Designation: E691 – 09

An American National Standard

Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method[1]

This standard is issued under the fixed designation E691; 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.

This standard has been approved for use by agencies of the Department of Defense.

INTRODUCTION

Tests performed on presumably identical materials in presumably identical circumstances do not, in general, yield identical results. This is attributed to unavoidable random errors inherent in every test procedure; the factors that may influence the outcome of a test cannot all be completely controlled. In the practical interpretation of test data, this inherent variability has to be taken into account. For instance, the difference between a test result and some specified value may be within that which can be expected due to unavoidable random errors, in which case a real deviation from the specified value has not been demonstrated. Similarly, the difference between test results from two batches of material will not indicate a fundamental quality difference if the difference is no more than can be attributed to inherent variability in the test procedure. Many different factors (apart from random variations between supposedly identical specimens) may contribute to the variability in application of a test method, including: *a* the operator, *b* equipment used, *c* calibration of the equipment, and *d* environment (temperature, humidity, air pollution, etc.). It is considered that changing laboratories changes each of the above factors. The variability between test results obtained by different operators or with different equipment will usually be greater than between test results obtained by a single operator using the same equipment. The variability between test results taken over a long period of time even by the same operator will usually be greater than that obtained over a short period of time because of the greater possibility of changes in each of the above factors, especially the environment.

The general term for expressing the closeness of test results to the“ true” value or the accepted reference value is accuracy. To be of practical value, standard procedures are required for determining the accuracy of a test method, both in terms of its bias and in terms of its precision. This practice provides a standard procedure for determining the precision of a test method. Precision, when evaluating test methods, is expressed in terms of two measurement concepts, repeatability and reproducibility. Under repeatability conditions the factors listed above are kept or remain reasonably constant and usually contribute only minimally to the variability. Under reproducibility conditions the factors are generally different (that is, they change from laboratory to laboratory) and usually contribute appreciably to the variability of test results. Thus, repeatability and reproducibility are two practical extremes of precision.

The repeatability measure, by excluding the factors *a* through *d* as contributing variables, is not intended as a mechanism for verifying the ability of a laboratory to maintain“ in-control” conditions for routine operational factors such as operator-to-operator and equipment differences or any effects of longer time intervals between test results. Such a control study is a separate issue for each laboratory to consider for itself, and is not a recommended part of an interlaboratory study.

The reproducibility measure (including the factors *a* through *d* as sources of variability) reflects what precision might be expected when random portions of a homogeneous sample are sent to random “in-control” laboratories.

To obtain reasonable estimates of repeatability and reproducibility precision, it is necessary in an interlaboratory study to guard against excessively sanitized data in the sense that only the uniquely best operators are involved or that a laboratory takes unusual steps to get“ good” results. It is also important to recognize and consider how to treat “poor” results that may have unacceptable assignable

causes (for example, departures from the prescribed procedure). The inclusion of such results in the final precision estimates might be questioned.

An essential aspect of collecting useful consistent data is careful planning and conduct of the study. Questions concerning the number of laboratories required for a successful study as well as the number of test results per laboratory affect the confidence in the precision statements resulting from the study. Other issues involve the number, range, and types of materials to be selected for the study, and the need for a well-written test method and careful instructions to the participating laboratories.

To evaluate the consistency of the data obtained in an interlaboratory study, two statistics may be used: the "*k*-value", used to examine the consistency of the within-laboratory precision from laboratory to laboratory, and the "*h*-value", used to examine the consistency of the test results from laboratory to laboratory. Graphical as well as tabular diagnostic tools help in these examinations.

[1] This practice is under the jurisdiction of ASTM Committee E11 on Quality and Statistics and is the direct responsibility of Subcommittee E11.20 on Test Method Evaluation and Quality Control.

Current edition approved Aug. 1, 2009. Published September 2009. Originally approved in 1979. Last previous edition approved in 2008 as E691 – 08. DOI: 10.1520/E0691-09.

1. Scope

1.1 This practice describes the techniques for planning, conducting, analyzing, and treating the results of an interlaboratory study (ILS) of a test method. The statistical techniques described in this practice provide adequate information for formulating the precision statement of a test method.

1.2 This practice does not concern itself with the development of test methods but rather with gathering the information needed for a test method precision statement after the development stage has been successfully completed. The data obtained in the interlaboratory study may indicate, however, that further effort is needed to improve the test method.

1.3 Since the primary purpose of this practice is the development of the information needed for a precision statement, the experimental design in this practice may not be optimum for evaluating materials, apparatus, or individual laboratories.

1.4 *Field of Application*—This practice is concerned exclusively with test methods which yield a single numerical figure as the test result, although the single figure may be the outcome of a calculation from a set of measurements.

1.4.1 This practice does not cover methods in which the measurement is a categorization, such as a go-no-go allocation (two categories) or a sorting scheme into two or more categories. For practical purposes, the discontinuous nature of measurements of these types may be ignored when a test result is defined as an average of several individual measurements. Then, this practice may be applicable, but caution is required and a statistician should be consulted.

1.5 The information in this practice is arranged as follows:

1.6 *This standard may involve hazardous materials, operations, and equipment. This standard does not purport to address all of the safety problems 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]

E177 Practice for Use of the Terms Precision and Bias in ASTM Test Methods

E456 Terminology Relating to Quality and Statistics

E1169 Practice for Conducting Ruggedness Tests

E2282 Guide for Defining the Test Result of a Test Method

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

3. Terminology

3.1 Terminology E456 provides a more extensive list of terms in E11 standards.

3.1.1 *accuracy*, *n*—the closeness of agreement between a test result and an accepted reference value. **E177**

3.1.2 *bias*, *n*—the difference between the expectation of the test results and an accepted reference value. **E177**

3.1.3 *interlaboratory study, (ILS) in ASTM* , *n*—a designed procedure for obtaining a precision statement for a test method, involving multiple laboratories, each generating replicate test results on one or more materials.

3.1.4 *precision*—the closeness of agreements between independent test results obtained under stipulated conditions. **E177**

3.1.5 *observation*—the process of obtaining information regarding the presence or absence of an attribute of a test specimen, or of making a reading on a characteristic or dimension of a test specimen. **E2282**

3.1.6 *repeatability*—precision under repeatability conditions. **E177**

3.1.7 *repeatability conditions*, *n*—conditions where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time. **E177**

3.1.8 *repeatability standard deviation, (S_r), n*—the standard deviation of test result obtained under repeatability conditions. **E177**

3.1.9 *reproducibility* , *n*—precision under reproducibility conditions. **E177**

3.1.10 *reproducibility conditions*, *n*—conditions where test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment. **E177**

3.1.11 *reproducibility standard deviation (S_R)*—the standard deviation of test results obtained under reproducibility conditions. **E177**

3.1.12 *test determination*, *n*—the value of a characteristic or dimension of a single test specimen derived from one or more observed values. **E2282**

3.1.13 *test method*, *n*—a definitive procedure that produces a test result. **E2282**

3.1.14 *test observation*, *n*—see **observation**.

3.1.15 *test result*, *n*—the value of a characteristic obtained by carrying out a specified test method. **E2282**

3.1.16 *test specimen*, *n*—the portion of a test unit needed to obtain a single test determination. **E2282**

3.1.17 *test unit*, *n*—the total quantity of material (containing one or more test specimens) needed to obtain a test result as specified in the test method. See test result. **E2282**

4. Summary of Practice

4.1 The procedure presented in this practice consists of three basic steps: planning the interlaboratory study, guiding the testing phase of the study, and analyzing the test result data. The analysis utilizes tabular, graphical, and statistical diagnostic tools for evaluating the consistency of the data so that unusual values may be detected and investigated, and also includes the calculation of the numerical measures of precision of the test method pertaining to both within-laboratory repeatability and between-laboratory reproducibility.

5. Significance and Use

5.1 ASTM regulations require precision statements in all test methods in terms of repeatability and reproducibility. This practice may be used in obtaining the needed information as simply as possible. This information may then be used to prepare a precision statement in accordance with Practice E177.

5.2 *Test Method and Protocol*—In this practice, the term "test method" is used both for the actual measurement process and for the written description of the process, while the term "protocol" is used for the directions given to the laboratories for conducting the ILS.

5.3 *Observations, Test Determinations and Test Results:*

5.3.1 A test method often has three distinct stages, the direct observation of dimensions or properties, the arithmetic combination of the observed values to obtain a test determination, and the arithmetic combination of a number of test determinations to obtain the test result of the test method. In the simplest of test methods a single direct observation is both the test determination and the test result. For example, the test method may require the measurement of the mass of a test specimen prepared in a prescribed way. Another test method may require the measurement of the area of the test specimen as well as the mass, and then direct that the mass be divided by the area to obtain the mass per unit area of the specimen. The whole process of measuring the mass and the area and calculating the mass per unit area is a test determination. If the test method specifies that only one test determination is to be made, then the test determination value is the test result of the test method. Some test methods require that several determinations be made and the values obtained be averaged or otherwise combined to obtain the test result of the test method. Averaging of several determinations is often used to reduce the effect of local variations of the property within the material.

5.3.2 In this practice, the term "test determination" is used both for the process and for the value obtained by the process, except when "test determination value" is needed for clarity.

5.3.3 The number of test determinations required for a test result should be specified in each individual test method. The number of test results required for an interlaboratory study of a test method is specified in the protocol of that study.

5.4 *Test Specimens and Test Units*—In this practice a test unit is the total quantity of material needed for obtaining a test result as specified by the test method. The portion of the test unit needed for obtaining a single test determination is called a test specimen. Usually a separate test specimen is required for each test determination.

5.5 *Precision, Bias, and Accuracy of a Test Method*:

5.5.1 When a test method is applied to a large number of portions of a material, that are as nearly alike as possible, the test results obtained nevertheless will not all have the same value. A measure of the degree of agreement among these test results describes the precision of the test method for that material.

5.5.2 Numerical measures of the variability between such test results provide inverse measures of the precision of the test method. Greater variability implies smaller (that is, poorer) precision and larger imprecision.

5.5.3 This practice is designed only to estimate the precision of a test method. However, when accepted reference values are available for the property levels, the test result data obtained according to this practice may be used in estimating the bias of the test method. For a discussion of bias estimation and the relationships between precision, bias, and accuracy, see Practice E177.

5.6 *Repeatability and Reproducibility*—These terms deal with the variability of test results obtained under specified laboratory conditions. Repeatability concerns the variability between independent test results obtained within a single laboratory in the shortest practical period of time by a single operator with a specific set of test apparatus using test specimens (or test units) taken at random from a single quantity of homogeneous material obtained or prepared for the ILS. Reproducibility deals with the variability between single test results obtained in different laboratories, each of which has applied the test method to test specimens (or test units) taken at random from a single quantity of homogeneous material obtained or prepared for the ILS.

5.6.1 *Repeatability Conditions*—The within-laboratory conditions specified above for repeatability. The single-operator, single-set-of-apparatus requirement means that for a particular step in the measurement process the same combination of operator and apparatus is used for every test result and on every material. Thus, one operator may prepare the test specimens, a second measure the dimensions and a third measure the breaking force. "Shortest practical period of time" means that the test results, at least for one material, are obtained in a time not less than in normal testing and not so long as to permit significant changes in test material, equipment or environment.

PLANNING THE INTERLABORATORY STUDY (ILS)

6. ILS Membership

6.1 *Task Group*[3]—Either the task group that developed the test method, or a special task group appointed for the purpose, must have overall responsibility for the ILS, including funding where appropriate, staffing, the design of the ILS, and decision-making with regard to questionable data. The task group should decide on the number of laboratories, materials, and test results for each material. In addition, it should specify any special calibration procedures and the repeatability conditions to be specified in the protocol (see 12.3 and 12.4).

6.2 *ILS Coordinator*—The task group must appoint one individual to act as overall coordinator for conducting the ILS. The coordinator will supervise the distribution of materials and protocols to the laboratories and receive the test result reports from the laboratories. Scanning the reports for gross errors and checking with the laboratories, when such errors are found, will also be the responsibility of the coordinator. The coordinator may wish to consult with the statistician in questionable cases.

6.3 *Statistician*:

6.3.1 The test method task group should obtain the assistance of a person familiar with the statistical procedures in this practice and with the materials being tested in order to ensure that the requirements outlined in this practice are met in an efficient and effective manner. This person should also assist the task group in interpreting the results of the data analysis.

6.3.2 When a person having adequate knowledge of both the materials and the proper statistical techniques is not available, the task group should obtain the services of a statistician who has experience in practical work with data from materials testing. The task group should provide the statistician with an opportunity to become familiar with the statistical procedures of this practice and with both the materials and the test method involved. The statistician should become a member of the task group conducting the ILS, (task group members need not be members of ASTM).

6.3.3 The calculations of the statistics (see Section 15) for each material can be readily done by persons not having statistical knowledge. (see 15.1.3 and 15.4.2.)

6.4 *Data Analyst*—This individual should be someone who is careful in making calculations and can follow the directions in Sections 15 through 17.

6.5 *Laboratory ILS Supervisor*—Each laboratory must have an ILS supervisor to oversee the conduct of the ILS within the laboratory and to communicate with the ILS Coordinator. The name of the supervisor should be obtained on the response form to the "invitation to participate" (see 9.4).

7. Basic Design

7.1 Keep the design as simple as possible in order to obtain estimates of within- and between-laboratory variability that are free of secondary effects. The basic design is represented by a two-way classification table in which the rows represent the laboratories, the columns represent the materials, and each cell (that is, the intersection of a row with a column) contains the test results made by a particular laboratory on a particular material (see Table 1).

8. Test Method

8.1 Of prime importance is the existence of a valid, well-written test method that has been developed in one or more competent laboratories, and has been subjected to a ruggedness test prior to the ILS.

8.2 A ruggedness test is a screening procedure for investigating the effects of variations in environmental or other conditions in order to determine how control of such test conditions should be specified in the written description of the method. For example, the temperature of the laboratory or of a heating device used in the test may have an effect that cannot be ignored in some cases but may be much less in others. In a ruggedness test, deliberate variations in temperature would be introduced to establish the allowable limits on control of temperature. This subject is discussed more fully in Refs (1,2 and 3) see also Guide E1169.

[3] To facilitate the preparation of the final report on the ILS, the task group can obtain the Research Report format guide from ASTM Headquarters.

TABLE 1 Glucose in Serum ILS Test Result Data

Laboratory	Material				
	A	B	C	D	E
1	41.03	78.28	132.66	193.71	292.78
	41.45	78.18	133.83	193.59	294.09
	41.37	78.49	133.10	193.65	292.89
2	41.17	77.78	132.92	190.88	292.27
	42.00	80.38	136.90	200.14	309.40
	41.15	79.54	136.40	194.30	295.08
3	41.01	79.18	132.61	192.71	295.53
	40.68	79.72	135.80	193.28	290.14
	42.66	80.81	135.36	190.28	292.34
4	39.37	84.08	138.50	195.85	295.19
	42.37	78.60	148.30	196.36	295.44
	42.63	81.92	135.69	199.43	296.83
5	41.88	78.16	131.90	192.59	293.93
	41.19	79.58	134.14	191.44	292.48
	41.32	78.33	133.76	195.12	294.28
6	43.28	78.66	137.21	195.34	297.74
	40.50	79.27	135.14	198.26	296.80
	42.28	81.75	137.50	198.13	290.33
7	41.08	79.76	130.97	194.66	287.29
	41.27	81.45	131.59	191.99	293.76
	39.02	77.35	134.92	187.13	289.36
8	43.36	80.44	135.46	197.56	298.46
	42.65	80.80	135.14	195.99	295.28
	41.72	79.80	133.63	200.82	296.12

8.3 As a result of carrying out the screening procedure, and of some experience with the test method in the sponsoring laboratory and one or two other laboratories, a written version of the test method must have been developed (but not necessarily published as a standard method). This draft should describe the test procedure in terms that can be easily followed in any properly equipped laboratory by competent personnel with knowledge of the materials and the property to be tested. The test conditions that affect the test results appreciably should have been identified and the proper degree of control of the test conditions specified in the description of the test procedure. In addition, the test method should specify how closely (that is, to how many digits) each observation in the test method is to be measured.

8.4 The test method should specify the calibration procedure and the frequency of calibration.

9. Laboratories

9.1 *Number of Laboratories*:

9.1.1 An ILS should include 30 or more laboratories but this may not be practical and some ILS have been run with fewer. It is important, that enough laboratories be included in the ILS to be a reasonable cross-section of the population of qualified laboratories; that the loss or poor performance of a few will not be fatal to the study, and to provide a reasonably satisfactory estimate of the reproducibility.

9.1.2 **Under no circumstances should the final statement of precision of a test method be based on acceptable test results for each material from fewer than 6 laboratories.** This would require that the ILS begin with 8 or more laboratories in order to allow for attrition.

9.1.3 The examples given in this practice include only 8 and 7 laboratories, respectively. These numbers are smaller than ordinarily considered acceptable, but they are convenient for illustrating the calculations and treatment of the data.

9.2 Any laboratory considered qualified to run the test routinely (including laboratories that may not be members of ASTM) should be encouraged to participate in the ILS, if the preparatory work is not excessive and enough suitably homogeneous material is available. In order to obtain an adequate number of participating laboratories, advertise the proposed ILS in where appropriate (for example, trade magazines, meetings, circulars, etc.).

9.3 "Qualified" implies proper laboratory facilities and testing equipment, competent operators, familiarity with the test method, a reputation for reliable testing work, and sufficient time and interest to do a good job. If a laboratory meets all the other requirements, but has had insufficient experience with the test method, the operator in that laboratory should be given an opportunity to familiarize himself with the test method and practice its application before the ILS starts. For example, this experience can be obtained by a pilot run (see Section 13) using one or two trial samples provided by the task group and returning the raw data and the test results to the task group. **The importance of this familiarization step cannot be overemphasized.** Many interlaboratory studies have turned out to be essentially worthless due to lack of familiarization.

9.4 Obtain written ensurance from each potential participating laboratory that it is properly equipped to follow all the details of the procedure and is willing to assign the work to a skilled operator in a timely manner. The decision of a laboratory to participate should be recorded on a response form to a written invitation. The invitation should include information covering the required time for calibrating the apparatus and for testing all of the materials, and other possible costs. The response form should include the name, address, and telephone number of the person supervising the ILS work within the laboratory, the address and other markings required to ensure the ILS sample material will be promptly delivered to the ILS supervisor, answers to brief questions concerning equipment, environment, and personnel, including previous use of the test method, upon which the apparent competence of the laboratory may be judged, and an affirmation that the laboratory understands what is involved and agrees to carry out its responsibilities with diligence.

9.5 The ILS should not be restricted to a group of laboratories judged to be exceptionally qualified and equipped for the ILS. Precision estimates for inclusion in a test method should be obtained through the efforts of qualified laboratories and personnel operating under conditions that will prevail when the test method is used in practice.

10. Materials

10.1 Material designates anything with a property that can be measured. Different materials having the same property may be expected to have different property levels, meaning higher or lower values of the property. Different dilutions of the same material or compound to be assayed are considered" different materials" for the purpose of this practice. The terminology" different levels of material" may be used, if appropriate.

10.2 The number and type of materials to be included in an ILS will depend on the range of the levels in the class of materials to be tested and likely relation of precision to level over that range, the number of different types of materials to which the test method is to be applied, the difficulty and expense involved in obtaining, processing, and distributing samples, the difficulty of, length of time required for, and expense of performing the test, the commercial or legal need for obtaining a reliable and comprehensive estimate of precision, and the uncertainty of prior information on any of these points.

10.2.1 For example, if it is already known that the precision is either relatively constant or proportional to the average level over the range of values of interest, a smaller number of materials will be needed than if it is merely known that the precision is different at different levels. The ruggedness test (see 8.2) and the preliminary pilot program (see Section 13) help to settle some of these questions, and may often result in the saving of considerable time and expense in the full ILS.

10.2.2 An ILS of a test method should include at least three materials representing different test levels, and for development of broadly applicable precision statements, six or more materials should be included in the study.

10.2.3 The materials involved in any one ILS should differ primarily only in the level of the property measured by the test method. When it is known, or suspected, that different classes of materials will exhibit different levels of precision when tested by the test method, consideration should be given to conducting separate interlaboratory studies for each class of material.

10.3 Each material in an ILS should be made to be or selected to be as homogeneous as possible prior to its subdivision into test units or test specimens. If the randomization and distribution of individual test specimens (rather than test units) does not conflict with the procedure for preparing the sample for test, as specified in the test method, greater homogeneity between test units can be achieved by randomizing test specimens. Then each test unit would be composed of the required number of randomized test specimens. (See Section 11 and 14.1 for the quantity of each material needed, its preparation and distribution.)

NOTE 1—It may be convenient to use established reference materials, since their homogeneity has been demonstrated.

11. Number of Test Results per Material

11.1 In the design of an ILS a sufficient total number of test results on each material must be specified to obtain a good estimate of the measure of repeatability, generally the repeatability standard deviation. In many cases, the standard deviation in question will be a function of the property level being measured. When this occurs, the standard deviation should be determined separately for each level. It is generally sound to limit the number of test results on each material in each laboratory to a small number, such as three or four. The minimum number of test results per laboratory will normally be three for a chemical test and three or four for a physical or optical test. The number may be as small as two when there is little danger that a test unit will be lost or questionable test results obtained, or as many as ten when test results are apt to vary considerably. Generally, the time and effort invested in an ILS is better spent on examining more materials across more laboratories than on recording a large number of test results per material within a few laboratories.

12. Protocol

12.1 In the protocol, cite the name, address, and telephone number of the person who has been designated ILS coordinator (see 6.2). Urge the laboratories to call the coordinator when any questions arise as to the conduct of the ILS.

12.2 Clearly identify the specific version of the test method being studied. If the test method allows several options in apparatus or procedure, the protocol should specify which option or options have been selected for the ILS. Test units and test data sheets must be provided for each option.

12.3 When special calibration procedures are required before every determination or every test result, they should be described specifically in the test method. If the test method specifies calibration only daily or less frequently, the ILS task group must decide whether to require recalibration before obtaining each test result. While doing so will eliminate calibration drift and help ensure relative independence of the test results, changes in calibration may increase the variability between test results.

12.4 Describe any special circumstances that must be addressed in implementing the repeatability conditions, such as the period of time between obtaining the test results for the same material; that is, not less than in normal testing and not so long as to likely permit significant changes in test material, equipment or environment.

12.5 Specify the required care, handling, and conditioning of the *materials* to be tested. Explain the coding system used in identifying the materials and the distinction between test units and test specimens, where appropriate.

12.6 Supply data sheets for each material for recording the raw data as observations are made. Give instructions on the number of significant digits to be recorded, usually one more, if possible, than required by the test method. Also, supply test result sheets on which test results can be calculated and reported. In many instances this can be combined with the raw data sheet. Specify the number of significant digits to be reported, usually two more than required by the test method. Request the laboratories send raw data and test result sheets as soon as the testing is completed, and at least weekly if testing will continue over several weeks.

12.7 Request that each laboratory keep a record (or log) of any special events that arise during any phase of the testing. This record, to be sent to the ILS coordinator, will provide a valuable source of information both in dealing with unusual data and in making improvements in the test method in future revisions.

12.7.1 Instruct the laboratories to notify the ILS coordinator promptly whenever an error in test procedure arises, so that a decision can be made as to whether a new set of test units should be sent to the laboratory for a complete retest of the material.

12.8 Enclose with the protocol a questionnaire requesting information on specific aspects of the apparatus, reagents,

calibration, or procedure, as well as any other information that might assist in dealing with data inconsistencies, or ensure the task group that the laboratory complied with the current requirements of the test method. Also obtain any other information that may be needed in preparing the final research report on the ILS (see Footnote [4]).

CONDUCTING THE TESTING PHASE OF THE ILS

13. Pilot Run

13.1 Before investing laboratory time in the full scale ILS, it is usually wise to conduct a pilot run with only one, or perhaps two, material(s) to determine whether the test method as well as the protocol and all the ILS procedures are clear, and to serve as a familiarization procedure for those without sufficient experience with the method (see 9.3). The results of this pilot run also give the task group an indication of how well each laboratory will perform in terms of promptness and following the protocol. Laboratories with poor performance should be encouraged and helped to take corrective action.

13.2 All steps of the procedures described in this practice should be followed in detail to ensure that these directions are understood, and to disclose any weaknesses in the protocol or the test method.

14. Full Scale Run

14.1 *Material Preparation and Distribution*:

14.1.1 *Sample Preparation and Labelling*— Prepare enough of each material to supply 50 % more than needed by the number of laboratories committed to the ILS. Label each test unit or test specimen with a letter for the material and a sequential number. Thus, for ten laboratories and two test results for each laboratory the test units for material B would be numbered from B1 to B30, or, if five test specimens per test unit are required, the test specimens may be numbered B1 to B150.

14.1.2 *Randomization*—For each material independently, allocate the specified number of test units or test specimens to each laboratory, using a random number table, or a suitable computerized randomization based on random numbers. See Ref. (**4**) for a discussion of randomization.

14.1.3 *Shipping*—Ensure that the test units are packaged properly to arrive in the desired condition. When the material is sensitive to the conditions to which it is exposed (light, heat, humidity, etc), place special directions for opening the package on a label outside the package. Clearly indicate the name of the person who has been designated as ILS supervisor at the laboratory on the address of each package. Follow each laboratory's instructions for ensuring prompt delivery of the package.

14.1.4 *Follow-up*—Once the test units have been shipped, the ILS coordinator should call each laboratory ILS supervisor within a week to ten days to confirm that all test units have arrived safely. If the task group has decided to intermingle test units from different materials in the order of testing, the testing should not start until all the test units have arrived at the laboratory so they can be tested in the specified order.

14.1.5 *Replacement Sets of Test Units*— As the ILS progresses, a laboratory may discover that the test method was not used properly on some test units. The laboratory ILS supervisor should discuss this with the ILS coordinator, who may send a replacement set of test units, replace the misused test units, or do nothing, as may seem desirable.

14.2 *Checking Progress*—From time to time, at intervals appropriate to the magnitude of the ILS, the coordinator should call each ILS supervisor to ascertain how the testing is progressing. By comparing the progress of all laboratories, the coordinator can determine whether some laboratories are lagging considerably behind the others and so advise these laboratories.

14.3 *Data Inspection*—The completed data sheets should be examined by the coordinator immediately upon receipt in order to detect unusual values or other deficiencies that should be questioned. Replacement sets of test units or of specific test units may be sent when there is missing or obviously erroneous data. The task group can decide later whether or not the additional data should be used in the estimation of the precision of the test method.

CALCULATION AND DISPLAY OF STATISTICS

15. Calculation of the Statistics

15.1 *Overview*—The analysis and treatment of the ILS test results have three purposes, to determine whether the collected data are adequately consistent to form the basis for a test method precision statement, to investigate and act on any data considered to be inconsistent, and to obtain the precision statistics on which the precision statement can be based. The statistical analysis of the data for estimates of the precision statistics is simply a one-way analysis of variance (within- and between-laboratories) carried out separately for each level (material). Since such an analysis can be invalidated by the presence of severe outliers, it is necessary to first examine the consistency of the data. The following paragraphs show, in terms of a numerical example, how the entire program is carried out:

15.1.1 The calculations are illustrated with test results from an ILS in which the concentration of glucose in serum (see Table 1) was measured at five different concentration levels by eight laboratories. Each laboratory obtained three test results at each concentration level.

15.1.2 For extended calculations it is usually necessary to retain extra significant digits in order to ensure that statistically important information is not lost in calculation by rounding off too soon. As a general rule, retain at least two more digits in the averages than in the reported test results and at least three significant figures in the standard deviations.

15.1.3 While the calculations described in this section are arranged for use of a hand calculator, they also can be readily programmed for the computer. A spreadsheet can be easily adapted to these calculations. In addition, a PC software package for Windows 3.1x, Windows 95, or OS2 is available from ASTM Headquarters.

[4] Available from ASTM International Headquarters. Order Adjunct No. ADJE0691.

15.2 *Table of ILS Test Results*—The test results received from the laboratories are usually best arranged in rows and columns as in Table 1. Each column contains the data obtained from all laboratories for one material, and each row contains the data from one laboratory for all materials. The test results from one laboratory on one material constitute a cell. Thus, the cell for Laboratory 2 and Material C contains the test results 132.92, 136.90 and 136.40. This cell is called C2, by material and laboratory. It helps in the interpretation of the data to arrange the materials in increasing order of the measured values.

15.3 *Worksheets*—Generally, it facilitates the calculations to prepare a separate calculation worksheet for each material, using Table 2 as a model but making appropriate changes for different numbers of laboratories, and test results per material. Enter the test result data for one material (from one column of Table 1) on a worksheet. Also enter the results of the following calculations for that material on the same worksheet, as illustrated in Table 2. Work on only one material at a time.

15.4 *Cell Statistics*:

15.4.1 *Cell Average, $\bar{x}$*—Calculate the cell average for each laboratory using the following equation:

$$\bar{x} = \sum_{1}^{n} x/n \tag{1}$$

where:

$\bar{x}$ = the average of the test results in one cell,
x = the individual test results in one cell, and
n = the number of test results in one cell.

TABLE 2 [A]Interlaboratory Study Worksheet for Glucose in Serum Initial Preparation of Test Result Data for Material *A*

Laboratory Number	Test Results, x			$\bar{x}$	s	d	h	k
	1	2	3					
1	41.03	41.45	41.37	41.2833	0.2230	−0.2350	−0.39	0.21
2	41.17	42.00	41.15	41.4400	0.4851	−0.0783	−0.13	0.46
3	41.01	40.68	42.66	41.4500	1.0608	−0.0683	−0.11	1.00
4	39.37	42.37	42.63	41.4567	1.8118	−0.0616	−0.10	1.70
5	41.88	41.19	41.32	41.4633	0.3667	−0.0550	−0.09	0.34
6	43.28	40.50	42.28	42.0200	1.4081	0.5017	0.83	1.32
7	41.08	41.27	39.02	40.4567	1.2478	−1.0616	−1.75	1.17
8	43.36	42.65	41.72	42.5767	0.8225	1.0584	1.75	0.77

[A]Average of cell averages, $\bar{\bar{x}} = 41.5183$
Standard deviation of cell averages, $s_{\bar{x}} = 0.6061$
Repeatability standard deviation, $s_r = 1.0632$
Reproducibility standard deviation, $s_R = 1.0632$

where:

x = individual test result,
$\bar{x}$ = cell average = $\sum_{1}^{n} x/n$ where n = number of test results per cell = 3,
$\bar{\bar{x}}$ = average of cell averages = $\sum_{1}^{p} \bar{x}/p$ where p = number of laboratories = 8,
s = cell standard deviation = $\sqrt{\sum_{1}^{n} (x - \bar{x})^2/(n-1)}$
d = cell deviation = $\bar{x} - \bar{\bar{x}}$
$s_{\bar{x}}$ = standard deviation of cell averages = $\sqrt{\sum_{1}^{p} d_2/(p-1)}$
s_r = repeatability standard deviation = $\sqrt{\sum_{1}^{p} s^2/p}$
s_R = reproducibility standard deviation = larger of s_r and $\sqrt{(s_{\bar{x}})^2 + (s_r)^2(n-1)/n}$
h = $d/s_{\bar{x}}$, and
k = s/s_r.

Thus from Table 2 for Material A, Laboratory 1 (that is, for Cell A1):

$$\bar{x} = (41.03 + 41.45 + 41.37)/3 = 41.2833.$$

15.4.2 *Cell Standard Deviation, s*—Calculate the standard deviation of the test results in each cell using the following equation:

$$s = \sqrt{\sum_{1}^{n} (x - \bar{x})^2/(n-1)} \tag{2}$$

The symbols have the same meaning as for Eq 1. Thus for Cell A1:

$$s = \sqrt{(41.03 - 41.2833)^2 + (41.45 - 41.2833)^2 + (41.37 - 41.2833)^2]/(3-1)}$$
$$= 0.2230$$

While Eq 2 shows the underlying calculation of the cell standard deviation, inexpensive pocket calculators are available that calculate both the average and the standard deviation directly. Check to be sure the calculator uses $(n-1)$ as the divisor in Eq 2, not n, and has adequate precision of calculation.

15.5 *Intermediate Statistics*:

15.5.1 *Average of the Cell Averages, $\bar{\bar{x}}$*—Calculate the average of all the cell averages for the one material using Eq 3.

$$\bar{\bar{x}} = \sum_{1}^{p} \bar{x}/p \tag{3}$$

where:

$\bar{\bar{x}}$ = the average of the cell averages for one material,
$\bar{x}$ = the individual cell averages, and
p = the number of laboratories in the ILS.

Thus for material A:

$$\bar{\bar{x}} = (41.2833 + 41.4400 + 41.4500 + 41.4567 + 41.4633 + 42.0200 + 40.4567 + 42.5767)/8$$
$$= 41.5183$$

15.5.2 *Cell Deviation, d*—For each laboratory calculate the cell deviation by subtracting the cell average from the average of the cell averages using the following equation:

$$d = \bar{x} - \bar{\bar{x}} \tag{4}$$

Thus for cell A1:

$$d = 41.2833 - 41.5183 = -0.2350 \tag{5}$$

15.5.3 *Standard Deviation of the Cell Averages, $s_{\bar{x}}$*—Calculate this statistic using the following equation:

$$s_{\bar{x}} = \sqrt{\sum_{1}^{p} d^2/(p-1)} \tag{6}$$

Thus for material A:

$$s_{\bar{x}} = \sqrt{[(-0.2350)^2 + (-0.0783)^2 + (-0.0683)^2 + (-0.0616)^2 + (-0.0550)^2 + (0.5017)^2 + (-1.0616)^2 + (1.0584)^2]/(8-1)}$$
$$= 0.6061$$

15.6 *Precision Statistics*—While there are other precision statistics, introduced later in this practice, the fundamental

precision statistics of the ILS are the repeatability standard deviation and the reproducibility standard deviation. The other statistics are calculated from these standard deviations.

15.6.1 *Repeatability Standard Deviation, s_r*—Calculate this statistic using the following equation:

$$s_r = \sqrt{\sum_1^p s^2/p} \qquad (7)$$

where:

s_r = the repeatability standard deviation,
s = the cell standard deviation (p of them from Eq 2), and
p = the number of laboratories.

Thus for material A:

$$s_r = \sqrt{[(0.2230)^2 + (0.4851)^2 + (1.0608)^2 + (1.8118)^2 + (0.3667)^2 + (1.4081)^2 + (1.2478)^2 + (0.8225)^2]/8}$$
$$= 1.0632$$

15.6.2 *Reproducibility Standard Deviation, s_R*—Calculate a provisional value of this statistic using the following equation:

$$(s_R)^* = \sqrt{(s_{\bar{x}})^2 + (s_r)^2 (n-1)/n} \qquad (8)$$

where: $s_{\bar{x}}$ and s_r are obtained from Eq 6 and Eq 7. The symbol, * indicates provisional value, (for more information see A1.1.2).

Thus for Material A:

$$(s_R)^* = \sqrt{(0.6061)^2 + (1.0632)^2 (3-1)/3} = 1.0588$$

Enter the larger of the values obtained by the use of Eq 7 and Eq 8 as the final value of s_R to be used for precision statements. In this case, Eq 7 yields the larger value. Therefore, $s_R = 1.0632$.

15.7 *Consistency Statistics, h and k*:

15.7.1 For each cell, calculate a value of h using the following equation:

$$h = d/s_{\bar{x}} \qquad (9)$$

where:

h = the between-laboratory consistency statistic,
d = the cell deviation (i.e., the deviation of the cell average from the average of the cell averages, from 15.5.2), and
$s_{\bar{x}}$ = the standard deviation of the cell averages (from 15.5.3).

Thus for Cell A1:

$$h = -0.2350/0.6061 = -0.39$$

Retain two decimal places in the computed values of h.

15.7.2 For each cell, use the following equation to calculate a value of k.

$$k = s/s_r \qquad (10)$$

where:

k = the within-laboratory consistency statistic,
s = the cell standard deviation for one laboratory (from 15.4.2), and
s_r = the repeatability standard deviation of the material (from 15.6.1).

Thus for Cell A1:

$$k = 0.2230/1.0632 = 0.21$$

Retain two decimal places in the computed values of k.

15.8 *Other Materials*—Repeat the steps described in 15.4 through 15.7 for each material, entering the calculation results on separate worksheets.

16. Tabular and Graphical Display of Statistics

16.1 *Material Order*—It is often useful to arrange the worksheets in order of increasing values of $\bar{\bar{x}}$, the material averages. This order may facilitate interpretation.

16.2 *Tables*—From the Table 2 results for each material, prepare tables of h and k as shown in Table 3 and Table 4 for the glucose in serum example.

16.3 *Graphs*—Prepare bar graphs for h and k in two ways: materials grouped by laboratory as in Fig. 1 and Fig. 2, and laboratories grouped by material as shown in Fig. 3 and Fig. 4. Arrange the laboratories and materials within and between each grouping in the same order as used in Table 1. Thus the materials will be arranged in order of increasing x from left to right, and the laboratories in order of laboratory code number.

DATA CONSISTENCY

17. Flagging Inconsistent Results

17.1 *Critical Values of the Consistency Statistics*—Table 5 lists critical values of the h and k consistency statistics at the 0.5 % significance level. The critical values for h (first column) depend on the number of laboratories (p, second column) participating in the ILS and the critical values for k (columns headed 2 through 10) depend both on the number of laboratories (p) and on the number of replicate test results (n) per laboratory per material. The 0.5 % level was chosen based on the judgment and experience that the 1.0 % resulted in too many cells being flagged and the 0.1 % level in too few. For further discussion see Appendix X1.

17.1.1 Obtain from Table 5 the appropriate critical values. For the glucose in serum example, the respective critical h and k values are 2.15 and 2.06. In Table 3 and Table 4 circle those values that exceed the critical values and underline those values that approach the critical values. On each graph draw a horizontal line for each critical value: two for h, since there are both positive and negative values of h, and one for k, as shown in Figs. 1-4.

17.1.2 The h and k graphs and the marked tables give a picture of the overall character of the variability of the test method as well as singling out particular laboratories or cells that should be investigated.

TABLE 3 Glucose in Serum-h[A]

Laboratory	Material				
	A	B	C	D	E
1	−0.39	−1.36	−0.73	−0.41	−0.46
2	−0.13	−0.45	0.10	0.15	1.64
3	−0.11	0.22	−0.21	−1.01	−0.68
4	−0.10	1.85	2.14	0.96	0.49
5	−0.09	−0.99	−0.71	−0.64	−0.34
6	0.83	0.21	0.55	0.97	0.17
7	−1.75	−0.16	−1.00	−1.33	−1.62
8	1.75	0.67	−0.15	1.31	0.79

[A]Critical value = 2.15.

TABLE 4 Glucose in Serum-*k*[A]

Laboratory	Material				
	A	B	C	D	E
1	0.21	0.11	0.22	0.02	0.18
2	0.46	0.89	0.79	1.78	2.33
3	1.00	0.56	0.63	0.61	0.69
4	1.70	1.85	2.41	0.74	0.22
5	0.34	0.52	0.44	0.72	0.24
6	1.32	1.09	0.47	0.63	1.03
7	1.17	1.38	0.77	1.45	0.84
8	0.77	0.34	0.36	0.94	0.42

[A]Critical value = 2.06.

17.2 *Plots by Laboratory*—In order to evaluate the differences between laboratories, use the following guidelines.

17.2.1 *h Graph*—There are three general patterns in these plots. In one, all laboratories have both positive and negative *h* values among the materials. In the second, the individual laboratories tend to be either positive or negative for all materials and the number of negative laboratories equals the number of positive laboratories, more or less. Neither of these patterns is unusual or requires investigation, although they may tell something about the nature of the test method variability. In the third pattern, one laboratory, with all *h* values positive (or negative), is opposed to all the other laboratories, with substantially all the *h* values negative (or positive). Such a pattern calls for an investigation of that laboratory.

17.2.1.1 Another kind of pattern to look for occurs within one laboratory, in which the *h* values for low property levels are of one sign, and for high property levels are of the opposite sign. If the values are extreme, this behavior should be investigated.

17.2.2 *k Graph*—Here the primary pattern to look for is that of one laboratory having large *k* values (or very small *k* values) for all or most of the materials. High *k* values represent within-laboratory imprecision. Very small *k* values may indicate a very insensitive measurement scale or other measurement problem.

17.3 *Plots by Material*—When a plot by laboratory shows several *h* or *k* values near the critical value line, look at the corresponding plot by material to see how that laboratory differs from the rest for a given material. Often a vertical line that seems strong in the plot by laboratory, because of its relation to the lines for the other materials, will turn out to be reasonably consistent with the other laboratories for the same material. Contrarywise, the *h* or *k* value for the one laboratory may be revealed as strongly different from the values for the other laboratories in the plot by material. If so, this behavior should be investigated.

18. Investigation

18.1 *Clerical and Sampling Errors*—Examine the laboratory report for each flagged cell. Try to locate where each test result in the flagged cell begins to deviate from the others. Is it in the original observations? Are the data rounded prematurely? Are the calculations correct? Then, look for signs of mislabeling of test units such that the test result for one material was reported as belonging to another material. Check these errors with the laboratories: do not assume them to be so.

18.2 *Procedural Errors*:

18.2.1 Study the laboratory reports again looking for deviations from either the test method or the protocol. For instance, variations in the number of significant digits reported in the test results may be a sign of incorrect rounding, or that the equipment in one laboratory is different from the rest. Also, study the event log for special comments relating to the flagged cells.

19. Task Group Actions

19.1 *General*—If the investigation disclosed no clerical, sampling or procedural errors, the unusual data should be retained, and the precision statistics based on them should be published. If, on the other hand, a cause was found during the investigation, the task group has several options to consider. If the laboratory clearly and seriously deviated from the test method, the test results for that laboratory must be removed from the ILS calculations. However, despite the danger of the recalcitrant laboratory having prior knowledge, it may be appropriate to ask the laboratory to retest one or more materials following the correct procedure, and then include the new set of test results in the ILS calculations. Of course, if the data have changed, recalculation of the *h* and *k* values must be made and the data consistency examined again.

19.2 *Exception*—When a large number of laboratories have participated in the ILS and no cause for some unusual cell values have been found during the investigation, it may be appropriate to delete a cell from the study if all of the other laboratories are in substantial agreement. The number of laboratories that can be considered large enough to support deletion of data without an identified cause cannot be stated exactly. Any action which results in discarding more than five percent of the ILS data likely will lead to the presentation of precision data that the test method cannot deliver in routine application.

19.3 *Test Method Vagueness*—One of the important things to be on the alert for during a laboratory investigation is for vagueness in the test method standard that permits a wide range of interpretation leading to loss of precision. Particular elements to check are lack of measurement tolerances, diversity of apparatus and insufficient direction for operator technique. These problems can be the basis for a revision of the standard.

20. Examples of Interlaboratory Studies

20.1 *Glucose in Serum*—The ILS is described in 15.1.1.

20.1.1 *h Statistic*—The overall impression given by Fig. 1 and Fig. 3 and Table 3 is one of reasonable consistency for variation among laboratories. Only Laboratory 4 stands out with large values for Materials B and C. The graph for Material C, in Fig. 3, shows that Laboratory 4 is distinctly different from the other laboratories. The graph for Material B, however, does not single out Laboratory 4.

20.1.2 *k Statistic*—Laboratories 2 and 4 stand out in Fig. 2 and Table 4. The laboratory plot, in Fig. 2, indicates Laboratory 4 has three high values, but a look at the material plots (Fig. 4) for *A* and *B* suggests that Laboratory 4 is not out of line for these two materials. On the other hand, the plot for Material C shows Laboratory 4 is different. Similarly, the plot for Material E shows Laboratory 2 is different for this material.