



Standard Practice for Intralaboratory Quality Control Procedures and a Discussion on Reporting Low-Level Data¹

This standard is issued under the fixed designation D 4210; 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.

^{ε1} NOTE—Keywords were added editorially in May 1996.

1. Scope

1.1 This practice is applicable to all laboratories that provide chemical and physical measurements in water, and provides guidelines for intralaboratory control and suggested procedures for reporting low-level data.

1.2 The use of this practice is based on the assumptions that the analytical method used is appropriate for the task, is either essentially bias-free or the bias is known, is capable of being brought into a state of statistical control, and possesses adequate sensitivity to determine the analytes at the levels of interest.

1.3 Further, it is assumed that quality assurance procedures for field operations such as sample collection, container selection, preservation, transportation, and storage are proper.

1.4 This practice is also predicated upon the laboratory already having established a quality control system with development of an adequate reporting system such that the laboratory's performance can be substantiated.

2. Referenced Documents

2.1 *ASTM Standards:*

D 1129 Terminology Relating to Water²

3. Terminology

3.1 *Definitions of Terms Specific to This Standard:*

3.1.1 *control charts*—a charting of the variability of a procedure such that when some limit in variability is exceeded the method is deemed to be out of control.

3.1.2 *control limits*—those upper and lower limits used to signal that a procedure is out of control.

3.1.3 *criterion of detection*—the minimum quantity (analytical result) which must be observed before it can be stated that a substance has been discerned with an acceptable probability that the statement is true (see 11.11). The criterion of detection must always be accompanied by the stated probability.

3.1.4 *in control*—once a reliable estimate of the population standard deviation is obtained, a deviation not exceeding 3σ is considered to be in control. Allowing deviations up to 3σ imply an α (alpha) = 0.0027 or about 3 chances in 1000 of judging an in control procedure to be out of control.

3.1.5 *limit of detection*—a concentration of twice the criterion of detection when it has been decided that the risk of making a Type II error is to be equal to a Type I error (see 11.11).

3.1.6 *Type I error, α (alpha) error*—a statement that a substance is present when it is not.

3.1.7 *Type II error, β (beta) error*—a statement that a substance is not present (was not found) when the substance was present.

3.2 *Definitions*—For definitions of other terms used in this practice, refer to Terminology D 1129.

4. Significance and Use

4.1 Any analytical procedure that is in statistical control will have an inherent variability as one of its characteristics. For a given procedure this variability is irreducible, that is, there is no identifiable factor or assignable cause that contributes to procedure variation.

4.2 The measure of procedure variability for this practice is the estimate of the population standard deviation. The specific population of interest can be either within an analytical set or between set analyses or both.

4.3 In considering low level reporting the question is: is the substance present? This practice will aid in determining the risk taken in assigning that a substance is present, when it is not, and provide an assessment of criterion of detection.

4.4 Procedure variability control limits are set by use of Shewhart control charts.³

5. Estimating Analytical Procedure Variability by Duplicate Analyses

5.1 For a crude estimate of population standard deviation, initially conduct 5 or 6 duplicate analyses from samples of nearly the same concentration. Accumulate additional data to obtain a reliable initial estimate of the population standard

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

Current edition approved Jan. 27, 1989. Published March 1989. Originally published as D 4210 – 83. Last previous edition D 4210 – 83.

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

³ "Presentation of Data and Control Chart Analysis," *ASTM STP 15-D*, ASTM, 1976, pp. 93–103.

deviation in which 40 to 50 data points (degrees of freedom) are needed. They may be analyses of duplicate samples or standards determined either within analytical-set or between sets depending on the information sought. However, with highly labile constituents only within set analyses would be appropriate.

5.2 After performing the duplicate analyses, determine the average difference between duplicates and divide this by 1.128 to estimate the standard deviation.³ For an example of this calculation refer to Annex A1.

5.3 Prepare necessary control charts as described in Section 9.

6. Estimating Analytical Procedure Variability Using a Stable Standard

6.1 Using a stable standard in replicate for 50 or more data points the procedure variability is estimated by calculating an estimate of the standard deviation in the usual way,

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

where:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

6.2 A discussion and illustration of the procedure is given in Annex A2.

6.3 Prepare a control chart with upper and lower limits as described in Section 9.

7. Pooling Estimates to Improve Estimation of Standard Deviation

7.1 As additional data are obtained initial estimates of variability can be put on a sounder footing by pooling with estimates from the new information, assuming that no substantial change is apparent. To test for significant change in variability the ratio of the two estimates s_1^2/s_2^2 is calculated and compared to appropriate values of the *F* distribution to test if pooling the estimates of variability is proper.

7.2 A discussion on and illustration of how to determine if the estimates of analytical procedure variance had changed to where they should not be combined is given in Annex A3.

7.3 If a procedure variability appears to have changed significantly, the procedure should be carefully reviewed to ascertain the cause.

7.4 When it appears that the variability of an analytical procedure has not changed, a pooled estimate of variability may be obtained.

8. Pooling Estimates of Variability

8.1 The pooling method consists of weighting the two *variance* estimates by the degrees of freedom of the respective data sets from which they were obtained, summing the weighted variance estimates, and dividing the sum by the sum of the degrees of freedom associated with the two estimates. The quotient which results is the pooled variance estimate, s^2 , from which the new, pooled estimate of the standard deviation, *s*, is obtained.

8.2 Using the data of A3.1

$$s^2 = [((df_1)s_1^2 + (df_2)s_2^2)/(df_1 + df_2)]$$

$$= [((n_1 - 1)s_1^2 + (n_2 - 1)s_2^2)/(n_1 + n_2 - 2)]$$

$$s^2 = [((60)s_1^2 + (40)s_2^2)/(60 + 40)]$$

$$= [(60(1.796)^2 + 40(2.145)^2)/(60 + 40)]$$

$$s^2 = (193.537 + 184.041)/100$$

$$s^2 = 3.776$$

$$s = 1.943 \mu\text{g/L}$$

When a pooled estimate of the procedure standard deviation is obtained, new control limits should be calculated using the revised estimate.

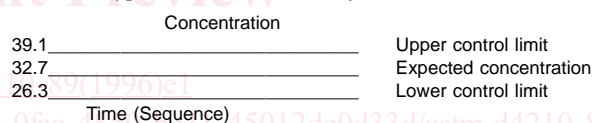
9. Setting Control Limits

9.1 There are two goals in setting control limits. They should be close enough to signal when there is trouble with a system, and they should be distant enough to discourage tinkering with a system that is operating within its capabilities. Since these two goals are in opposition, a compromise is necessary. The compromise which has been found satisfactory in a great many applications is the use of 3σ control limits, and they are illustrated here in 9.2. Warning control limits are described in 9.5.1.

9.2 *Use of a Standard:*

9.2.1 Consider a sample whose concentration was prepared as 32.7 $\mu\text{g/L}$ and is analyzed by a procedure whose estimated standard deviation is 2.131 $\mu\text{g/L}$. The control limits are therefore $32.7 \pm 3 \times 2.131$ or 26.31 and 39.09. Assuming that results can be read to tenths of a microgram, a result ≥ 26.3 and ≤ 39.1 is judged acceptable.

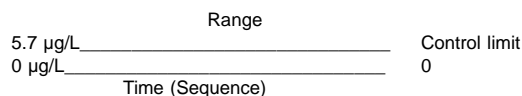
9.2.2 *Typical Control Chart for Standards:*



9.3 *Use of an Unknown Duplicate:*

9.3.1 Suppose an unknown duplicate sample is analyzed in separate runs by a procedure whose estimated standard deviation is 1.537 $\mu\text{g/L}$. The control limit for the *range* of the two analyses is 1.537×3.686 or 5.67 (3.686 is the proper factor for duplicate ranges).² Assuming that results can be read to tenths of a microgram, an absolute difference between the duplicates (their range) ≤ 5.7 is judged acceptable.

9.3.2 *Typical Control Chart for Duplicate Analyses Ranges:*



9.4 *A Special Case, Use of Recovery Data:*

9.4.1 The use of recovery data from spiked samples for control purposes presents some special problems which are dealt with in Annex A4. Begin with the estimation of the variability associated with the determination of recoveries.

9.4.2 If the spiking recovery demonstrates a bias, the control limits must be centered about the estimate of the bias.

9.4.3 Suppose the calculated estimation of spike population variation expressed as a standard deviation is found to be 0.1532 mg/L as illustrated in Annex A4, then control limits would be $\pm 3 \times 0.1532$ or -0.46 mg/L and $+0.46 \text{ mg/L}$.

9.5 Warning Limits:

9.5.1 Some analysts prefer to use warning limits 2σ , along with the typical 3σ limits previously described. For 2σ limits the factors (f) to use times the standard deviation $[(f)s]$ are respectively (9.2), $f = 2$; (9.3), $f = 2.834$; (9.4), $f = 2$.

10. Recommended Control Sample Frequency

10.1 Until experience with the method dictates otherwise, to monitor accuracy, one quality control sample of expected value should be included with every ten analyses or with each batch, whichever results in the greater frequency.

10.2 To monitor precision, one quality control sample should be included with every 10 analyses or with each batch of analyses run at the same time, whichever results in the greater frequency. If duplicates are used to monitor precision, they should be analysed in different runs when a between run measure of variability is employed in setting control limits. If the method demonstrates a high degree of reliability, control sample frequency can be appropriately relaxed.

11. A Discussion on Reporting Low-Level Data

11.1 There are specific problems in the reporting of low-level data which are associated with the question: is a substance present?

11.2 In answering the question “is a substance present?”, there are two possible correct conclusions which may be reached. One may conclude that the substance is present when it is present, and one may conclude that the substance is not present (see Note 1) when it is not present. Conversely, there are two possible erroneous conclusions which may be reached. One may conclude that the substance is present when it is not, and one may conclude that the substance is not present when it is. The first kind of error, finding something which is not there, is called a TYPE I ERROR. The second kind of error, not finding something which is there, is called a TYPE II ERROR.

NOTE 1—Since Avogadro’s number is very large, one could argue that one should never claim that a substance is not present. A common sense meaning of not present is intended here, that is, if measurement is being made in micrograms per litre the presence of a few nanograms per litre is irrelevant.

11.3 These two types of errors are illustrated in the material that follows, using the result which might be obtained from a

single analysis when the substance is not present to illustrate Type I error and the inferences that might be drawn from a single analysis at two different actual concentrations to illustrate Type II error. Of course inferences as to water quality are seldom, if ever, based on the result of a single analysis. A single result is used here to simplify the exposition.

11.4 If the standard deviation, σ , of an analytical procedure has been determined at low concentrations including 0, then the probability of making a Type I error can be set by choosing an appropriate α (alpha) level to determine the criterion of detection (see 3.1.3).

11.5 For example, suppose that the standard deviation, σ , of an analytical procedure is $6 \mu\text{g/L}$ and that an α (alpha) of 0.05 is deemed acceptable so that the probability of making a Type I error is set at 5%. The criterion of detection can then be found from a table of cumulative normal probabilities to be $1.645 \sigma = 1.645 (6 \mu\text{g/L}) \approx 10 \mu\text{g/L}$ (see Fig. 1).

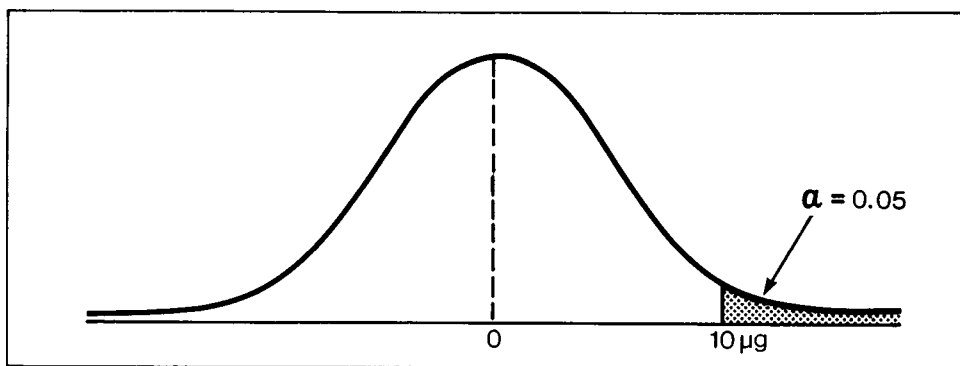
11.6 Any value observed below $10 \mu\text{g/L}$ would be reported as less than the criterion of detection, since to report such a value otherwise would increase the probability of making a Type I error beyond 5%.

11.7 Note that the context of decision is the analytical result produced by the laboratory. A result is obtained and a response made to it. Nothing has been said concerning the ability to detect a substance which is present at a specified concentration.

11.8 Once the criterion of detection has been set, the probability of making a Type II error, β (beta), or its complement $1-\beta$, the probability of discerning the substance when it is present, can be determined for given true situations. (The probability $1-\beta$ is sometimes called the power of the test).

11.9 Consider the same analytical procedure as described in this section with a criterion of detection of $10 \mu\text{g/L}$. Suppose that the concentration of the sample being analyzed is $10 \mu\text{g/L}$, that is, the concentration is equal to the criterion of detection and if all analytical results below the criterion of detection were reported as such, then the probability of discerning the substance would be 0.5 or 50% (see Fig. 2).

11.10 Conversely, the probability of making a Type II error and failing to discern the substance would also be 0.5. From this example it can be seen that the probability of discerning a substance when its concentration is equal to the criterion of detection is hardly overwhelming. In order for the probability



Normal Frequency Curve
 FIG. 1 Probability of Type I Error

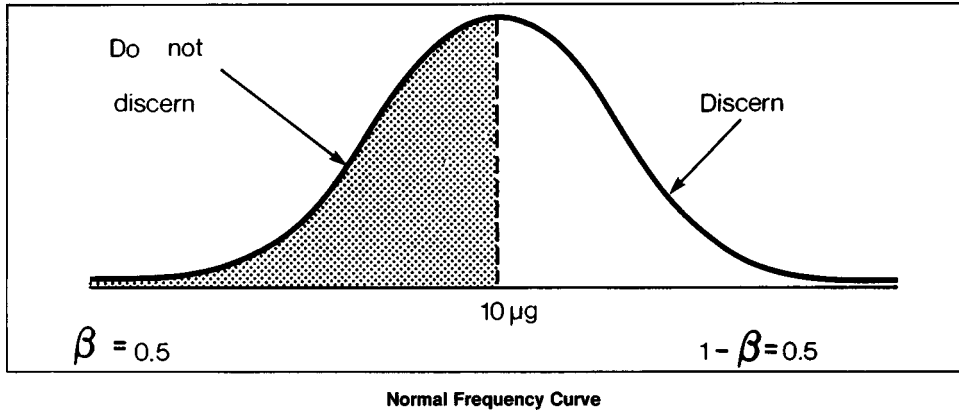


FIG. 2 Probability of Type II Error, True Value = Criterion of Detection

of a Type II error to be equal to the probability of a Type I error, $\beta(\text{beta}) = \alpha(\text{alpha})$, the concentration of the sample being analyzed must be twice the criterion of detection.

11.10.1 This concentration of twice the criterion of detection is the *limit of detection* when it has been decided that the risk of making a Type II error is to be equal to the risk of making a Type I error (see Fig. 3).

11.11 The concept of Type II error has been emphasized because generally, attention is paid to the avoidance of Type I error with no consideration given to the probability of making a Type II error. It should also be recognized that when the probability of a Type I error is decreased by selecting a lower $\alpha(\text{alpha})$ -level, the probability of making a Type II error is increased.

11.11.1 Having clarified the conceptual context in which an $\alpha(\text{alpha})$ -level is set and the difference between the criterion of detection and the limit of detection, the reporting of low-level data can be considered.

11.12 Results reported as “less than” or “below the criterion of detection,” are virtually useless for either estimating outfall and tributary loadings or concentrations for example.

12. Two Codes, “W” and “T,” Are Suggested for Low-Level Reporting

12.1 The T code has the following meaning: “Value reported is less than criterion of detection.” The use of this code warns

the data user that the individual datum with which it is associated does not, in the judgment of the laboratory that did the analysis, differ significantly from 0.

12.2 It should be recognized an implied significance test which fails to reject the null hypothesis, that a result does not differ from a standard value, in no way diminishes the value of the result as an estimate. To illustrate: A result of 9 μg on a test whose $\sigma = 6 \mu\text{g}$ cannot be regarded as significantly different from 0 for any $\alpha(\text{alpha})$ -level less than 0.067; however, if a significance test were made with $\alpha(\text{alpha}) = 0.1$, then the null hypothesis would be rejected and the result deemed significantly different from 0.

12.2.1 So the result, 9 μg, could be reported as “below the criterion of detection” for all $\alpha(\text{alpha})$ less than 0.067 and could be reported as simply “9 μg” for all $\alpha(\text{alpha})$ greater than 0.067. But however reported, the result of 9 μg remains the best estimate of the true value since changing the risk of making a Type I error neither augments or diminishes the value of an estimate. In practice, this consideration means that if a number can be obtained, it may be reported along with the appropriate codes and their definition.

12.2.2 It may be added that low-level results are better estimates, in the sense of being more precise in an absolute value, than higher results since for many analytical tests with which one is acquainted the standard deviation of the test

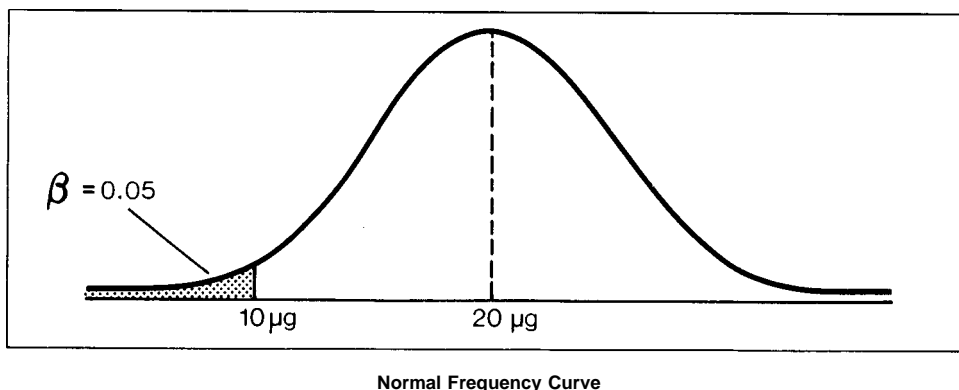


FIG. 3 Probability of Type II Error, True Value = Twice Criterion of Detection