Designation: D 2777 – 98

Standard Practice for Determination of Precision and Bias of Applicable Test Methods of Committee D-19 on Water[1]

This standard is issued under the fixed designation D 2777; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 This practice establishes uniform standards for estimating and expressing the precision and bias of applicable test methods for Committee D-19 on Water.

1.2 Except as specified in 1.3, 1.4, and 1.5, this practice requires the task group proposing a new test method to carry out a collaborative study from which statements for precision (overall and single-operator standard deviation estimates) and bias can be developed. This practice provides general guidance to task groups in planning and conducting such determinations of precision and bias.

1.3 If a full-scale collaborative study is not technically feasible, due to the nature of the test method or instability of samples, the largest feasible scaled-down collaborative study shall be conducted to provide the best possible limited basis for estimating the overall and single-operator standard deviations.

1.3.1 Examples of acceptable scaled-down studies are the local-area studies conducted by Subcommittee D19.24 on microbiological methods because of inherent sample instability. These studies involve six or more completely independent local-area analysts who can begin analysis of uniform samples at an agreed upon time.

1.3.2 If uniform samples are not feasible under any circumstances, a statement of single-operator precision will meet the requirements of this practice. Whenever possible, this statement should be developed from data generated by independent multiple operators, each doing replicate analyses on independent samples of a specific matrix type, which generally fall within specified concentration ranges (see 7.2.5.2(*3*)).

1.3.3 This practice is not applicable to methodology involving continuous sampling or measurement, or both, of specific constituents and properties.

1.3.4 This practice is also not applicable to open-channel flow measurements.

1.4 A collaborative study that satisfied the requirements of the version of this practice in force when the study was conducted will continue to be considered an adequate basis for the precision and bias statement required in each test method. If the study does not satisfy the current minimum requirements for a collaborative study, a statement listing the study's deficiencies and a reference to this paragraph shall be included in the precision and bias statement as the basis for an exemption from the current requirements.

1.5 This paragraph relates to special exemptions not clearly acceptable under 1.3 or 1.4. With the approval of Committee D-19 on the recommendation of the Results Advisor and the Technical Operations Section of the Executive Subcommittee of Committee D-19, a statement giving a compelling reason why compliance with all or specific points of this practice cannot be achieved will meet both ASTM requirements (**1**)[2] and the related requirements of this practice. Precision and bias statements authorized by this paragraph shall include the date of approval by Committee D-19.

1.6 In principle, all test methods are covered by this practice.

1.7 In Section 11 this practice shows exemplary precision and bias statement formats for: (*1*) test methods yielding a numerical measure, (*2*) test methods yielding a non-numerical report of success or failure based on criteria specified in the procedure, and (*3*) test methods specifying that procedures in another ASTM test method are to be used with only insignificant modifications.

1.8 All studies, even those exempt from some requirements under 1.3 or 1.5, shall receive approval from the Results Advisor before being conducted (see Section 8) and after completion (see Section 12).

2. Referenced Documents

2.1 *ASTM Standards:*

D 1129 Terminology Relating to Water[3]

D 1141 Specification for Substitute Ocean Water[3]

D 1193 Specification for Reagent Water[3]

D 4375 Terminology for Basic Statistics in Committee D-19 on Water[3]

D 5790 Test Method for Measurement of Purgeable Organic Compounds in Water by Capillary Column Gas

[1] This practice is under the jurisdiction of ASTM Committee D-19 on Water and is the direct responsibility of Subcommittee D19.02 on General Specifications, Technical Resources, and Statistical Methods.

Current edition approved Jan. 10, 1998. Published October 1998. Originally published as D 2777 – 69 T. Last previous edition D 2777 – 96.

[2] The boldface numbers in parentheses refer to the list of standards at the end of this practice.

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

Chromatography/Mass Spectrometry[4]
D 5905 Specification for Substitute Wastewater[3]
E 177 Practice for Use of the Terms Precision and Bias in ASTM Test Methods[5]
E 178 Practice for Dealing with Outlying Observations[5]
E 456 Terminology Relating to Quality and Statistics[5]
E 1169 Guide for Conducting Ruggedness Tests[5]

3. Terminology

3.1 *Definitions*—For definitions of terms used in this practice, refer to Terminologies D 1129, D 4375 and E 456D 1129D 1193E 177, and Practice E 177D 5790.

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *accuracy*—a measure of the degree of conformity of a single test result generated by a specific procedure to the assumed or accepted true value and includes both precision and bias.

3.2.2 *bias*—the persistent positive or negative deviation of the average value of a test method from the assumed or accepted true value.

3.2.3 *laboratory*—a single and completely independent analytical system with its own specific apparatus, source of reagents, set of internal standard operating procedures, etc. Different laboratories will differ from each other in all of these aspects, regardless of how physically or organizationally close they may be to each other.

3.2.4 *operator*—usually the individual analyst within each laboratory who performs the test method throughout the collaborative study. However, for complicated test methods, the operator may be a team of individuals, each performing a specific function throughout the study.

3.2.5 *precision*—the degree of agreement of repeated measurements of the same property, expressed in terms of dispersion of test results about the arithmetical mean result obtained by repetitive testing of a homogeneous sample under specified conditions. The precision of a test method is expressed quantitatively as the standard deviation computed from the results of a series of controlled determinations.

4. Summary of Practice

4.1 After the task group has assured itself that the test method has had all preliminary evaluation work completed, it should prepare the test method write-up in final form. The plan for collaborative study is developed in accordance with this practice and submitted along with the test method write-up to the Results Advisor for concurrence except as specified in 1.3, 1.4, and 1.5. Upon receipt of concurrence, the collaborative test is conducted, data analyzed, and precision and bias statements formulated by the task group. The final precision and bias statistics must be based on retained data from at least six independent laboratories. The statements, with backup data including the reported results summary, the calculations leading up to the statements, and the test method write-up with precision and bias statements included are submitted to the subcommittee vice-chairman who in turn sends a copy of it to the Results Advisor for concurrence before balloting. This assures having an acceptable copy of the collaborative study results to send to ASTM for items on the main committee ballot. In most instances, the collaborative study shall be complete before a subcommittee ballot. If the collaborative study is not complete, the test method may go on the ballot as a provisional test method rather than a standard test method. Copies of the test data, approved calculations, and statistical results shall be filed at ASTM Headquarters when the test method is submitted by the subcommittee chairman as an item for the main committee ballot.

4.1.1 The appendix shows an example of "Form A—Approval of Plans for Interlaboratory Testing," as Fig. X1.1.

4.1.2 For an example of a data reporting form, see Fig. X2.1.

4.1.3 In addition, the appendix shows a sample calculation of precision and bias from real collaborative test data, the related table of statistics, and the related precision and bias statement.

5. Significance and Use

5.1 Following this practice should result in precision and bias statements which can be achieved by any laboratory properly using the test method studied. These precision and bias statements provide the basis for generic limits for use in the Quality Control section of the test method.

5.2 The method specifies the media for which the test method is appropriate. The collaborative test corroborates the write-up within the limitations of the test design. An extensive test can only use representative media so that universal applicability cannot be implied from the results.

5.3 The fundamental assumption of the collaborative study is that the media tested, the concentrations tested, and the participating laboratories are a representative and fair evaluation of the scope and applicability of the test method as written.

6. Preliminary Studies

6.1 Considerable pilot work on a test method must precede the determination of its precision and bias (**2,3**). This pilot work should explore such variables as preservation requirements, reaction time, concentration of reagents, interferences, calibration, and sample size. Potentially significant factors must be investigated and controlled in the written test method in advance of the collaborative test. Also, disregard of such factors may introduce so much variation among operators that results are misleading or inconclusive (**4**) (see 9.3 and 9.4). A ruggedness study conducted in a single laboratory is particularly useful for such investigations and should be conducted to prove a test method is ready for interlaboratory testing (see Guide E 1169E 178 for details).

6.2 Only after a proposed test method has been tried, proved, and reduced to unequivocal written form should a determination of its precision and bias be attempted.

7. Planning the Collaborative Test

7.1 Based upon the task group's knowledge of a test method and having the unequivocal write-up, several factors must be

[4] *Annual Book of ASTM Standards*, Vol 11.02.
[5] *Annual Book of ASTM Standards*, Vol 14.02.

considered in planning the collaborative test to properly assess the precision of the test method. The testing variables that must be considered in planning are discussed below. It is generally not acceptable to control significant sources of variability in the collaborative study which cannot be controlled in routine use of the test method, because this leads to false estimates of the test method precision and bias. In addition, the task group must determine within the resources available how to best estimate the bias of the test method.

7.2 *Testing Variables*:

7.2.1 It is desirable to develop a statement of precision of a test method that indicates the contribution to overall variation of selected causes such as laboratory, operator, sample matrix, analyte concentration, and other factors that may or have been shown to have strong effects on the results. Since any test method can be tried in only a limited number of applications, the standard deviation calculated from the results of a study can be only an estimate of the universe standard deviation. For this reason, the symbol s (sample standard deviation) is used herein. The precision estimates generated from the study data will usually be the overall standard deviation (s_T) and the pooled single-operator standard deviation (s_o) for each sample matrix and concentration studied.

7.2.2 Laboratories, operators, sample matrices, and analyte concentrations are the only sources of variability represented in the precision and bias statements resulting from the usual collaborative study. They may not represent the additional influence that can arise from differences in sample splitting, field preservation, transportation, etc., all of which may influence routine analytical results as shown in the general precision definitions in Terminology D 1129D 1129.

7.2.3 *Laboratories*—The final precision and bias statistics for each analyte, matrix, and concentration must be based on data from at least six laboratories that passed all of the outlier tests (see 10.3 and 10.4), that is, retained data. To be assured of meeting this requirement, it is recommended that usable data be obtained from a minimum of eight independent laboratories. To guarantee eight providing usable data, it will often be necessary to get ten or more laboratories to agree to participate, because some may not provide data and others may not provide usable data. Maximizing the number of participating laboratories is often the most important thing that can be done to guarantee a successful study.

7.2.4 Even if the single-operator standard deviation is the only statistic to be estimated in the study (see 1.3.2), there should be a minimum of eight operators providing usable data, so you are assured of data from six operators after all outlier removal.

7.2.5 *Sample Matrices*—The collaborative study shall be conducted with at least one representative sample matrix, which should be reproducible by subsequent user-laboratories so that they can compare their results with the results of the collaborative study.

7.2.5.1 Typically, a reagent water prepared according to Specification D 1193D 1193 or a synthetic medium, such as the substitute wastewater described in Specification D 5905D 5905 or the substitute ocean water described in Specification D 1141D 1141 , is used as the reference matrix. Analytes may be supplied separately as concentrates for addition to this matrix by each laboratory or the reference matrix containing the analyte(s) may be supplied to each participant. Information on how the reference matrix was prepared in the study shall be clear in the precision and bias statement of the test method so users can reproduce it properly.

7.2.5.2 Additional collaborative testing should also be conducted using other matrices specified in the scope of the test method. Since these matrices must be the same for each study participant, they may have to be prepared (or obtained from a single source), preserved, and distributed to all laboratories. As with the reference matrix, analytes may be supplied in a separate spiking solution or already added to the matrix. A particularly attractive matrix might be a standard material available from an organization such as the National Institute of Standards and Technology (NIST). Use of uniform sample matrices is necessary in these studies since they enable a more certain comparison with the reference matrix than is possibly with matrices supplied separately by each participant.

(1) Use of matrices with naturally occurring, non-zero background levels of the analyte(s) being studied will result in precision and bias estimates that will be much more difficult to properly compare with estimates from the reference matrix.

(2) Any matrix spiking that may be necessary shall not significantly change the natural characteristics of the matrix.

(3) With the exception of the kind of limited study described in 1.3.2, the matrix-of-choice approach, in which each participant is expected to acquire their own sample of a designated type, should not be used. Such studies are basically incompatible with the statistical approaches employed in this practice; both the ranking test and the individual outlier test are incapable of distinguishing laboratory effects from matrix effects. In addition, the presence of variable background concentrations prevents the assignment of a proper mean concentration level to each precision estimate produced in the study.

7.2.5.3 The same study design should be used for all sample matrices. A separate precision and bias statement should be generated for each sample matrix with a brief description of the matrix tested.

7.2.5.4 When studies are available indicating the applicability of the test method for matrices untested in 7.2.5.1 and 7.2.5.2 and not meeting the other requirements of this practice, at the discretion of the task group responsible for the test method and the Results Advisor, and providing the data are analyzed in accordance with Section 10 of this practice, this supporting data may be included in a separate section of the precision and bias statement. A clear but brief description of the matrices shall be included and the study protocol employed. It is the intent of this practice that ultimately, data concerning the precision and bias of the test method in the full range of matrices covered in the scope and analyzed in accordance with this practice, will be made available to the users of the test method.

7.2.6 *Analyte Concentrations*—If pilot work has shown that precision is linear with increasing analyte concentrations, at least three Youden pairs (**5**), that is, six concentrations, covering the range of the test method should be included for

each matrix. If the pilot work suggests that precision is other than constant or linear, more concentration levels should be analyzed. The study concentrations should generally be rather uniformly distributed over the range of the test method.

7.2.6.1 Study samples with concentrations at or near the detection limit of a test method are likely to produce non-quantitative results from many of the participating laboratories if participants are permitted to use their detection limit to censor their results. Zeroes or less thans that result from this censoring process are non-quantitative results and cannot be included in the statistical analysis of study results specified later in this practice. Conducting the specified statistical analysis on whatever quantitative data are available under such circumstances can produce misleading precision and bias estimates. If it is considered necessary to include samples at or near the detection limit, such samples shall be in addition to the minimum required three Youden pairs at concentrations that can be readily measured by qualified laboratories. Data from analyses of the basic three or more Youden pairs that can be quantified can then be statistically analyzed as specified to produce a proper traditional precision and bias statement for the test method. Results from analyses of Youden pairs at or near the detection limit can be included in this traditional statistical analysis if it turns out that most laboratories report quantified results. Otherwise, results for low-level samples must be statistically analyzed using specialized procedures, for example, procedures similar to those under development in Subcommittee D19.02, which are beyond the scope of this practice.

7.2.7 Since the order of analyses should not be a source of systematic variability in the study, each participant should either be told to randomize the order of study sample analyses or be given a specific random order for their analyses.

7.2.7.1 Whenever the time of analyses has been shown to influence the analytical results, close control over the time of analyses will be essential.

7.2.8 If pilot work has shown that the sample container must be of a specific material prepared in a specific manner prior to use, the variation in containers obtained and prepared by the participants will be a random variable and should be treated as such in the planning of the study and in the statistical analysis of the data.

7.2.9 The manner of preservation or other treatment of the sample prior to typical use of the test method, if known to affect the precision or bias, or both, of results, shall be incorporated into the collaborative study design.

7.3 *Measurement of Precision*:

7.3.1 Every interlaboratory study done to provide precision and bias estimates for a D-19 test method must use a Youden-pair design rather than a replicate sample design. Justifiable exceptions to this requirement shall be approved through the process provided in 1.5. In a Youden-pair design, each participant receives (or prepares from a concentrate) a separate sample *for each analysis required in the study*. There are no replicate analyses; each participant analyzes each study sample once and only once, per analyte if appropriate. Among the set of samples each laboratory analyzes for a specific matrix, there are pairs of samples containing similar but usually different analyte concentrations that differ from each other by up to 20 %. As a matter of convenience to whomever is preparing the samples or spiking concentrates, up to half the Youden pairs may have the same concentration, that is, be blind duplicates, but the participants must have no basis for comparing their single test results from analyses of different study samples.

7.3.2 The only difference in treatment of data from a Youden-pair study is the calculation used to estimate the means and standard deviations; these calculations may be found in Youden and Steiner (**6**). Once developed, these mean and standard deviation estimates are treated the same as statistics from a study with the usual replicate design. A detailed example with and without raw experimental data is given in Refs. (**7**) and (**8**), respectively.

7.3.3 The value of the nonreplicate design is that the single-operator standard deviation estimates are free of any conscious or unconscious analyst bias. The procedures for calculating overall and single-operator standard deviations are given in 10.4 and 10.5 and illustrated in Appendix X3.

7.4 *Measurement of Bias*:

7.4.1 The concept of accuracy comprises both precision and bias (see Terminology D 1129D 1129 and Practice E 177E 177). As discussed in Practice E 177E 177, there is not a single form for statements of accuracy that can be universally recommended. Since the accuracy of a measurement process is affected by both random and systematic sources of error, measures of both kinds of error are needed. The standard deviation is a universal measure of random sources of error (or precision). Bias is a measure of the systematic errors of a test method.

7.4.2 A collaborative study evaluation of bias for a specific matrix produces a set of analyte/sample means. The difference between a true value (however defined) and the related mean is an estimate of the average systematic error, that is, bias of the test method.

7.4.3 There are three major approaches commonly used to test a measurement procedure: (*1*) measurement of known materials, (*2*) comparison with other measurement procedures, and (*3*) comparison with modifications of the procedure itself (**9**). The third approach may involve the standard addition technique or the simultaneous analysis of several aliquots of different sizes (for example, 0.5, 1, 1.5, 2, 2.5 units). The task group will select the approach that best suits its needs within the resources available to it.

7.4.4 The most likely task group approach will be the use of known materials. Since reference standards are unlikely to be available, the task group will prepare its samples with added (therefore known to them) quantities of the constituent(s) being tested. The best available chemical and analytical techniques for preparing, stabilizing, if necessary, storing and shipping the prepared samples should be known within the task group and will not be addressed in this practice. However, if the sample preparation and handling techniques used for the study are different from those expected to be used for samples during routine application of the test method, those differences shall be pointed out in the precision and bias statement. Future users

of the test method may decide that these differences had an effect on the precision or bias results, or both, from the study.

7.5 *Quality Control During the Study*:

7.5.1 The Quality Control section to appear in the test method must be drafted before the collaborative study design is finalized and the study design must assure that the collaborative study will produce any background data not otherwise available to properly complete the final Quality Control section. Each part of the draft Quality Control section must be used during the collaborative study unless insufficient background data exist to establish credible interim required performance criteria for that part.

7.5.2 All quality control data/information produced to meet the requirements of 7.5.1 shall be reported to the task group chair along with results from analyses on the study samples.

8. Collaborative Study Design Approval

8.1 After approval by the task group, the task group chair (or designee) will summarize the proposed design of the collaborative study. This summary will include: (*1*) the test method to be tested in ASTM format and as approved by the task group; (*2*) the analytes to be included in the study; (*3*) the number of samples in accordance with the paired-sample plan of 7.3.1; (*4*) the approach for determining the bias of the test method as exemplified in the collaborative study; (*5*) the range of concentration covered, and approximate concentration of material in each sample or set; (*6*) the approximate number of laboratories and analysts; (*7*) the matrices and QC samples being tested; (*8*) plans for developing study samples; and (*9*) a copy of the instruction and data reporting package to be given to each study participant. This summary should be presented to the Results Advisor in the form of a letter.

8.1.1 As an aid, the task group chairman may use, "Approval of Plans for Interlaboratory Testing," Form A, and in Appendix X1 (a completed example is shown in Fig. X1.1).

8.2 Upon review of the plan, the Results Advisor will advise the task group chairman whether the plan meets the requirements of this practice or what changes are necessary to meet the requirements of this practice.

8.3 Upon receipt of approval of the collaborative test plan by the Results Advisor, the task group chairman (or designee) will conduct the collaborative test.

9. Conducting the Collaborative Study

9.1 A single entity, acting for the task group, will prepare the samples for the collaborative study and ship them to the participants with instructions for the study, a copy of the exact test method (if not already supplied), and the participant reporting form (or reporting instructions).

9.1.1 The instructions for the collaborative study shall require sufficient preliminary work by potential collaborators to adequately familiarize them with the test method prior to study measurements. This is necessary to ensure that each collaborative study is made by a peer group and that a learning experience is not included in the statistics of the collaborative study. The task group may also develop procedures to qualify prospective collaborators, and this approach is strongly recommended.

9.1.2 Each laboratory should usually supply its own calibration materials, as independent calibration materials are a significant source of interlaboratory variability. However, if the cost of availability of calibration materials is judged to be a significant deterrent to participation or if currently available materials are inadequate and not considered typical for subsequent routine use of the test method, these materials may be distributed with the study samples. If calibration standards are provided, the Precision and Bias section of the test method should so note, including the concentrations and matrix of the standards and any specific instructions for their use.

9.1.3 As an aid, the task group chairman may use Form B, "Data Report from Individual Laboratories," as in Appendix X2 (a completed example is shown in Fig. X2.1).

9.2 The batch of samples containing a specific member of a Youden pair should be clearly marked with a common unique code, informative to the distributors but not informative to the study participants. Samples should be sized to supply more than the minimum amount necessary to participate in the study (with reasonable allowance for pipetting, rinsing, etc.) to allow for trial runs and analytical restarts that may be necessary. A separate set of samples shall be provided for each operator. Sample concentrations should not be easily surmised values (1, 5, etc.). The assignment of samples to the participating laboratories should be randomized within each concentration level. The above recommendations should help assure statistical independence of results.

9.3 A copy of the test method under investigation, the written instructions for carrying out his/her part of the program, and the necessary study samples should be supplied to each operator. No supplementary instructions or explanations such as by telephone or from a task group member within a cooperating laboratory should be supplied to one participant if not to all. Study materials should be distributed from one location, and the operator's reports should be returned to one location.

9.4 The written instructions should cover such items as: (*1*) directives for storing and subdividing the sample; (*2*) preparation of sample prior to using the test method; (*3*) order of analyses of samples (random order within each laboratory is often best); (*4*) details regarding the reporting of study results on the reporting form; and (*5*) the time limit for return of the reporting form.

9.4.1 Laboratories shall be required to report all figures obtained in making measurements, instead of rounding results before recording them. This may result in recording one or more significant figures beyond what may be usual in the Report section of the test method. A decision about rounding all data can be made by the task group when the final statistical analyses are performed.

9.4.2 The laboratories shall report results from analyses of study samples without background subtraction and shall also report background levels for every matrix that they use in the study. The task group will make any background corrections that may be necessary.

9.4.3 Zeros and negative numbers should be reported whenever they represent the actual test results produced. Test results should never be censored by a participant. The reporting of less

than or greater than results negates the objectivity of subsequent statistical calculations and should be avoided. Never report zero in place of a less-than or other nonquantitative test result.

9.5 The task group chair (or designee) should monitor the collaborative study to assure that results are reported back within the agreed upon time limit and are free of obvious procedural, transcription, clerical, or calculation errors. Careful design of the reporting form (or reporting instructions) will facilitate this task.

10. Collaborative Study Data Analysis

10.1 For each matrix/analyte, the steps involved by the task group chair in the data analysis consist of: (*1*) tabulating the data; (*2*) eliminating any laboratories that did not follow significant study instructions, were not in control during the study, or were so consistently high or low that their results are unreasonable (see 10.3); (*3*) eliminating any individual outlier data points (10.4); (*4*) for each matrix and analyte concentration studied, calculating the overall and single-operator standard deviations and means from the retained data and calculating the bias from each mean spike recovery (must subtract the mean reported background value whenever necessary); (*5*) tabulating the statistics; (*6*) assembling information required for the research report; and, if desired, (*7*) summarizing these results in a graph or regression equation for the test method statement.

10.1.1 As an aid to following the steps, the task group chair may find it helpful to review the sample calculations of precision and bias given in Appendix X3.

10.2 *Tabulation of Data*—The data reported by the laboratories shall be made consistent in reporting units and, if possible, in the number of reported values per operator or laboratory (**10**). Before beginning, remove any unusable data sets generated by laboratories that did not follow significant study instructions or used an unacceptable variation of the test method being studied. Unless each laboratory used its own matrix with a unique background concentration, all outlier testing and precision estimates are to be based on the concentration reported rather than on background-corrected results.

10.2.1 Sometimes looking at the histogram of a set of data will help one recognize or understand, or both, the cause of unusual data.

10.3 *Rejection of Outlier Laboratories*— If one or more laboratory's data for an analyte in a specific matrix are so consistently high or low that there must be a large systematic error specific to that laboratory, all the data from the laboratory for that analyte/matrix should be rejected. Identify outlier laboratories by applying the Youden laboratory ranking test (**11**) at the 5 % significance level.

10.3.1 For example, say *n* laboratories reported results for a specific matrix and analyte. Within the data set reported for each concentration, assign a rank score from 1 for the highest result to *n* for the lowest result.

10.3.1.1 For this test, all *n* rank scores for each concentration shall be assigned, even if one or more of the laboratories did not report a result for this particular concentration. The rank of any missing results should be the mean rank of the actual data reported by that laboratory for the other concentrations of the same analyte and matrix. Also, assign an appropriate rank to nonquantitative results.

10.3.1.2 Identical results would each be given the average of the ranks the group is entitled to receive.

10.3.2 For the matrix/analyte, total the rank scores for each laboratory over all of the *q* concentrations. If the total rank sum for any particular laboratory is designated as *R*, then if either:

$R <$ the lower value in Table 1, or

$R >$ the upper value in Table 1,

that laboratory is a candidate to be marked as an outlier and ignored in subsequent calculations with a 5 % risk of this judgement being incorrect.

10.3.2.1 If more than 20 % of the laboratories reporting usable data for the matrix/analyte are outlier candidates, order the candidate laboratories according to the difference between their total rank sum and the nearest critical value given above, and reject individual or tied groups of laboratories until rejection of the next laboratory would exceed the 20 % limit. If rejection of a group of laboratories with equal distances would cause the 20 % limit to be exceeded, randomly reject laboratories from the group until rejection of the next laboratory would exceed the 20 % limit. Data from laboratories ultimately marked as outliers should be ignored for subsequent calculations.

10.3.3 Repeat 10.3 for every matrix and analyte studied.

10.4 *Rejection of Unusable Data and Individual Outlier Results, and Calculation of Final Mean and Overall Standard Deviation Estimates*:

10.4.1 Reject nonquantitative responses since they are useless for subsequent calculations. These rejections do not count against the 10 % limit in 10.4.4 because such responses are unusable. It is the task group's responsibility to judge whether reported zeros are truly quantitative analytical results, and this should usually be done after consulting with each laboratory that reported a zero, whenever that is possible.

10.4.2 Let the remaining data reported for a specific matrix/analyte/concentration be designated x_i, $i = 1$ to n. Then calculate the mean ($\bar{x}$) and overall standard deviation (s_T) as follows:

$$\bar{x} = \frac{\left(\sum_{i=1}^{n} x_i \right)}{n} \tag{1}$$

and

$$s_T = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}} \tag{2}$$

10.4.3 Calculate the *T* value for the most extreme remaining value (x_e) as follows:

$$T = (x_e - \bar{x}) / s_T \tag{3}$$

If the absolute value of *T* is greater than the critical value for *n* measurements from Table 2, x_e is considered an outlier value and ignored for subsequent calculations (**12**,**13**,**14**).

10.4.4 If an outlier was just removed in 10.4.3, return to 10.4.2 unless the removal of one more individual outlier would exceed 10 % of the usable data originally reported for this