	1523.5	56.1
	2	289.5	290.6	1.1	1743.5	1759.4	15.9	243.2	242.8	0.4	1565.3	1529.6	35.7
Total		$\Sigma R = 35.8$			$\Sigma R = 411.2$			$\Sigma R = 33.4$			$\Sigma R = 488.6$		
Number of runs		$n = 22$			$n = 22$			$n = 22$			$n = 22$		
Average range		$\bar{R} = 1.63$			$\bar{R} = 18.69$			$\bar{R} = 1.52$			$\bar{R} = 22.21$		

TABLE 5 Outliers Between Day Averages

Laboratory	Dodecanol			Ethylene Glycol			Nonylphenol			Pentaerythritol		
	Day 1	Day 2	Range	Day 1	Day 2	Range	Day 1	Day 2	Range	Day 1	Day 2	Range
A	293.3	292.3	1.0	1778.5	1782.1	3.6	249.4	247.8	1.6	1548.4	1553.2	4.8
B	290.0	287.2	2.8	1784.7	1752.4	32.3	244.2	246.4	2.2	1500.0	1492.3	7.7
C	290.7	290.4	0.3	1803.5	1784.5	19.0	262.6	272.0	9.4	1564.3	1562.7	1.6
D	297.0	300.0	3.0	1824.4	1833.0	8.6	251.1	250.0	1.1	1476.9	1573.0	96.1
E	310.0	304.0	6.0	1785.0	1785.0	0.0	249.5	245.5	4.0	1551.5	1579.5	28.0
F	289.2	289.5	0.3	1716.6	1723.7	7.1	244.8	245.8	1.0	1492.4	1484.8	7.6
G	295.4	293.8	1.6	1771.0	1769.0	2.0	247.7	248.4	0.7	1554.5	1560.0	5.5
H	296.4	293.6	2.8	1809.6	1809.6	0.0	249.4	246.6	2.8	1588.9	1552.1	36.8
I	295.3	295.2	0.1	1797.4	1776.8	20.6	247.2	246.8	0.4	1531.0	1537.0	6.0
J	291.8	295.5	3.7	1779.5	1782.7	3.2	245.7	246.6	0.9	1533.8	1535.1	1.3
K	290.6	290.0	0.6	1767.0	1751.4	15.6	243.6	243.0	0.6	1551.6	1547.4	4.2
Total	$\Sigma R = 22.2$			$\Sigma R = 112.0$			$\Sigma R = 24.7$			$\Sigma R = 199.6$		
Number of runs	$n = 11$			$n = 11$			$n = 11$			$n = 11$		
Average range	$\bar{R} = 2.02$			$\bar{R} = 10.18$			$\bar{R} = 2.25$			$\bar{R} = 18.15$		

rejectable at a 0.01 significance level.

21. Outliers Between Laboratory Averages

21.1 Calculate the laboratory averages (to 0.1 unit) and tabulate (Table 6).

21.2 Determine the standard deviation of the laboratory averages for each material using the calculating form of the formula given in Table 6.

21.3 Calculate the test criteria:

$$T_n = (X_n - \bar{X})/s \tag{5}$$

and

$$T_1 = (\bar{X} - X_1)/s \tag{6}$$

(see Table 6)

where:

X_n = largest laboratory average,

X_1 = smallest laboratory average,

\bar{X} = grand average of all laboratories, and

s = standard deviation of the laboratory averages.

21.4 From Table 7 obtain the critical value of T at the 0.05 significance level for $n = 11$. Comparing the observed with the critical values, the data show:

Material	Critical T	Observed T_n or T_1	Suspect Laboratory
Dodecanol	2.36	2.49	E
Ethylene glycol	2.36	(2.15, max)	none
Nonylphenol	2.36	2.88	C
Pentaerythritol	2.36	(1.86, max)	none

The data from the indicated laboratories are suspect as

TABLE 6 Outliers Between Laboratory Averages

Laboratory	Dodecanol		Ethylene Glycol		Nonylphenol		Pentaerythritol	
	Actual	Actual	$X - 1700^A$	Actual	$X - 200^A$	Actual	$X - 1400^A$	
A	292.8	1780.3	80.3	248.6	48.6	1550.8	150.8	
B	288.6	1768.6	68.6	245.3	45.3	1496.2	96.2	
C	290.6	1794.0	94.0	267.3	67.3	1563.5	163.5	
D	298.5	1828.7	128.7	250.6	50.6	1525.0	125.0	
E	307.0	1785.0	85.0	247.5	47.5	1565.5	165.5	
F	289.4	1720.2	20.2	245.3	45.3	1488.6	88.6	
G	294.6	1770.0	70.0	248.0	48.0	1557.2	157.2	
H	295.0	1809.6	109.6	248.0	48.0	1570.5	170.5	
I	295.2	1787.1	87.1	247.0	47.0	1534.0	134.0	
J	293.6	1781.1	81.1	246.2	46.2	1534.4	134.4	
K	290.3	1759.2	59.2	243.3	43.3	1549.5	149.5	
$\Sigma X =$	3235.6		883.8		537.1		1535.2	
$\Sigma X^2 =$	952006.02		78767.20		26638.37		221744.24	
$(\Sigma X)^2 =$	10469107.36		781102.44		288476.41		2356839.04	
$(\Sigma X)^2/n =$	951737.03		71009.31		26225.13		214258.09	
$s =$	$\sqrt{952006.02 - 951737.03 / 11 - 1}$ $\sqrt{78767.20 - 71009.31 / 11 - 1}$ $\sqrt{26638.37 - 26225.13 / 11 - 1}$ $\sqrt{221744.24 - 214258.09 / 11 - 1}$							
$s =$	5.19		27.9		6.43		27.4	
$\bar{X} =$	294.1		1780.3		248.8		1539.6	
$T_n =$	$307.0 - 294.1 / 5_{.19} = 2.49$		$1828.7 - 1780.3 / 27_{.9} = 1.73$		$267.3 - 248.8 / 6_{.43} = 2.88$		$1570.5 - 1539.6 / 27_{.4} = 1.13$	
$T_I =$	$294.1 - 288.6 / 5_{.19} = 1.06$		$1780.3 - 1720.2 / 27_{.9} = 2.15$		$248.8 - 243.3 / 6_{.43} < 1$		$1539.6 - 1488.6 / 27_{.4} = 1.86$	

^ATo avoid handling large numbers and thus simplify the calculations, the data have been "coded" by subtracting the indicated constant (*K*) from each value. The coded values were used to calculate the standard deviation directly. The mean, \bar{X} , is obtained by the following equation:

$$\bar{X} = \Sigma X/n + K$$

Example: Ethyleneglycol
 $\bar{X} = 883.8/11 + 1700 = 1780.3$

TABLE 7 Critical Values for *T* When Standard Deviation is Calculated from Present Sample

NOTE 1—From Table 1 of Practice E 178. Based on available literature (8), these significance levels have been doubled to take account of the fact that in actual practice the criterion is applied to either the smallest or the largest observation (or both) as the case happens to be. Adjustment of these values was also made for division by *n* - 1 instead of *n* in calculating *s*.

Number of Observations, <i>n</i>	0.05 Significance Level	0.01 Significance Level
3	1.15	1.15
4	1.48	1.50
5	1.71	1.76
6	1.89	1.97
7	2.02	2.14
8	2.13	2.27
9	2.21	2.39
10	2.29	2.48
11	2.36	2.56
12	2.41	2.64
13	2.46	2.70
14	2.51	2.76
15	2.55	2.81
16	2.59	2.85
17	2.62	2.89
18	2.65	2.93
19	2.68	2.97
20	2.71	3.00
21	2.73	3.03
22	2.76	3.06
23	2.78	3.09
24	2.80	3.11
25	2.82	3.14

rejectable at a 0.05 significance level.

21.5 Practice E 178 also indicates, in 4.3, that an alternative system based entirely on ratios of simple differences among the observations is given in the literature (6, 7, 11). This system

may be used if it is felt highly desirable to avoid calculation of *s*.

22. Summary

22.1 The results of Sections 19, 20, and 21 can be summarized as follows:

Material	Test Results Regarded as Suspect		
	Runs (0.001)	Days (0.01)	Laboratory Averages (0.05)
Dodecanol	none	none	E
Ethylene glycol	B	B	none
Nonylphenol	none	C	C
Pentaerythritol	B and E	D	none

23. Discussion

23.1 When the above operations show any set of data from a laboratory to be suspect, every effort should be made to find an assignable cause that will justify rejection.

23.2 As Practice E 180 does not provide procedures for the analysis of data in which values are missing, rejection in any one of the three categories (runs, day, or laboratories) makes it necessary to exclude from the analysis of variance all of the data from that laboratory pertinent to the material or sample in question.

NOTE 7—Only the outliers between runs need be eliminated from the repeatability calculations, as illustrated in 25.2.7.

23.3 Although rejected data are usually excluded before performing the analysis of variance, it is advisable to perform the analysis using the entire set, as well as after the elimination of the suspect data. With a calculator, this will entail relatively little additional work and the comparative data are often helpful in appraising the results of the entire program, as well as in deciding whether or not the rejection is justified.

PART E—STATISTICAL ANALYSIS OF COLLABORATIVE DATA

24. Scope

24.1 This part demonstrates the statistical analysis of typical data obtained with the design of Section 16.

24.2 The abridged analysis of variance gives the basic information needed for calculating within-laboratory, between days variability and reproducibility as defined in this practice. It determines the between-laboratories and within-laboratory, between-days variances for each level and combines them to give the two pertinent standard deviations or coefficients of variation.

24.3 Because it disregards interactions, this simplified procedure sacrifices information that could be developed by using conventional methods for the analysis of variance. Task groups capable of handling such procedures are referred to the literature (1, 2, 3, 5, 13) and specifically to the *ASTM Manual for Conducting an Interlaboratory Study of a Test Method (STP 335) (5)*.

25. Analysis of Variance

25.1 The abridged analysis of variance is illustrated in the following sections by two examples representing collaborative studies of single methods involving several levels or materials and an adequate number of laboratories, with one qualified analyst in each carrying out two determinations (paired duplicates) on each of two days. Although by some definitions the repeatability estimate can be based on the variation between paired duplicates, experience in chemical testing shows that such estimates are usually more optimistic and imply a superior level of precision than when they are derived from independent determinations performed on different days. To conform to the definitions for repeatability and reproducibility conditions in Terminology E 456, this practice uses the duplicate results for calculating the repeatability standard deviation (or coefficient of variation) (see 25.2.7 and 25.2.9.1). Estimates of the within-laboratory, between-days variability and reproducibility are based on the averages of the duplicate determinations obtained on each of two days. Accordingly, the analysis of variance determines the within-laboratory, between-days variance and the between-laboratories variance for each sample and provides for combining (pooling) the data for all samples to give overall standard deviations (or coefficients of variation) which are used to calculate the within-laboratory, between-days variability and reproducibility of the method.

25.2 *Example A*—This example illustrates the use of coefficients of variation. See Example B for a case where the

standard deviations can be used directly.

25.2.1 *Specific Example*—Four materials (dodecanol, nonylphenol, pentaerythritol, and ethylene glycol) were analyzed for hydroxyl number by a single analyst, in each of eleven laboratories. The entire set of data is shown in Table 8. Only the results for dodecanol are used in the following sections to demonstrate the analysis of variance technique.

25.2.2 *Homogeneity of Data and Testing for Outliers*—The usual tests for homogeneity and normality are beyond the scope of this simplified procedure.⁹ On applying the tests for outliers 21.4, the results of Laboratory E were excluded because of a divergent value among the laboratory averages. Table 8 shows the remaining data (as the averages of the duplicate determinations).

25.2.3 *Coded Data*—To avoid handling large numbers in the analysis of variance, the data are coded by subtracting 280 from each value, as shown in Table 9.

25.2.4 *Analysis of Variance*—Perform the following operations on the coded data.

25.2.4.1 Square the individual values and add them, as follows:

$$\begin{aligned} &13.3^2 + 10.0^2 + 10.7^2 + \dots + 15.2^2 \\ &\quad + 15.5^2 + 10.0^2 \\ &= 3505.0600 \end{aligned} \tag{7}$$

25.2.4.2 Square the column totals, add, divide by the number of values in each column, as follows:

$$25.6^2 + 17.2^2 + \dots + 27.3^2 + 20.6^2/2 = 3483.8200 \tag{8}$$

25.2.4.3 Add the individual values, square this total, divide by the number of values, as follows:

$$(13.3 + 10.0 + \dots + 15.5 + 10.0)^2/20 = 3307.5920 \tag{9}$$

25.2.4.4 Using Eq 7, Eq 8, and Eq 9 to complete the analysis of variance as shown in Table 10, the components of variance should then be calculated as follows:

$$\begin{aligned} s_a^2 &= 2.1240 \text{ and } s_a = \sqrt{2.1240} = 1.46 \tag{10} \\ s_b^2 &= (19.5809 - s_a^2)/2 \\ &= (19.5809 - 2.1240)/2 \\ &= 17.4569/2 \\ &= 8.7284 \\ s_{a+b}^2 &= s_a^2 + s_b^2 = 2.1240 + 8.7284 = 10.8524 \end{aligned}$$

⁹ Refer to any standard textbook on statistics, specifically to the sections on the Homogeneity of Variances, Bartlett Test, etc.

TABLE 8 Averages of Duplicate Determinations-Dodecanol

Laboratory	A	B	C	D	F	G	H	I	J	K
Day No. 1	293.3	290.0	290.7	297.0	289.2	295.4	296.4	295.3	291.8	290.6
Day No. 2	292.3	287.2	290.4	300.0	289.5	293.8	293.6	295.2	295.5	290.0
Totals	585.6	577.2	581.1	597.0	578.7	589.2	590.0	590.5	587.3	580.6

TABLE 9 Data from Table 8 Coded

Laboratory	A	B	C	D	F	G	H	I	J	K
Day No. 1	13.3	10.0	10.7	17.0	9.2	15.4	16.4	15.3	11.8	10.6
Day No. 2	12.3	7.2	10.4	20.0	9.5	13.8	13.6	15.2	15.5	10.0
Totals	25.6	17.2	21.1	37.0	18.7	29.2	30.0	30.5	27.3	20.6

TABLE 10 Analysis of Variance—Example A

Source of Variance	Sum of Squares, SS	Degrees of Freedom, DF	Mean Square	Expected Mean Square
Between laboratories	Eq 8 – Eq 9	$m - 1$	SS/DF	$s_a^2 + ns_b^2$
Within laboratory, between days	Eq 7 – Eq 8	$m(n - 1)$	SS/DF	s_a^2
Total	Eq 7 – Eq 9	$mn - 1$		

where:

m = number of columns (laboratories),

n = number in each column (days),

SS = sum of squares,

DF = degrees of freedom,

s_b^2 = variance due to differences between columns (laboratories),

and

s_a^2 = variance due to differences within columns

(days).

Example for Dodecanol

Source of Variance	Sum of Squares, SS	Degrees of Freedom, DF	Mean Square	Expected Mean Square
Between laboratories	3483.8200 – 3307.5920 = 176.2280	10 – 1 = 9	176.2280/9 = 19.5809	$s_a^2 + 2s_b^2$
Within laboratory, between days	3505.0600 – 3483.8200 = 21.2400	10(2 – 1) = 10	21.2400/10 = 2.1240	s_a^2
Total	3505.0600 – 3307.5920 = 197.4680	(10 × 2) – 1 = 19		

$$s_{a+b} = \sqrt{10.8524} = 3.29$$

where:

s_a = estimated standard deviation of a single result (average of duplicates) within-laboratory, between-days, based on 10 degrees of freedom, and

s_{a+b} = estimated standard deviation of a single result (average of duplicates) in any laboratory, based on approximately 9 degrees of freedom.

25.2.4.5 The mean square for between laboratories (19.5809 in the dodecanol example, Table 10) is expected to be significantly greater than that for between days (2.1240, Table 10) because of the additional variability due to laboratories. This condition is generally true, but should be verified with the F-test which is the ratio of the mean square for between laboratories to the mean square for between days. For the example, $F = 19.5809/2.1240 = 9.22$. The critical value for F with 9 and 10 DF at the 0.05 level of significance is 3.02. The critical F value is obtained from tables in any standard statistical text book. In this example, the critical value is exceeded, and the mean square for between laboratories is considered significantly greater than that for between days. This means that calculations for s_b^2 , s_{a+b}^2 , and s_{a+b} in 25.2.4.4

are valid. If the critical value for F is not exceeded, the mean square for between laboratories has not been shown to be significantly greater than that for between days. This means that the between-laboratory effect is not considered to be significant, and s_b^2 is zero. In this case, the values for s_{a+b}^2 and s_{a+b} are set equal to s_a^2 and s_a , respectively.

25.2.4.6 Calculate the coefficient of variation percents (CV %) as follows:

$$CV_a \% = \left(\frac{s_a \times 100}{\bar{X}} \right) \quad (11)$$

$$CV_{a+b} \% = \left(\frac{s_{a+b} \times 100}{\bar{X}} \right) \quad (12)$$

25.2.5 *Other Materials*—Perform analyses of variance on the data for the other three materials, using the above example as a model. These are not illustrated, but the results are shown in Table 11.

25.2.6 *Pooling of Data*—The tabulated values should exhibit one of the following three patterns: (1) the s_a or the s_{a+b} values, or both, in good agreement for the four materials, (2) the coefficients of variation agreeing for the four materials, or (3) neither showing the desired uniformity. In Table 11, it is

TABLE 11 Summary of Data for Four Materials—Example A

Material	Average OH Number	Within-Laboratory, Between Days			Single Result, Any Laboratory		
		Degrees of Freedom, DF	s_a	Coefficient of Variation, %	Degrees of Freedom, DF	s_{a+b}	Coefficient of Variation, %
Dodecanol	292.9	10	1.46	0.50	9	3.29	1.13
Nonylphenol	247.0	10	1.32	0.53	9	2.25	0.91
Pentaerythritol	1543.6	8	9.76	0.63	7	26.53	1.72
Ethylene glycol	1781.5	10	7.68	0.43	9	29.59	1.66

evident that the standard deviations differ widely and, therefore, cannot be pooled. The coefficients of variation for the between-days, within-laboratories data are in excellent agreement and an overall coefficient can be calculated by pooling them as follows:

$$CV_a \% \text{ (overall)} = \sqrt{\frac{(DF_1 \times CV_1^2 \%) + \dots (DF_n \times CV_n^2 \%)}{DF_1 + \dots DF_n}} \tag{13}$$

$$= \sqrt{\frac{(10 \times 0.50^2) + (10 \times 0.53^2) + (8 \times 0.63^2) + (10 \times 0.43^2)}{10 + 10 + 8 + 10}}$$

$$= 0.52 \% \tag{14}$$

The between-laboratories data show good agreement in the coefficients of variation for dodecanol and nonylphenol, as well as good agreement between those for pentaerythritol and ethylene glycol, but there is a significant spread between the two groups and most task groups would hesitate to combine such data for the entire set. Therefore, the proper action is to report separate coefficients of variation for the two groups.

NOTE 8—The following statistical tests are useful for determining whether or not the standard deviations can be pooled:

Cochran Test: Eisenhard, C., Hastay, M. W., and Wallis, W. A., "Techniques of Statistical Analysis," McGraw-Hill Book Co., Inc., New York, NY, 1947, p. 388.

Hartley Test: Bowker, A. H., and Lieberman, G. J., "Handbook of Industrial Statistics," Prentice-Hall, Inc., Englewood Cliffs, NJ, 1955, p. 952.

The coefficient of variation for hydroxyl values in the 250 to 300 range is calculated as follows:

$$CV_{a+b} \% = \sqrt{\frac{(9 \times 1.13^2) + (9 \times 0.91^2)}{9 + 9}}$$

$$= 1.03 \% \tag{15}$$

Similarly, the coefficient for values in the 1500 to 1800 range is calculated as follows:

$$CV_{a+b} \% = \sqrt{\frac{(7 \times 1.72^2) + (9 \times 1.66^2)}{7 + 9}}$$

$$= 1.69 \% \tag{16}$$

NOTE 9—If the s_a and s_{a+b} values (rather than the coefficients) should show good agreement, the mathematical procedure for pooling them is analogous to that shown in 25.3.3.

25.2.7 *Repeatability*—A useful precision estimate can be obtained from the values for the duplicate determinations in the form of the permissible range for such paired determinations. The standard deviation for duplicates can be calculated from the original data for paired determinations as illustrated for dodecanol in Table 12.

$$s \text{ (from duplicates)} = \sqrt{\frac{\text{sum of the squares of all differences}}{2 \times \text{number of sets}}} \tag{17}$$

$$= \sqrt{\frac{87.40}{2 \times 22}} \tag{18}$$

$$= 1.41, \text{ based on 22 degrees of freedom} \tag{19}$$

The data for the other three materials are analyzed similarly, after eliminating outliers between runs (19.3). These operations

TABLE 12 Results of Duplicate Runs—Example A

Run No. 1	Run No. 2	Difference	Difference Squared
292.0	294.6	2.6	6.76
291.2	293.4	2.2	4.84
292.1	288.0	4.1	16.81
287.2	287.2	0.0	...
290.3	291.1	0.8	0.64
291.6	289.2	2.4	5.76
297.1	296.9	0.2	0.04
298.6	301.4	2.8	7.84
309.0	311.0	2.0	4.00
305.0	303.0	2.0	4.00
289.8	288.7	1.1	1.21
289.4	289.6	0.2	0.04
295.9	294.9	1.0	1.00
294.2	293.5	0.7	0.49
296.2	296.7	0.5	0.25
292.3	294.8	2.5	6.25
294.8	295.8	1.0	1.00
296.3	294.0	2.3	5.29
291.4	292.2	0.8	0.64
297.6	293.4	4.2	17.64
291.2	289.9	1.3	1.69
289.5	290.6	1.1	1.21
Total			87.40

are not illustrated, but the results are summarized in Table 13. As was the case in 25.2.6, the full set cannot be pooled, but the coefficients of variation for dodecanol and nonylphenol can be combined to give an overall value for the 250 to 300 range, and the pentaerythritol and ethylene glycol coefficients can be combined for the 1500 to 1800 range. Using the first pair as an example,

$$CV \% = \sqrt{\frac{(22 \times 0.48^2) + (22 \times 0.50^2)}{22 + 22}}$$

$$= \sqrt{\frac{22 \times (0.48^2 + 0.50^2)}{44}}$$

$$= \sqrt{\frac{0.4804}{2}}$$

$$= \sqrt{0.2402}$$

$$= 0.49 \% \tag{20}$$

25.2.8 *Degrees of Freedom*—Calculation of the exact number of degrees of freedom applicable to the pooled coefficient of variation (or to the pooled standard deviation) is a complex procedure that is beyond the scope of this practice. Concerning the reproducibility in a universe of laboratories based on a study among m laboratories, a conservative estimate of $(m - 1)$ degrees of freedom is used. For the within-laboratory, between-days variability of the method, the available degrees of freedom can be approximated from the following equation:

TABLE 13 Standard Deviation and Coefficients of Variation for Repeatability (from Duplicates)—Example A

Material	Average OH Number	Degrees of Freedom, DF	Standard Deviation	Coefficient of Variation, %
Dodecanol	294.15	22	1.41	0.48
Nonylphenol	248.84	22	1.24	0.50
Pentaerythritol	1539.56	20	15.53	1.01
Ethylene glycol	1781.67	21	14.00	0.79

$$DF = k \text{ materials or levels} \times m \text{ laboratories} \times (n - 1) \text{ days} \quad (21)$$

In view of the fact that tests for outlying observations may reject some data and result in different values of m for each level of material, it is more correct to calculate the total degrees of freedom by adding the DF values for the pertinent materials or levels. For the example cited, the within-laboratory, between-days DF values of Table 11 are used. With regard to checking limits for duplicates, the available DF can be approximated as follows:

$$DF = k \text{ materials or levels} \times m \text{ laboratories} \times n \text{ days} \times (r - 1) \text{ multiples} \quad (22)$$

where r = number of replications (always two in this practice)

These values are shown in Table 13.

25.2.9 Calculation of Precision Limits—The following precision estimates should be calculated from the pertinent coefficients of variation of the preceding paragraphs, illustrated as follows:

25.2.9.1 Repeatability (95 % Probability)—Multiply the coefficient of variation for duplicate runs by 2.8 ($\approx 1.96 \sqrt{2}$). For the example cited in 25.2.7, where $CV\% = 0.49\%$, $0.49 \times 2.8 = 1.4\%$ relative, at the 250 to 300 level, the 95 % limit of range for duplicate values.

25.2.9.2 Within-Laboratory, Between-Days Variability (95 % Probability)—Similarly, multiply the overall coefficient of variation for the within-laboratory, between-days data by 2.8. In this case, where $CV_a\% = 0.52$, $0.52 \times 2.8 = 1.5\%$ relative, the 95 % limit of the range between two values (each the average of duplicates obtained by the same analyst on different days).

25.2.9.3 Reproducibility (95 % Probability)—These values are calculated in accordance with 25.2.9.2 except that the over-all coefficient of variation for the between-laboratories data is multiplied by 2.8. For the example cited at the 250 to 300 level, where the pooled coefficient of variation = 1.03 % relative, the 95 % limit of the range of two values = $1.03 \times 2.8 = 2.88\%$ relative.

NOTE 10—In the above examples, the coefficients of variation were multiplied by 2.8 because these had been pooled in 25.2.6. If the standard deviations had proven poolable, the overall s_a and s_{a+b} values would have been used. These operations are illustrated in 25.3.

25.3 Example B—The following example illustrates a case where the standard deviations are in agreement and are pooled to give overall standard deviations and precision statements on an absolute basis.

25.3.1 Specific Example—Three materials containing 24, 12, and 0 % levels of Component X were analyzed by one analyst in each of ten laboratories, who performed duplicate determinations and repeated the entire series one day later.

25.3.2 Summary of Data—To conserve space, the individual

results and the analysis of variance are not shown. The results are summarized in Table 14.

25.3.3 Pooling of Data—It is obvious that the standard deviations show excellent agreement. Accordingly, the overall standard deviations are obtained by pooling as follows:

$$\begin{aligned} s_a(\text{overall}) &= \sqrt{\frac{(DF_1 \times (s_{a1})^2) + \dots (DF_n \times (s_{an})^2)}{DF_1 + \dots DF_n}} \\ &= \sqrt{\frac{(10 \times 0.16^2) + (10 \times 0.20^2) + (10 \times 0.14^2)}{10 + 10 + 10}} \\ &= 0.17 \end{aligned} \quad (23)$$

$$\begin{aligned} s_{a+b}(\text{overall}) &= \sqrt{\frac{(DF_1 \times (s_{a+b1})^2) + \dots (DF_n \times (s_{a+b}n)^2)}{DF_1 + \dots DF_n}} \\ &= \sqrt{\frac{(9 \times 0.39^2) + (9 \times 0.30^2) + (9 \times 0.34^2)}{9 + 9 + 9}} \\ &= 0.35 \end{aligned} \quad (24)$$

25.3.4 Calculation of Precision Estimates—The precision estimates are calculated as shown in 25.2.9, except that the standard deviations are used instead of the coefficients of variation. These estimates and the pertinent data are shown in Table 15.

PART F—FORMAT OF PRECISION STATEMENTS

26. Principle

26.1 The formal statements of repeatability and reproducibility of methods for industrial chemicals should include the estimated standard deviations or coefficients of variation, the degrees of freedom, and the 95 % limits on the difference (range) between two test results.

26.2 These estimates should be obtained by the procedures outlined in Part E or by equivalent statistical methods.

27. Example (Using the Data of Table 15, Example B)

27.1 The following form and typical wording are recommended for the precision statements that appear in the Precision and Bias section of the test method:

28. Precision and Bias

28.1 Precision—The following criteria should be used for judging the acceptability of results (see Note 11):

28.1.1 Repeatability (Single Analyst)—The standard deviation for a single determination has been estimated to be 0.22 % absolute at 60 DF. The 95 % limit for the difference between two such runs is 0.6 % absolute.

28.1.2 Laboratory Precision (Within-Laboratory, Between-Days Variability, Formerly Called Repeatability)—The standard deviation of results (each the average of duplicates),

TABLE 14 Summary of Data for Three Levels—Example B

Mean Level Component X, %	Within-Laboratory, Between Days			Single Result, Any Laboratory		
	Degrees of Freedom, DF	s_a	Coefficient of Variation, %	Degrees of Freedom, DF	s_{a+b}	Coefficient of Variation, %
24.5	10	0.16	0.65	9	0.39	1.5
12.1	10	0.20	1.65	9	0.30	2.5
0.2	10	0.14	70	9	0.34	17

TABLE 15 Summary of Precision Estimates—Example B

Precision Estimates	Pertinent Standard Deviation	Degrees of Freedom, DF	95 % Range	
			Factor	$s \times$ Factor, % Absolute
Repeatability	0.22	60	2.8	0.6
Within-laboratory, between-days variability	0.17	30	2.8	0.5
Reproducibility	0.35	9	2.8	1.0

obtained by the same analyst on different days, has been estimated to be 0.17 % absolute at 30 DF. The 95 % limit for the difference between two such averages is 0.5 % absolute.

28.1.3 *Reproducibility (Multilaboratory)*—The standard deviation of results (each the average of duplicates), obtained by analysts in different laboratories, has been estimated to be 0.35 % absolute at 9 DF. The 95 % limit for the difference between two such averages is 1.0 % absolute.

NOTE 11—See 34.1 for the wording of this note.

29. Example (Using Data From Table 11 and Sections 25.2.7, 25.2.9.1, 25.2.9.2, and 25.2.9.3; Example A)

29.1 The following form and typical wording are recommended for the precision statements that appear in the Precision and Bias section of the test method:

30. Precision and Bias

30.1 *Precision*—The following criteria should be used to judge the acceptability of results (see Note 12):

30.1.1 *Repeatability (Single Analyst)*—The coefficient of variation for a single determination has been estimated to be 0.49 % relative at 44 DF. The 95 % limit for the difference between two such runs is 1.4 % relative, at the 250 to 300 level.

30.1.2 *Laboratory Precision (Within-Laboratory, Between-Days Variability, Formerly Called Repeatability)*—The coefficient of variation of results (each the average of duplicate determinations), obtained by the same analyst on different days, has been estimated to be 0.52 % relative at 38 DF. The 95 % limit for the difference between two such averages is 1.5 % relative.

30.1.3 *Reproducibility (Multilaboratory)*—The coefficient of variation of results (each the average of duplicate determinations), obtained by analysts in different laboratories, has been estimated to be 1.03 % relative at 9 DF. The 95 % limit for the difference between two such averages is 2.9 % relative.

NOTE 12—This note would be similar to Note 13 in 34.1.

PART G—BIAS (SYSTEMATIC ERROR)

31. Principle

31.1 In testing chemicals, the true or exact value is seldom known and appraisals of systematic error often are based on an expected value, such as a theoretical value calculated for a purified or standard sample. In other cases, the bias of a method is evaluated by comparing the determined average with the average obtained using a standard or referee method. Again, the recoveries of known amounts of the constituent in question from a prepared series of standards may be used for this purpose. The following are suggested ways of expressing the expected bias of analytical methods:

32. Examples

32.1 *Example No. 1*—Examples of expressing the expected bias referring to Test Method D 1013, are as follows:

The average value obtained in the analysis of a National Institute of Standards and Technology standard sample of acetanilide was 10.29 ± 0.04 %¹⁰ versus a theoretical nitrogen content of 10.36 %.

The average value obtained in the analysis of a purified melamine sample was 66.28 ± 0.11 % versus a theoretical nitrogen content of 66.67 %.

32.2 *Example No. 2*—An example referring to Test Method D 1727, is as follows:

The determined values for urea content averaged 0.2 % absolute higher than the expected values based on the total nitrogen content of the urea resin solution, as determined by Test Method D 1727. This was true for all three levels (0, 12, and 24 %) used in the interlaboratory test.

32.3 *Example No. 3*—An example referring to a hypothetical case is as follows:

Recoveries of known amounts of Constituent X in a series of prepared standards were as follows:

Amount Added, ppm	Recovery, percent relative
10.0	98
50.0	97
100.0	98

The limit of detectability was found to be 2 ppm.

PART H—PRESENTATION OF DATA

33. Experimental Data

33.1 When a method is submitted to a letter ballot for acceptance as an ASTM standard, the collaborative data used in determining its precision and bias should be sent to ASTM Headquarters. The precision and bias statement in the standard should have a footnote that informs the reader that the supporting data is on file in the Research Reports file at ASTM and that copies are available by request to ASTM. (For example, see Footnote 11.)

34. Statistical Data

34.1 Details of the statistical analysis should not be included in the draft, but should be referred to the Subcommittee on Precision and Bias when the method is submitted for editorial review. However, the draft of the method should contain a brief statement describing the interlaboratory study in sufficient detail so that the design will be apparent to anyone statistically interested. This can be done conveniently by adding a note to the section on Precision, as in the following example:

NOTE 13—These precision estimates are based on an interlaboratory study performed in 1967 on three samples, containing approximately 24, 12, and 0 % of Component X. One analyst in each of ten laboratories performed duplicate determinations and repeated one day later, for a total of 120 determinations.¹¹ Practice E 180 was used in developing these precision estimates.

¹⁰ The limits of uncertainty of the averages were calculated by the procedure given in the *ASTM Manual on Presentation of Data and Control Chart Analysis, STP 15D*, Part 2, p. 52. 1976.

¹¹ Supporting data are available from ASTM Headquarters. Request RR:E 15-1005.

REFERENCES

- (1) Finkner, Morris D., "The Reliability of Collaborative Testing for AOAC Methods," *Journal, Association of Official Agricultural Chemists*, Vol 40, 1957, pp. 882–892.
- (2) Youden, W. F., and Steiner, E. H., *Statistical Manual of the Association of Official Analytical Chemists*, P.O. Box 540, Benjamin Franklin Station, Washington, DC 20044.
- (3) Youden, W. J., "Graphic Diagnosis of Interlaboratory Test Results," *Industrial Quality Control*, Vol XV, No. 11, May 1959, pp. 24–28.
- (4) Murphy, R. B., "On the Meaning of Precision and Accuracy," *Materials Research & Standards*, ASTM, Vol 1, No. 4, April 1961, p. 264.
- (5) *ASTM Manual for Conducting an Interlaboratory Study of a Test Method, STP 335*; Out of Print. Available from University Microfilms, Inc., 300 N. Zeeb Rd., Ann Arbor, MI 48103.
- (6) Dixon, W. J., and Massey, F. J., *Introduction to Statistical Analysis*, 2nd Ed., McGraw-Hill Book Co., New York, NY, 1957.
- (7) Proschan, F., "Testing Suspected Observations," *Industrial Quality Control*, Vol XIII, No. 7, January 1957, pp. 14–19.
- (8) Grubbs, F. E., *Annals of Mathematical Statistics*, Vol 21, March 1950, pp. 27–58.
- (9) Davies, O. L., "Design and Analysis of Industrial Experiments," 2nd Ed., Hafner Publishing Co., 1956, p. 588.
- (10) Youden, W. J., "Statistical Methods for Chemists," John Wiley & Sons, New York, NY, 1951, pp. 74–79.
- (11) Barnett and Lewis, *Outliers in Statistical Data*, John Wiley & Sons, New York, NY, 1978.
- (12) Mandel, J., and Lashof, T.W., "The Nature of Repeatability and Reproducibility," *Journal of Quality Technology*, Vol 19, No. 1, January 1987.
- (13) Box, G. E. P., Hunter, W. G., and Hunter J. S., *Statistics for Experimenters*, John Wiley & Sons, New York, NY, 1978.
- (14) *ASTM Manual on Presentation of Data and Control Chart Analysis: 6th Edition*, ASTM Manual Series MNL 7, 1990, (Revision of *Special Technical Publication (STP) 15D*).

ASTM International takes no position respecting the validity of any patent rights asserted in connection with any item mentioned in this standard. Users of this standard are expressly advised that determination of the validity of any such patent rights, and the risk of infringement of such rights, are entirely their own responsibility.

This standard is subject to revision at any time by the responsible technical committee and must be reviewed every five years and if not revised, either reapproved or withdrawn. Your comments are invited either for revision of this standard or for additional standards and should be addressed to ASTM International Headquarters. Your comments will receive careful consideration at a meeting of the responsible technical committee, which you may attend. If you feel that your comments have not received a fair hearing you should make your views known to the ASTM Committee on Standards, at the address shown below.

This standard is copyrighted by ASTM International, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959, United States. Individual reprints (single or multiple copies) of this standard may be obtained by contacting ASTM at the above address or at 610-832-9585 (phone), 610-832-9555 (fax), or service@astm.org (e-mail); or through the ASTM website (www.astm.org).