Designation: D 5847 – 02

An American National Standard

Standard Practice for Writing Quality Control Specifications for Standard Test Methods for Water Analysis[1]

This standard is issued under the fixed designation D 5847; 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 provides specific, mandatory requirements for incorporating quality control (QC) procedures into all test methods under the jurisdiction of Committee D-19.

1.2 ASTM has adopted the following:

Policy on implementation of requirements for a quality control section in standard test methods generated by Committee D-19 on Water.

GENERAL—By July 29, 1998, or at the next reapproval or revision, whichever is later, every D-19 Standard Test Method shall contain a QC section that is in full compliance with the requirements of this practice.

NEW COLLABORATIVE TESTING—As of July 29, 1998, each collaborative study design shall include a QC section as part of the method to be tested. Prior to approval of the study design, the Results Advisor shall ascertain the appropriateness of the QC section in meeting the requirements of this Practice and Practice D-2777, and shall advise the designer of the study of any changes needed to fulfill the requirements of these practices. Before a collaborative study may be conducted, approval of the study design by the Results Advisor must be obtained.

OLDER VALIDATED METHODS—Standard test methods that were validated using D-2777-77, D-2777-86, or D-2777-94, when balloted for reapproval or revision, shall contain a QC section based upon the best information from the historical record. Where appropriate, information derived from the record of the collaborative study shall be utilized for this purpose. The introduction of the QC section into these standard test methods shall not be construed as a requirement for a new collaborative study, though the Subcommittee may opt for such a study. Any information available regarding QC or precision/bias testing shall be included in the appropriate sections of the published method.

1.3 Required QC sections in all applicable test methods are intended to achieve two goals. First, users of Committee D-19 test methods will be able to demonstrate a minimum competency in the performance of these test methods by comparison with collaborative study data. Second, all users of test methods will be required to perform a minimum level of QC as part of proper implementation of these test methods to ensure ongoing competency.

1.4 This practice contains the primary requirements for QC of a specific test method. In many cases, it may be desirable to implement additional QC requirements to assure the desired quality of data.

1.5 The specific requirements in this practice may not be applicable to all test methods. These requirements may vary depending on the type of test method used as well as the analyte being determined and the sample matrix being analyzed. See Explanation 1 in Appendix X1.

1.5.1 If there are compelling reasons why any of the specific QC requirements listed in this practice are not applicable to a specific test method, these reasons must be documented in the QC section of the test method.

1.5.2 With the approval of Committee D-19 on the recommendation of the D-19 Results Advisor and the Technical Operations section of the Executive Subcommittee, a statement giving the compelling reasons why compliance with all or specific points of this practice cannot be achieved will meet the requirements of both ASTM and this practice.

1.5.3 Test Methods developed prior to the approval of this practice with a QC Section that meet the requirements of Specification D 5789 are considered in compliance with this Practice.

1.6 This practice is for use with quantitative methods and may not be applicable to qualitative test methods.

1.7 Presently, this practice is applicable primarily to chemical test methods. It is intended that, in future revisions, the practice will be expanded to include other methods such as microbiological methods.

2. Referenced Documents

2.1 *ASTM Standards:*

D 1129 Terminology Relating to Water[2]

D 1193 Specification for Reagent Water[2]

D 2777 Practice for Determination of Precision and Bias of Applicable Methods of Committee D-19 on Water[2]

D 3648 Practices for the Measurement of Radioactivity[3]

[1] This practice is under the jurisdiction of ASTM Committee D19 on Water and is the direct responsibility of Subcommittee D19.02 on General Specifications, Technical Resources and Statistical Methods.

Current edition approved July 10, 2002. Published July 2002. Originally published as D 5847 – 99. Last previous edition D 5847 – 99a$^{\epsilon 1}$.

[2] *Annual Book of ASTM Standards,* Vol 11.01.

[3] *Annual Book of ASTM Standards,* Vol 11.02.

D 3856 Guide for Good Laboratory Practices in Laboratories Engaged in Sampling and Analysis of Water[2]

D 4375 Terminology for Basic Statistics in Committee D-19 on Water[2]

D 5789 Writing Quality Control Specifications for Standard Test Methods for Organic Constituents[2]

D 5810 Guide for Spiking Into Aqueous Samples[2]

3. Terminology

3.1 *Definitions:*

3.1.1 For definitions of terms used in this practice, refer to Terminology D 1129 and Terminology D 4375.

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *batch*—a set (group) of samples analyzed such that results of analysis of the QC samples (laboratory control sample, method blank, matrix spike, and duplicate or matrix spike duplicate) analyzed with the batch are indicative of the quality of the results of analysis of samples in the batch. The number of samples in the batch is defined by the task group responsible for the method. See 6.4 and Explanation 2 in Appendix X1.

3.2.1.1 *Discussion*—When results from tests of any of the QC samples associated with batch the fail to meet the performance criteria, the test method should define the appropriate corrective action. To make such a response valid, the batch must be constructed in such a way as to assure that all variables affecting the batch will affect all samples in the batch in a statistically equivalent manner.

3.2.2 *calibration standard*—a solution containing the analyte of interest at a known concentration either purchased from an external source or prepared in-house from materials of known purity or concentration, or both, and used to calibrate the measurement system.

3.2.3 *detection limit*—the minimum concentration or amount of a substance that can be detected with a known degree of confidence.

3.2.4 *independent reference material (IRM)*—a material of known purity and concentration obtained either from the National Institute of Standards and Technology (NIST) or other reputable supplier. The IRM shall be obtained from a different lot of material than is used for calibration.

3.2.5 *laboratory control sample (LCS)*—a sample of known concentration and composition that is taken through the entire test method to determine whether the analytical system is in control. The LCS must be prepared in the appropriate ASTM-grade water from a material that sufficiently challenges the test. See Explanation 3 in Appendix X1. The LCS can be an IRM obtained from an outside source or prepared in-house from materials of known purity and concentration. Alternatively, the LCS may be a real sample of the matrix that is typically analyzed and which has been fully characterized.

3.2.5.1 *Discussion*—The LCS may also be commonly known as a "quality control sample" or an "ongoing precision and recovery sample" (OPR).

3.2.6 *matrix spike (MS)*—addition of a known concentration of analyte to a routine sample representing a specific matrix for the purpose of evaluating interference from matrix components. (See Guide D 5810.)

3.2.7 *method blank (blank)*—reagent water (see Specification D 1193) either known to be free of the constituent(s) of interest or containing only a low, known concentration of the constituent(s) of interest not exceeding five times the estimated detection limit.

3.2.7.1 *Discussion*—The purpose of analysis of the method blank is to confirm that the reagents or analytical system, or both, do not contribute a measurable amount of the constituent(s) of interest during analysis of routine samples or, if they do, to determine what the contribution is.

3.2.8 *quantitation limit*—the minimum concentration or amount of a substance that can be measured with a known degree of confidence.

3.2.9 *sample pretreatment (pretreatment)*—any handling, manipulation or treatment of a sample prior to subjecting the sample to the analysis. Examples are filtration, digestion, dilution, pH adjustment and extraction.

4. Summary of Practice

4.1 This practice provides the writer of a test method in Committee D19 specific steps to be included in the QC section of the test method. A QC section is required in all applicable standard test methods that mandates use of the following QC measures:

4.1.1 Periodic calibration or verification of calibration of the measurement system,

4.1.2 Initial demonstration of laboratory capability,

4.1.3 Analysis of at least one blank per batch,

4.1.4 Analysis of at least one LCS per batch,

4.1.5 Analysis of at least one MS per batch, where applicable, and

4.1.6 Periodic analysis of an IRM.

4.2 Duplicate analysis of at least one sample per batch is suggested. The duplicate analysis may be of a sample or of a matrix spike (matrix spike duplicate; MSD). See Explanation 4 in Appendix X1.

4.3 If there are valid reasons why any of the above QC requirements are inapplicable to a specific test method (see Section 1.), these reasons must be documented in the QC section of the test method. See 1.5 and Explanation 1 in Appendix X1.

5. Significance and Use

5.1 In order to be certain that the end user of analytical results obtained from using an ASTM Committee D-19 test method can be confident that the values have been obtained through a competent application of the test method, a demonstration of the proficiency of the analytical system must be performed. Appropriate proficiency is demonstrated by achievement of performance criteria derived from results of the test method collaborative study. The QC measures specified in this practice must be included in each ASTM test method, as applicable, to ensure the quality of measurements.

5.2 In order for users of D-19 test methods to achieve consistently valid results, a minimum level of QC must be performed. This minimum level of QC is stipulated in this practice and by the taskgroups developing D-19 test methods.

If the specific requirements outlined in this practice are not applicable to the test method, alternative QC must be defined in the test method.

6. Requirements for QC Specifications in Test Methods

6.1 Every test method must have a quality control (QC) section. Listed below are requirements applicable to nearly all chemical test methods and that must be followed to ensure that the test method is in control and to validate the accuracy of data generated for a specific matrix.

6.1.1 The measures that must be specified in the QC section of test methods and the reasons for these measures are as follows:

6.1.1.1 Calibration and calibration verification are necessary to ensure that the analytical system is properly calibrated during the period that the analysis is performed.

6.1.1.2 An initial demonstration of laboratory capability is necessary to prevent errors as a result of unfamiliarity with the test.

6.1.1.3 Analysis of a blank with each batch may indicate that analytes in a test sample are the result of contamination.

6.1.1.4 An LCS is run with each batch to determine that the measurement system is in control at the time samples are being analyzed.

6.1.1.5 An MS (recovery check) provides information on the bias of the test method in a specific matrix.

6.1.1.6 A duplicate analysis (Dup) or duplicate of the MS (matrix spike duplicate; MSD) indicates the repeatability of the method for a specific matrix.

6.1.1.7 An IRM is analyzed periodically to validate the accuracy of the test system and standards used for calibration.

6.1.2 In addition to the QC measures required above, each test method should contain a detection limit and a quantitation limit so that there is an indication of the lowest level at which the substance(s) determined by the test method can be detected and measured.

6.1.3 Statistical tests should be done at a significance level of $\alpha \leq 0.01$, that is, ≥ 99 % confidence level. If other levels are specified, the reason for deviation should be delineated in the method.

6.1.4 The operational principles and characteristics of detectors used for radioactivity measurements are somewhat different from those of instruments used for measurements of chemical and physical properties. Therefore, authors of ASTM test methods for radioactivity measurements should provide specific guidance within each test method, practice or guide relative to applicable QC program requirements. Guidance on the preparation and use of instrument tolerance and control charts can be found in Practices D 3648 and D 3856, and in ASTM MNL 7.[4]

6.2 *Calibration and Calibration Verification*—For test methods requiring calibration of instrumentation, an appropriate number of calibration standards must be analyzed during day that an analysis is performed to confirm that the instrument is properly set up and required sensitivity is being obtained. The actual number of standards required will depend on the requirements of the test method. For tests run infrequently, analysis of a single calibration standard to verify an existing calibration curve may suffice. For tests run frequently, it may be necessary to intersperse verification standards with test samples. Under these circumstances, it is recommended that a different standard concentration be used each time calibration is verified. Raw data (absorbance, intensity, etc.) should be compared to data generated in the past under the same conditions and should fall within three standard deviations of the mean value found in the past based on the pooled single operator precision. Alternatively, data should be compared to the calibration limits stated in the test method or should be developed from collaborative study data. Refer to Guide D 3856 and Practice D 3648 for further information on calibration checks.

6.2.1 For titrimetric test methods, titrants must be standardized on a scheduled basis against a standard solution of known concentration in duplicate or triplicate. The average normality/molarity is then used for calculation. The frequency of standardization is left to the judgment of the writer of the test method and should be based on the stability of the titrant.

6.2.2 An alternate calibration procedure, such as an internal standard, external standard, or single-point calibration procedure, must be specified in the test method.

6.2.3 The test method must establish the frequency of calibration and calibration verification.

6.3 *Initial Demonstration of Laboratory Capability*—A test must be included in the test method to confirm that the laboratory is capable of running the test method and generating acceptable data. This test of laboratory capability will vary depending on the test method. Whenever appropriate, a precision and bias (as recovery) test is performed. For most test methods this can be done by analyzing at least seven replicates of a standard solution prepared from a reference material containing the analyte at one of the concentration levels used in the collaborative study. The matrix and chemistry of the solution should be such that, when spiked, results statistically equivalent to results produced in the collaborative study should be produced. Each of the replicates should be presented to the operator as unknowns and should be interspersed with other samples following the procedures used in the collaborative study. For some test methods, fewer replicates may be used, however, the statistical power of the test is dependent on the number of replicates, and the meaningfulness of the study is reduced when fewer than seven replicates are used. (For the examples in this practice, fewer than seven replicates are used for convenience.) Each replicate must be taken through the complete analytical test method including any pretreatment. The mean and standard deviation of these results are then calculated as described in Terminology D 4375 and compared to the single operator precision and recovery found in the collaborative study.

NOTE 1—*Initial Demonstration of Laboratory Capability*—The type of test designed to assess the capability of a laboratory or operator is at the discretion of the method writer. It can be designed any way the method writer believes is appropriate for the test method so long as it provides meaningful data to ensure that the laboratory or operator is capable of generating results that are valid and accurate within the confidence limits

[4] ASTM Manual on Presentation of Data and Control Chart Analysis, ASTM MNL 7.

defined in the precision and bias statement of the test method.

6.3.1 To establish that results produced by a laboratory will be acceptable, the test method writer must prepare a table containing a upper limit for acceptable precision and a range for acceptable recovery for the analytes determined by the test method. The limit for acceptable precision is established by carrying out a one-sided F test at the $\alpha = 0.01$ significance level, and the range for acceptable recovery is established by carrying out a two-sided Student's t test. Instructions for performing these calculations are provided in 6.3.1.1 and 6.3.1.2. An example is given as Example 1 in Appendix X2.

6.3.1.1 The single-sided F test for a limit on precision is carried out using the square of the standard deviation found by the operator, S_A, and the square of the expected pooled single operator standard deviation reported in the collaborative study, S_O, at the concentration level at which the precision study was carried out, and dividing the square of S_A by the square of S_O. The resulting value must be less than or equal to the F value at the 0.01 significance level (99 % confidence level) for the number of degrees of freedom in the operator's study and the number of degrees of freedom in the collaborative study. The following formula is used:
Eq 1:

$$\frac{(S_A)^2}{(S_O)^2} \leq F_{0.99} \text{ at } (df_{S_A}, df_{S_O}) \quad (1)$$

where:
S_A = standard deviation found by operator,
S_O = single operator standard deviation reported in collaborative study,
$F_{0.99}$ = F value at 99 % confidence level,
df_{S_A} = degrees of freedom in laboratory's study (usually 6 because 7 replicates are usually run), and
df_{S_O} = Degrees of freedom for the single operator standard deviation estimate from the collaborative study

If $S_A < S_O$, S_A/S_O is inverted to S_O/S_A in Formula 1. See Example 1 in Appendix X2.

6.3.1.2 The two-sided Student's t test for a recovery range is carried out using Eq 2:

$$\left| \frac{\bar{X}_A - \bar{X}}{\sqrt{(S_T)^2 - \frac{(n-1)(S_O)^2}{n}}} \right| \leq t_{0.99} \text{ at } df \quad (2)$$

where:
$\bar{X}_A$ = mean value found by laboratory,
$\bar{X}$ = mean value found in collaborative study,
S_T = overall standard deviation found in collaborative study, and
S_O = single operator standard deviation found in collaborative study.

NOTE 2—If $S_O > S_T$ from the collaborative study, let $S_O = S_T$:

n = number of replicates used in laboratory's precision study (usually 7),
$t_{0.99}$ = student's t value at 99 % confidence level and
df = degrees of freedom for the overall standard deviation estimate from the collaborative study (one less than the number of laboratories that provided usable data at the concentration being tested.)

See Example 1 in Appendix X2.

6.3.2 The test method shall contain the requirement that the initial demonstration must be repeated until the results fall within these criteria.

6.4 *Batch QC*—The QC for routine operation is governed by a batch. A batch consists of a set of samples accompanied by QC samples. The QC samples are an LCS, blank, MS, and optionally, a Dup or MSD. The result obtained for the QC samples that accompany each batch must meet performance criteria developed from collaborative study data using the procedures in this practice or such as those found in Practice D 5789. The control limits are included in each test method. The taskgroup must specify in the test method the consequence of a result for a QC sample that fails to meet a performance criterion.

6.4.1 The size and frequency of the batch is determined by identifying the key variables affecting the batch and selecting a batch size and frequency so that these variables do not vary - are controlled - during analysis of the batch. The taskgroup may specify any batch size or frequency, or both, so long as the results of analysis of the LCS, blank, MS, and Dup or MSD can be assured to be indicative of the variables affecting the remaining samples in the batch; that is, all samples in the batch are subject only to the same set of random variables. If the risk or consequence of failure of a QC sample is high, the batch size should be small; if the risk is low, the batch size may be large. The Taskgroup must establish a maximum time between QC samples or the maximum number of samples in the batch, or both, or instruct the method user of the risk. See Explanation 2 in Appendix X1.

6.4.2 *Method Blank (Blank)*—Each test method shall require that, where applicable, a blank must be analyzed with each batch, as appropriate to the method. The blank is taken through all the steps of the test method including any preservation and pretreatment that may be necessary for samples. The value found for the blank should be below the detection limit of the test method or significantly below the confidence limits of the known concentration of the analyte in the associated test sample.

6.4.3 *Laboratory Control Sample (LCS)*—Each test method shall require that, where applicable, an LCS must be run with each batch, preferably at both the beginning and end of the batch, to determine if the measurement system is in control.

6.4.3.1 The LCS must be prepared in the appropriate ASTM-grade water from a material that sufficiently challenges the test method (see Explanation 3 in Appendix X1). The LCS must be taken through all steps of the test method. The concentration of the LCS must be known within a specified range of error. It is recommended that an independent reference material be used as the LCS, where possible.

6.4.3.2 Selecting an analyte concentration for the LCS other than the one employed in the collaborative study will require, for purposes of comparison, using a mean and standard deviation obtained from the collaborative test regression expressions at the selected true concentration. In this instance, a

procedure different from that in Example 1 in Appendix X2 must be used to determine the degrees of freedom for the Student's t value for the two-sided test.

6.4.4 *Matrix Spike (MS)*—The MS tests the bias of the test method in the matrix being analyzed. A portion of at least one sample from each batch is spiked with a known concentration of the analyte and the sample is taken through the test method including any sample pretreatment that may be required. Guidance on spiking can be found in Guide D 5810. The concentration of the analyte in the spiked sample should be at least double, but not over five times, the concentration of the analyte in the unspiked sample. For multi-analyte methods, such as gas chromatography (GC) or inductively coupled plasma (ICP) methods, it may be complicated to spike all analytes at a concentration in the range of 2 to 5 times the concentration of the analytes in the unspiked sample. For this condition, the analytes may be spiked at a fixed concentration or groups of analytes may be spiked at a few concentrations. The spike concentration plus the concentration found in the unspiked sample must fall within the demonstrated working range for the test method.

6.4.4.1 Selecting an analyte concentration for the MS other than the one employed in the collaborative study will require, for purposes of comparison, using a mean and standard deviation obtained from the collaborative test regression expressions at the selected true concentration. In this instance, a procedure different from that in Example 1 in Appendix X2 must be used to determine the degrees of freedom for the Student's t value for the two-sided test.

6.4.4.2 Two choices are available for development of performance criteria for MS recovery when multiple matrices are evaluated: (1) develop overall performance criteria by pooling data across all matrices. These criteria will reflect the performance of the test method across all matrices but will be broader than criteria developed for a specific sample matrix; (2) develop performance criteria for each matrix and include a table of matrices and their respective performance criteria in the test method. Use the test data from each matrix to develop the performance criteria for that matrix.

6.4.4.3 If, after the test method is balloted and approved, the test method will be applied to a matrix considerably different from those used to create the performance criteria included in the test method, it may be appropriate for the Taskgroup to develop additional performance criteria and add these criteria to the test method. Also, if the test method will be applied to a matrix considerably different from that used in the collaborative study, the Taskgroup may stipulate in the test method that the method user may develop performance criteria as specified in Practice D 3856. In this event, the taskgroup must also stipulate that if the performance criteria developed by the method user are less stringent than those specified in the test method, the client or data user must be informed that less stringent performance criteria are being used. See Explanation 5 in Appendix X1.

6.4.4.4 The following procedure is used for development of performance criteria for recovery. An example is given as Example 2 in Appendix X2.

6.4.4.5 Include a test for percent recovery (P) of the spike using Eq 3:

$$P = 100\,\frac{|A(V_s + V) - BV_s|}{CV} \tag{3}$$

where:

A = estimated concentration obtained from analysis of the spiked sample,
B = estimated concentration obtained from analysis of the unspiked sample,
C = known concentration of analyte in the spiking solution,
V_s = volume of sample used, and
V = volume of spiking solution added.

Because both A and B are experimentally determined, the mean percent spike recovery (P) must be estimated as follows:

$$\bar{P} = (100/CV)(\bar{x}_T(V_s+V)), \tag{4}$$

where:

$\bar{x}_T$ = the expected mean of analytical results at concentration T, when $T = CV/(V_s+V)$

and the standard deviation of such percent spike recoveries (s_P) is estimated as follows:

$$s_P = (100/CV)(s_A{}^2(V_s+V)^2+s_B{}^2(V_s)^2)^{1/2} \tag{5}$$

where:

s_A = the expected standard deviation of analytical results at measured concentration A, and
s_B = the expected standard deviation of analytical results at measured concentration B.

6.4.4.6 A specific P value is acceptable if it is in the following interval developed from the collaborative test:

$$(\bar{P} - 3(s_P)) \le P \le (\bar{P}+3(s_P)) \tag{6}$$

If P does not fall within these limits, a matrix interference may be present in the sample selected for spiking. Under these circumstances, the test method should state one or more of the following corrective actions: the selected sample and all samples in the batch should be reanalyzed, the selected sample and all samples in the batch should be analyzed by a test method not affected by the matrix interference, the matrix interference should be removed, or the related analytical results must be qualified an indication that they do not fall within the performance criteria of the test method or that a matrix effect exists for the affected samples.

6.5 *Duplicate (Dup)*—As an ongoing check on the precision of the analyses of samples, the method writer should include the requirement that a sample be analyzed in duplicate with each batch. If the sample contains the analyte at a level greater than five times the detection limit of the method, the sample and dup may be analyzed unspiked; otherwise, an MSD should be used.

6.5.1 Two choices are available for development of a performance criterion for precision of the dup (or MSD) when multiple matrices are evaluated. The choice must be consistent for development of performance criteria for recovery of the MS (6.4.4).

6.5.1.1 Develop an overall performance criterion by pooling data across all matrices. This criterion will reflect the performance of the test method across all matrices but will likely be broader than criteria developed for a specific sample matrix.

6.5.1.2 Develop a separate performance criterion for each matrix and include a table of matrices and their respective performance criteria in the test method. Use the test data from each matrix to develop the performance criterion for that matrix.

6.5.2 An appropriate statistical test such as an *F* test at the $\alpha = 0.01$ significance level (99 % confidence level) shall be applied to compare the precision of the sample analyses with the single operator precision in a collaborative study for similar concentrations. This is done to determine whether the precision of routine analyses is satisfactory. Refer to Example 3 in Appendix X2 for information on carrying out the F test. In order to properly carry out this comparison, the concentration of the routine sample selected must be within the concentration range studied in the collaborative study. As sufficient data are accumulated from the duplicate analyses performed by the laboratory, a relationship between single operator precision and concentration within the laboratory could be developed and used instead of the precision found in the collaborative study whenever the laboratory's precision is better. Refer to Guide D 3856 for more information on determining the acceptability of accumulated duplicate results. See Explanation 5 in Appendix X2.

6.6 Independent reference material (IRM) - In order to verify the quantitative value of the laboratory's calibration standards, each test method shall contain a requirement for periodic analysis of an IRM (if available) submitted as a regular sample (when practical) to the laboratory. This may be a standard reference material (SRM) from NIST, a reference material from a government agency, or a reputable commercial source. Results from analysis of the IRM must be within the control limits specified by the outside source or those used to evaluate the laboratory's routine calibration checks. Refer to Guide D 3856 for further information on calibration checks.

7. Approval

7.1 For a test method that is required to contain a QC section, the final QC section to appear in the method, along with documentation of all related calculations, must be reviewed and approved by the D-19 Results Advisor before it can appear on a committee ballot.

7.2 When an interlaboratory study has been conducted, the final QC section and all related calculations are submitted for approval at the same time as the precision and bias statement.

7.3 After approval, the D-19 Results Advisor shall send all materials submitted to him to ASTM for filing.

8. Keywords

8.1 bias; precision; quality control

APPENDIXES

(Nonmandatory Information)

X1. Explanations

X1.1 *Explanation 1—Reasons for Inapplicability of This Practice*:

X1.1.1 If the laboratory participates in a quality assurance/quality control program that includes the extensive laboratory auditing and performance evaluation that occurs with some radiochemistry programs, the QC requirements listed in this practice may not be necessary.

X1.1.2 Portions or all of the QC required by this practice may be inapplicable to certain test methods. For example, the MS is not applicable to pH because the buffering capacity of a sample cannot be determined readily. Therefore, a test method for determination of pH would not be required to contain a requirement or a performance specification for recovery of the MS.

X1.1.3 Performance criteria for a test method that have been developed using Practice D 5789.

X1.2 *Explanation 2—Batch Size and Frequency of QC Samples*:

X1.2.1 The batch size and frequency of QC samples will be dependent on the number and frequency of analysis of test samples. For example, if samples are analyzed monthly and there are ten samples in the batch, ten months would elapse between analysis of QC samples. If any QC sample fails (LCS, Blank, MS, Dup or MSD), results from analysis of all samples during the ten-month period would be suspect. Because it would likely be impossible to reanalyze many of the intervening samples within the holding time or reporting period, recovery from the QC failure would not be possible.

X1.2.2 Because the consequence of failure of a QC sample at the end of a batch may be severe, great pressure could be brought to bear on the analyst. History has shown that, under this pressure, some analysts have manipulated QC results to meet performance criteria. Because increasing the batch size or the time between calibration verifications or QC sample batches increases the financial loss that will occur if the QC is failed and the batch must be reanalyzed, the Taskgroup should weigh the economic and legal consequences as a component of the decision on the appropriate batch size and the frequency of QC samples.

X1.2.3 If the test method or practice will be used for reporting results to a regulatory authority for permitting or regulatory compliance purposes, the task group should consider batch size and frequency requirements that will satisfy the regulatory authority. For example, EPA has established a batch

size of 10 or 20 samples and a frequency in the range to 8 to 12 h as reasonable for a batch.

X1.3 *Explanation 3—Examples of Reference Materials that Challenge Test Methods*:

X1.3.1 The analytes selected for evaluating a test method should sufficiently challenge the test. The following examples illustrate this challenge:

X1.3.1.1 An amino acid should be used for checking a Kjeldahl nitrogen test because an ammonia standard would not sufficiently challenge the test.

X1.3.1.2 Various forms and species of metals should be used in checking whether a test method for total metals recovers all forms and species.

X1.3.1.3 Various species of cyanides should be used in checking whether a test method for total cyanide recovers all species.

X1.3.2 For some test methods, a more suitable material may be more applicable or appropriate than reagent water. The following examples illustrate alternatives:

X1.3.2.1 Ocean water (ASTM D 1141) for tests to be performed in a seawater matrix.

X1.3.2.2 Methanol as a conventional turbidity blank.

X1.3.2.3 A filter and/or suspended solid material for total suspended solids (TSS).

X1.4 *Explanation 4—Duplicate or Matrix Spike Duplicate*:

X1.4.1 The determination that a duplicate analysis is required must be made by the task group responsible for the method. The purpose of the duplicate is to determine the precision of measurements of the analyte(s) when the test method is applied to a specific sample in the batch and the result is applied to the validity of the test method for analysis of all samples in the batch.

X1.5 *Explanation 5—Applicability of Performance Criteria to Sample Matrices*:

X1.5.1 Committee D-19 test methods are typically validated in a variety of matrices. From this validation, a composite precision and bias statement is prepared. These tested matrices are considered to be those matrices for which the test method has been validated. Validation assures that the precision and bias of results on a given matrix is known (characterized) and of sufficient quality for its intended use. So long as the test remains in statistical control, further testing of the characterized matrix should result in a similar precision and bias.

X1.5.2 Performance specifications for the MS in a test method are applicable to the matrices tested in the collaborative study. This applicability may be extended to other matrices that present less of a challenge to the test method. For example, a test method validated on wastewaters from a variety of industries can be assumed to be applicable to drinking water. The taskgroup should recognize this applicability and not unnecessarily restrict the test method to only those matrices on which the method has been validated.

X1.5.3 It is an objective of this practice to establish absolute standards of performance for test methods so that data users know the limits within which the test method is being operated. Allowing development of less stringent performance criteria compromises this standard. For some intractable matrices, this compromise may be desirable. If the taskgroup expects that such matrices will be encountered in the use of a test method, the taskgroup should evaluate the intractable matrices and either find the means for overcoming the matrix problem or develop a separate set of MS and Dup performance criteria to allow for the matrix. Alternatively, if the Taskgroup believes that it is appropriate to allow for development of less stringent performance criteria by the method user, the taskgroup should insert the necessary language in the QC section of the test method that the method user must document the justification for use of less stringent performance criteria and make this documentation available to the user or client to whom the data will be reported.

X2. EXAMPLES

X2.1 *Example 1—Initial Demonstration of Performance*: Suppose a collaborative study is run on a new test method at the 10 mg/L level with 17 df, the pooled single operator precision at this level is found to be 0.4 mg/L. (The degrees of freedom for a single operator precision estimate are equal to the total number of analytical results actually used to produce the estimate minus the number of laboratories that generated these data.) A laboratory performs a precision study on the method at the 10 mg/L level using seven replicates (6 df). The standard deviation of these replicates is calculated to be 0.8 mg/L. The *F* value as obtained from Fig. X2.1 is 4.10. To determine whether this precision is satisfactory, use Eq X2.1:

$$\frac{(0.8)^2}{(0.4)^2} \leq 4.10 \qquad \text{(X2.1)}$$

$$\frac{0.64}{0.16} = 4.00 < 4.10$$

From this test, it can be seen that the novice operator's precision is satisfactory. Using this approach, a table of acceptable precision ranges can be prepared for different numbers of replicates. Using the data in this example, the table would appear as in Table X2.1.

X2.1.1 The data in Table X2.1 are based on applying Eq X2.1 to determine the highest acceptable standard deviation for the novice operator, (S_A). By transposing Eq X2.1, it can be seen that $(S_A)^2 \leq (S_O)^2F$. For example, in determining the highest acceptable standard deviation for a duplicate (two replicates; 1 df), the *F* value in Fig. X2.1 at the intersection of 1 and 17 is found to be 8.40. Using the data in X2.1, $S_O = 0.4$. Therefore, $(S_A)^2 \leq (0.4)^2(8.40) \leq 1.344$.

$$S_A \leq \sqrt{1.344} = 1.159 \qquad \text{(X2.2)}$$

Because 1.16 would not be acceptable, the value is rounded down to 1.15.

Degrees of Freedom for Denominator (dfS_0)	Degrees of Freedom for Numerator (dfS_1)										
	1	2	3	4	5	6	7	8	9	10	12
1	4052.	4999.	5403.	5625.	5764.	5859.	5928.	5981.	6022.	6056.	6106.
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.34	27.23	27.05
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.65	14.54	14.37
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89
6	13.74	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16
13	9.07	6.70	5.74	5.20	4.86	4.62	4.44	4.30	4.19	4.10	3.96
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46
18	8.28	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18

NOTE 1—Source—E. S. Pearson and H. O. Hartley, "Biometrika Tables for Statisticians," Vol II, Cambridge University Press, 1972.

FIG. X2.1 Critical Values of F at 1 % Significance (99 % Confidence) Level (One-Sided)

TABLE X2.1 Acceptable Precision Ranges

Number of Replicates in Precision Study	Acceptable Range of Standard Deviation at 10 mg/L (mg/L)
2	≤1.15
3	≤0.99
4	≤0.91
5	≤0.86
6	≤0.83
7	≤0.81
8	≤0.79
9	≤0.77
10	≤0.76

X2.1.2 Using the same example described in X2.1, suppose the laboratory finds a mean value of 11.4 mg/L when performing seven replicate determinations of a 10 mg/L solution. The mean value that ten laboratories found for the 10 mg/L solution in the collaborative study was 9.1 mg/L. The critical value of *t* at 9 df at the 99 % confidence level is 3.250 as shown in Table X2.2. The overall standard deviation (S_T) found in the collaborative study was 0.8 mg/L. Using Eq X2.3 to determine the acceptability of this result:

$$\left|\frac{11.4 - 9.1}{\sqrt{(0.8)^2 - \frac{(6)\,(0.4)^2}{7}}}\right| = 3.24 < 3.250 \qquad \text{(X2.3)}$$

The *t* test shows that the laboratory's mean value is acceptable.

X2.1.3 Using this approach, a table of acceptable mean concentration ranges can be prepared for different numbers of replicates. Using the data in this example, the table would appear as in Table X2.3. If necessary, this study should be repeated until the single operator precision and mean value obtained by the laboratory are within established limits.

TABLE X2.2 Critical Values of *t* at 1 % Significance (99 % Confidence Level)[A]

D.F.	*t* Value	D.F.	*t* Value
1	63.657	21	2.831
2	9.925	22	2.819
3	5.841	23	2.807
4	4.604	24	2.797
5	4.032	25	2.787
6	3.707	26	2.779
7	3.499	27	2.771
8	3.355	28	2.763
9	3.250	29	2.756
10	3.169	30	2.750
11	3.106	40	2.704
12	3.055	50	2.678
13	3.012	60	2.660
14	2.977	120	2.617
15	2.947	∞	2.576
16	2.921	. . .	. . .
17	2.898	. . .	. . .
18	2.878	. . .	. . .
19	2.861	. . .	. . .
20	2.845	. . .	. . .

[A] Source—*Statistical Methods for Chemists* by W. J. Youden, John Wiley & Sons, New York.

TABLE X2.3 Acceptable Mean Concentration Ranges

Number of Replicates in Precision Study	Acceptable Range of Mean Concentration, mg/L
2 or 3	6.7 to 11.5
≥4	6.8 to 11.4

X2.1.4 The data in Table X2.3 are based on applying Eq X2.3 to determine the range of the acceptable mean concentrations found by the novice operator ($\bar{X}_A$). By transposing Eq X2.3:

$$\bar{X} - \left|t_{0.99}\sqrt{S_T^2 - \frac{(n-1)\,S_O^2}{n}}\right| \le \bar{X}_A \le \bar{X} + \left|t_{0.99}\sqrt{S_T^2 - \frac{(n-1)\,S_O^2}{n}}\right| \qquad \text{(X2.4)}$$

These equations define the upper and lower bounds of the acceptable range. For example, in determining the acceptable range when seven replicates are used in the precision study, the following calculation would be used:

$$9.1 - \left|3.25\sqrt{(0.8)^2 - \frac{6\,(0.4)^2}{7}}\right| \le \bar{X}_A \le 9.1 + \left|3.25\sqrt{(0.8)^2 - \frac{6\,(0.4)^2}{7}}\right| \qquad \text{(X2.5)}$$

The lower end of the range calculates to be 6.795 and the high end of the range is found to be 11.405. The data in Table X2.3 are the nearest decimal values within the calculated limits.

X2.2 *Example 2—Example QC Test for MS Recovery*:

X2.2.1 Suppose a sample is analyzed and found to contain 8.2 mg/L of analyte. To check recovery, 2 mL of a 500 mg/L solution is added to 100 mL of sample and this spiked solution is analyzed. A value of 16.0 mg/L is found in the spiked sample. Percent recovery of the spike is calculated using Eq X2.6:

$$P = (100/500\ \text{mg/L}\ (0.002\ \text{L}))(16.0\ \text{mg/L}\ (0.100\ \text{L} + 0.002\ \text{L}) \qquad \text{(X2.6)}$$

$$- 8.2\ \text{mg/L}\ (0.100\ \text{L})) = (100\ \text{mg–1})(0.812\ \text{mg}) = 81.2\ \%$$

X2.2.2 Using the relationships given in the "Precision and Bias" section of the test method, at true concentration *T:*

$$\bar{x} = 0.990\ T + 0.10\ \text{mg/L} \qquad \text{(X2.7)}$$

and:

$$s = 0.050\ T = 0.050\ (\bar{x} - 0.10\ \text{mg/L})/0.990 = 0.0505\ (\bar{x} - 0.101\ \text{mg/L}) \qquad \text{(X2.8)}$$

X2.2.3 Then, produce the following estimates:

$$T = 500\ \text{mg/L}\ (0.002\ \text{L})\ /\ (0.100\ \text{L} + 0.002\ \text{L}) = 9.80\ \text{mg/L} \qquad \text{(X2.9)}$$

so:

$$\bar{x}_T = 0.940\ (9.80\ \text{mg/L}) + 0.10\ \text{mg/L} = 9.31\ \text{mg/L} \qquad \text{(X2.10)}$$

and:

$$\bar{P} = (100/500\ \text{mg/L}\ (0.002\ \text{L}))\ (9.31\ \text{mg/L}\ (0.100\ \text{L} + 0.002\ \text{L})) = 95\ \% \qquad \text{(X2.11)}$$

and:

$$s_A = 0.0505\ (16.0\ \text{mg/L} - 0.101\ \text{mg/L}) = 0.803\ \text{mg/L} \qquad \text{(X2.12)}$$

and:

$$s_B = 0.0505\ (8.2\ \text{mg/L} - 0.101\ \text{mg/L}) = 0.408\ \text{mg/L} \quad \text{(X2.13)}$$

so:

$$s_P = (100/1.00\ \text{mg})\ ((0.803\ \text{mg/L})^2\ (0.102\ \text{L})^2 \quad \text{(X2.14)}$$
$$+ (0.408\ \text{mg/L2}\ (0.100\ \text{L})2)1/2 = (100\ \text{mg–1})\ (0.00837\ \text{mg2})1/2 = 9.15\ \%$$

X2.2.4 Therefore:

$$\bar{P} - 3\ (s_P) = 95\ \% - 3\ (9.15\ \%) = 67\ \% \quad \text{(X2.15)}$$

and:

$$\bar{P} + 3\ (s_P) = 123\ \% \quad \text{(X2.16)}$$

Because $P = 81.2\ \%$ is within the recovery limits, the spike recovery is acceptable, indicating that there is no matrix effect.

X2.3 *Example 3—Example QC Test for Duplicates*:

X2.3.1 Suppose a sample is analyzed in duplicate and the values found are 8.5 and 12.5 mg/L. The single operator precision found in the collaborative study at the 10.5 mg/L level was 0.80 mg/L with 6 df. To determine whether the duplicate values are acceptable as compared to the single operator precision, an *F* test is used (see Eq 1). In the *F* test, S_A = standard deviation of the duplicate sample analysis values and S_O = the single operator precision found in the collaborative study:

S_A = 2.83 mg/L
S_O = 0.80 mg/L
$F_{0.99}$ at (1,6) = 13.74

$$\frac{(S_A)^2}{(S_O)^2} = \frac{8.01}{0.64} = 12.52 < 13.74 \quad \text{(X2.17)}$$

The duplicate values are acceptable.

X3. SUGGESTED WORDING FOR THE QC SECTION IN EACH TEST METHOD

X3.1 The following is suggested wording for the quality control (QC) section for each test method. This wording will vary from test method to test method and should be viewed as a guide:

NOTE X3.1—"X" represents the section number in the test method.

X. Quality Control (QC)

X.1 In order to be certain that analytical values obtained using this test method are valid and accurate within the confidence limits of the test, the following QC procedures must be followed when running the test:

X.2 Calibration and Calibration Verification (for Instruments and Analytical Systems Requiring Calibration or Standardization)

X.2.1 *Instrument:*

X.2.1.1 Analyze at least (number) calibration standards containing (concentration) of (analyte) prior to analysis of samples to calibrate the instrument.

X.2.1.2 Verify instrument calibration (frequency) by analyzing a standard at the concentration of one of the calibration standards (X.1.2.1.1). The (response (absorbance, intensity, etc) for external standard calibration) (response factor for internal standard calibration) shall fall within the limits in the following table (or within x % of the response or response factor) from the calibration).

(insert table)

X.2.1.3 If calibration cannot be verified, recalibrate the instrument.

X.2.2 Standardization (for analytical systems requiring standardization)

X.2.2.1 Standardize the analytical system on a (frequency) basis with the (normality/compound) titrant as follows: Transfer (number) mL of (standard solution) to a (container) and titrate with (normality/compound). The average (normality/molality) is used to calculate of the concentration of (analyte) in a sample.

X.2.2.2 Verify analytical system calibration (frequency) by analyzing an independent reference material at the concentration of the titrant (X.1.2.1.2). The (normality/molality) shall fall within the limits in the following table:

(insert table)

X.2.2.3 If analytical system standardization cannot be verified, re-standardize the system.

X.3 Initial Demonstration of Laboratory Capability

X.3.1 If a laboratory has not performed the test before or if there has been a major change in the measurement system, for example, new analyst, new instrument, etc., a precision and bias study must be performed to demonstrate laboratory capability.

X.3.2 Analyze seven replicates of a standard solution prepared from an IRM containing (concentration) of (analyte). The matrix and chemistry of the solution should be equivalent to the solution used in the collaborative study. Each replicate must be taken through the complete analytical test method including any sample preservation and pretreatment steps. The replicates may be interspersed with samples.

X.3.3 Calculate the mean and standard deviation of the seven values and compare to the acceptable ranges of precision and bias in the following table:

(insert table)

This study should be repeated until the single operator precision and the mean recovery are within the limits given in the table above. If a concentration other than the recommended concentration is used, refer to Test Method D 5847 for information on applying the F test and t test in evaluating the acceptability of the mean and standard deviation.

X.4 Laboratory Control Sample (LCS)

X.4.1 To ensure that the test method is in control, analyze an LCS containing (concentration) of (analyte) with each batch of (number of samples). If large numbers of samples are analyzed in the batch, analyze the LCS after every (number) samples. The LCS must be taken through all of the steps of the analytical method including sample preservation and pretreatment. The result obtained for the LCS shall fall within the limits in the following table:

(insert table)

X.4.2 If the result is not within these limits, analysis of samples is halted until the problem is corrected, and either all samples in the batch must be reanalyzed, or the results must be qualified with an indication that they do not fall within the performance criteria of the test method.

X.5 Method Blank (Blank)

X.5.1 Analyze a reagent water test blank with each batch. The concentration of (analyte) found in the blank must be less than (concentration). If the concentration of the (analyte) is found above this level, analysis of samples is halted until the contamination is eliminated and a blank shows no contamination at or above this level, or the results must be qualified with an indication that they do not fall within the performance criteria of the test method.

X.6 Matrix Spike (MS)

X.6.1 To check for interferences in the specific matrix being tested, perform an MS on at least one sample from each batch by spiking an aliquot of the sample with a known concentration of (analyte) and taking it through the analytical method.

X.6.2 The spike concentration plus the background concentration of (analyte) must not exceed (concentration of analyte). The spike must produce a concentration in the spiked sample 2 to 5 times the background concentration or 10 to 50 times the detection limit of the test method, whichever is greater.

X.6.3 Calculate the percent recovery of the spike (P) using the following formula:

$$P = 100\,\frac{|A(V_s + V) - BV_s|}{CV} \qquad \text{(X3.1)}$$

where:

A = concentration found in spiked sample,
B = concentration found in unspiked sample,
C = concentration of analyte in spiking solution,
V_s = volume of sample used, and
V = volume of spiking solution added.

X.6.4 The percent recovery of the spike shall fall within the limits in the following table:

(insert table)

If the percent recovery is not within these limits, a matrix interference may be present in the sample selected for spiking. Under these circumstances, one of the following remedies must be employed: the matrix interference must be removed, all samples in the batch must be analyzed by a test method not affected by the matrix interference, or the results must be qualified with an indication that they do not fall within the performance criteria of the test method.

X.7 Duplicate

X.7.1 To check the precision of sample analyses, analyze a sample in duplicate with each batch. If the concentration of the analyte is less than five times the detection limit for the analyte, an MSD should be used.

X.7.2 Calculate the standard deviation of the duplicate values and compare to the single operator precision in the collaborative study using an *F* test. Refer to 6.4.4 of Test Method D 5847 for information on applying the *F* test.

X.7.3 If the result exceeds the precision limit, the batch must be reanalyzed or the results must be qualified with an indication that they do not fall within the performance criteria of the test method.

X.8 Independent Reference Material (IRM)

X.8.1 In order to verify the quantitative value produced by the test method, analyze an IRM submitted as a regular sample (if practical) to the laboratory at least once per quarter. The concentration of the reference material should be in the range of (concentration of analyte) to (concentration of analyte). The value obtain must fall within the control limits specified by the outside source.