Designation: D 4210 – 89 (Reapproved 1996)ε1

AMERICAN SOCIETY FOR TESTING AND MATERIALS
100 Barr Harbor Dr., West Conshohocken, PA 19428
Reprinted from the Annual Book of ASTM Standards. Copyright ASTM

Standard Practice for Intralaboratory Quality Control Procedures and a Discussion on Reporting Low-Level Data[1]

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.

ε[1] 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 Water[2]

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,

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

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

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

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

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

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

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

$$s^2 = 3.776$$

$$s = 1.943 \text{ µg/L}$$

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

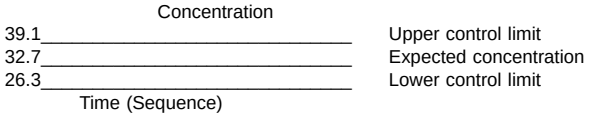
9. Setting Control Limits

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 µg/L and is analyzed by a procedure whose estimated standard deviation is 2.131 µg/L. The control limits are therefore 32.7 ± 3 × 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.

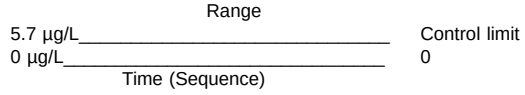
9.2.2 *Typical Control Chart for Standards*:



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 µg/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 ±3 × 0.1532 or − 0.46 mg/L and + 0.46 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 appropriateα (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 µg/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 µg/L)≃ 10 µg/L (see Fig. 1).

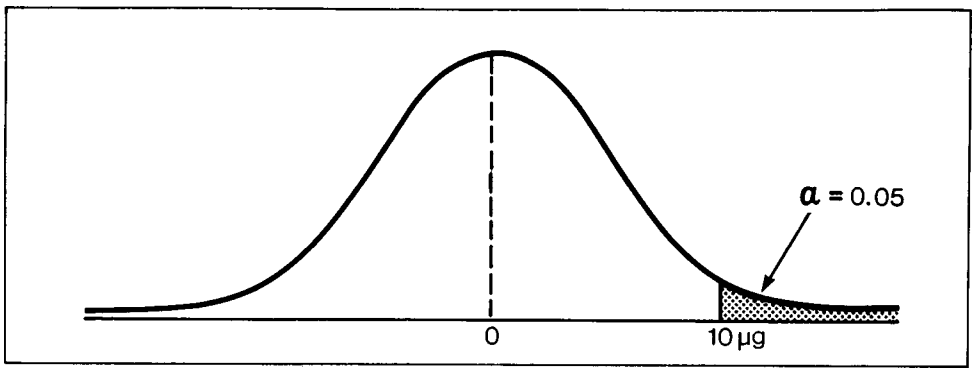
11.6 Any value observed below 10 µg/L would be reported as less than the criterion of detection, since to report such a value otherwise would increase the probability of making a Type I error beyond 5 %.

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-β, the probability of discerning the substance when it is present, can be determined for *given true situations.* (The probability 1-β 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 µg/L. Suppose that the concentration of the sample being analyzed is 10 µg/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

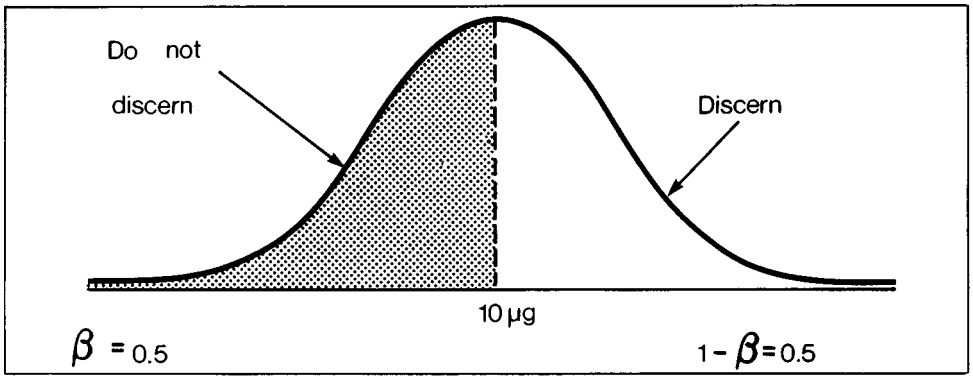


FIG. 2 Probability of Type II Error, True Value = Criterion of Detection

of a Type II error to be equal to the probability of a Type I error, β(beta) = α(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 α(alpha)-level, the probability of making a Type II error is increased.

11.11.1 Having clarified the conceptual context in which an α(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 σ = 6 µg cannot be regarded as significantly different from 0 for any α(alpha)-level less than 0.067; however, if a significance test were made with α(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 α(alpha) less than 0.067 and could be reported as simply "9 µg" for all α(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

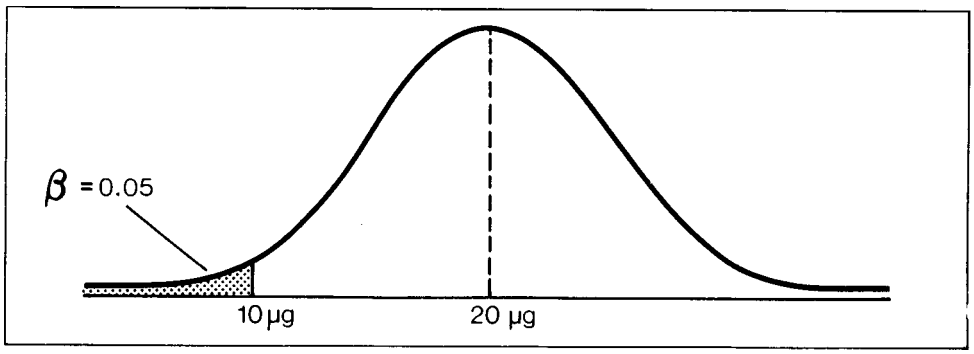


FIG. 3 Probability of Type II Error, True Value = Twice Criterion of Detection

increases by some function with the concentration.

12.3 The *W* code has the following meaning: "Value observed is less than lowest value reportable under *T* code." This code is used when a positive value is not observed or calculated for a result. In these cases the lowest reportable value, which is the lowest positive value which is observable, is reported with the *W*.

12.3.1 The following example illustrates the use of the codes: Suppose that a laboratory has determined that its criterion of detection for total phosphorus is 10 µg/L, and suppose in addition that the smallest increment that can be read on the analytical device corresponds to a concentration of 2 µg/L. Given these conditions, any value observed >10 µg/L would be reported without an accompanying code; any value observed >2 µg and <10 µg would be reported with the *T* code; if no instrument response were observed, the result would be reported as *W*, 2.

13. Reporting Negative Results

13.1 With many analytical procedures there will always be an instrument response, so the *W* code will not apply. In particular, this lack of applicability will occur when a result is obtained through subtraction of a blank value. In this case negative results will often be obtained; in fact, if the constituent of interest is not present, one would expect negative results to occur as often as positive.

13.2 In order that valid inferences may be made from data sets, it is important that negative results be reported as such. Consider the following three different ways of reporting the same results. The left hand column gives results in a heavily censored form; the center column has negative results censored; the right hand column gives the results as obtained.

<3 µg	2 µg	2 µg
<3	0	−2
<3	0	−1
4	4	4
3	3	3
<3	0	−3
<3	1	1
<3	0	−1
<3	0	0
<3	2	2

13.3 Nothing can be done with the results in the left hand column except to conclude that we don't know whether the constituent is present or not.

13.4 If the results in the center column were taken at face value, one could conclude that the mean concentration was 1.2 µg with a standard error of the mean of 0.467 and 95 % confidence limits for the mean of 0.14 µg and 2.26 µg. Since the confidence limits do not include zero, it would appear that the evidence supports the presence of the constituent.

13.5 Analysis of the uncensored results of the right hand column gives a mean concentration of 0.5 µg, a standard error of the mean of 0.719, and 95 % confidence limits for the mean of − 1.13 µg and 2.13 µg. The correct conclusion can be drawn that the evidence is insufficient to support the presence of the constituent.

13.6 Note that the censored data of the center column distort both the mean and the standard error of the data, making the data appear more precise than they are. Logically any result of 0 or less which is reported should be reported with the *T* code.

14. Keywords

14.1 estimating analytical variability; quality control; reporting low-level data

ANNEXES

(Mandatory Information)

A1. ESTIMATING ANALYTICAL PROCEDURE VARIABILITY BY DUPLICATE ANALYSES

A1.1 In using duplicates to estimate population standard deviation, an example is provided in Table A1.1. Consider the pairs of results, in micrograms per litre, on duplicates which were analysed in different runs.

A1.2 Two of the ranges obtained, 12 and 18, strongly suggest that the analytical system was out of control. The two extreme ranges may be tested by obtaining the average range, $\bar{R}$, for all duplicate pairs.

$$4 + 1 + 3 \ldots . + 3 + 0 = 131$$

$$\bar{R} = 131/50 = 2.62$$

A1.3 An estimate of the standard deviation, *s*, is obtained from the average range of duplicate analyses by dividing by 1.128, the proper factor for acquiring a standard deviation estimate from ranges derived from duplicates. [3]

$$s = \frac{2.62}{1.128} = 2.323 \text{ µg/L}$$

A1.4 Multiplying this standard deviation estimate by 3.686, the factor for the 3σ control limit for ranges from duplicates, gives $2.323 \times 3.686 = 8.56$. Since the extreme range, 18, is greater than 8.56, this range is discarded. Since the other extreme range, 12, is also greater than 8.56, it too is discarded. However, if the second extreme range had been 8 instead of 12, it would be necessary to perform a sequential recalculation with the set of 49 ranges to see if it too should be discarded.

A1.5 The remaining 48 ranges are now summed and the average range found

$$\bar{R} = 101/48 = 2.104$$

A1.6 Dividing, as before, by 1.128 gives the estimate of the standard deviation,

$$s = \frac{2.104}{1.128} 1.865 \text{ µg/L}$$

A1.7 The 3σ control limit for the range is now 1.865

TABLE A1.1 Estimating Analytical Procedure Variability by Duplicate Analyses

1st Result	*2nd Result*	*Range*	*1st Result*	*2nd Result*	*Range*
50	46	4	39	42	3
37	36	1	24	25	1
22	19	3	20	18	2
17	20	3	12	10	2
32	34	2	28	32	4
46	46	0	35	40	5
26	28	2	22	22	0
26	30	4	26	25	1
61	58	3	41	40	1
44	45	1	20	21	1
40	44	4	22	40	18
36	35	1	37	35	2
29	31	2	29	26	3
26	38	12	34	35	1
36	36	0	17	19	2
47	45	2	43	44	1
16	20	4	56	53	3
18	21	3	30	32	2
26	22	4	20	21	1
35	36	1	36	32	4
26	25	1	43	39	4
49	51	2	22	21	1
33	32	1	35	36	1
40	38	2	53	50	3
16	13	3	47	47	0

(3.686) = 6.874. Note that the remaining 48 ranges are all less than this limit so no further discarding is necessary.

A2. ESTIMATING ANALYTICAL PROCEDURE VARIABILITY BY MULTIPLE ANALYSES OF A STABLE STANDARD

A2.1 In using multiple analyses of a stable standard to estimate population standard deviation, an example is given in Table A2.1.

A2.1.1 The estimate of the standard deviation, *s*, is obtained in the usual way:

$$s^2 = \frac{\sum x^2 i - n\bar{x}^2}{n - 1}$$

$$s^2 = \frac{59\,540.6 - 50\,(34.368)^2}{49}$$

$$s^2 = 9.84957$$

$$s = 3.1384$$

A2.2 The two values 24.7 and 49.6 strongly suggest that the procedure was out of control. They are tested sequentially beginning with 49.6 since it is the farthest value from the mean.

TABLE A2.1 Estimating Analytical Procedure Variability by Multiple Analyses of Stable Standard

35.1	31.8	36.4	33.8	33.0	34.2
33.2	35.0	32.1	34.3	37.2	33.7
33.7	31.4	24.7	32.9	34.3	33.9
35.9	35.6	38.2	34.2	32.7	35.6
33.5	30.2	33.1	35.6	34.1	40.1
34.5	32.7	34.9	31.5	35.8	34.6
34.4	31.1	36.2	36.4	33.9	33.8
49.6	34.8	34.0	32.6	35.5	33.0
34.3	35.3				

FIG. A2.1 Mean of all values = 34.368.

A2.3 The absolute difference from the mean is, 49.6 − 34.368 = 15.232; this difference is greater than 3 times the estimated standard deviation, 3 (3.1384) = 9.415, so the value 49.6 is discarded.

A2.4 The new mean for the remaining 49 results is 34.05714 with an estimated standard deviation of 2.2633.

A2.5 The absolute difference between the revised mean and the second questionable result is, 34.05714 − 24.7 = 9.3514; this difference is greater than 3 times the revised estimated standard deviation, 3 (2.2633) = 6.79, so the value 24.7 is discarded.

A2.6 The new mean for the now remaining 48 results is 34.25208 with an estimated standard deviation of 1.8248. The 3σ control limits are now 34.25208 ± 3 (1.8248) or 28.8 and 39.7.

A2.7 On examining the remaining 48 results one finds another result, 40.1, which must be discarded since it is greater than 39.7. The process is reiterated once again with the remaining 47 results and gives a mean of 34.12766 and an estimated standard deviation of 1.6257. The new control limits 29.3 and 39.0 encompass the 47 values remaining in the data set so further reiteration is not necessary.

A2.8 While some analysts may prefer 2σ control limits, 3σ control limits were selected in this example since they are close

enough to signal when there is trouble with a system but distant enough to discourage tinkering with a system that is operating within its capabilities.

A2.9 Note that if the three omitted values had been included in the calculation, the estimated standard deviation would have been a badly inflated 3.138 µg/L. [4]

A2.10 It should be noted that *s* is expressed in absolute rather than relative terms. If variability were fully proportional to concentration, then the relative standard deviation (coefficient of variation) would be appropriate, but many analytical procedures are not so characterized. It appears that for any given practical working range variability may be treated as a constant with minimal ill effects. However, if very different ranges are employed to determine the same constituent an estimate of the standard deviation will be required for each range. One would not expect the variability that characterizes analyses in the range from 0 to 100 µg to also pertain to analyses in the range from 0 to 10 mg.

[4] Grant E. L., and Leavenworth, R. S. "Statistical Quality Control," 4th edition, McGraw-Hill Book Co., pp. 137–150.

A3. METHOD FOR TESTING CHANGE IN PROCEDURE VARIABILITY

A3.1 Suppose an initial estimate of an analytical procedure's standard deviation is obtained, $s_1 = 1.796$ µg/L, based on a data set of 61 items and therefore having associated with the estimate 60 degrees of freedom. A new estimate, $s_2 = 2.145$ µg/L, is then obtained based on 41 additional measurements, and thus having 40 degrees of freedom. The ratio of the two estimates of the *variance* is found as follows:

$$\frac{{s_1}^2}{{s_2}^2} = \frac{1.796^2}{2.145^2} = \frac{3.225616}{4.601025} = 0.701$$

and the ratio compared to appropriate values of the *F* distribution.

A3.2 Testing at an α(alpha)-level = 0.05, the appropriate upper value is simply the tabulated value for the upper 2.5 % point of the *F* distribution with 60 and 40 degrees of freedom; this tabulated value is 1.80. Obtaining the appropriate *lower* value requires a little arithmetic. The tabulated value for the upper 2.5 % point of the *F* distribution with 40 and 60 degrees of freedom (note the reversal) is found and its reciprocal taken, 1/1.74 = 0.575, to give the required value.

A3.3 Since the ratio of the two estimates of the analytical procedure variance, 0.701, lies between the values 0.575 and 1.80, one would *not* conclude that the variability of the procedure had changed.

A4. ESTIMATING ANALYTICAL PROCEDURE VARIABILITY BY USING SPIKE RECOVERIES

A4.1 Consider the following data set, values in milligrams per litre in Table A4.1.

A4.2 In column five you will note there are 3 deviations from expected recoveries which appear extreme: 1.19, 1.33 and − 0.97; these results are discarded. From the remaining 41 results in the 5th column of the data set an estimate of the standard deviation of the spiking recovery procedure is calculated in the usual way and found to be $s = 0.1532$ mg/L. (Since the deviations from expected results represent the difference between two analytical determinations, we would expect the standard deviation of the spiking recovery procedure to be greater than the standard deviation of a single determination by a factor of $\sqrt{2}$.)

A4.3 The mean of the deviations from the expected results is − 0.0061 mg/L. Since the absolute value of this mean is less than the standard error of the mean of the spiking recovery procedure, $s_m = 0.1532/\sqrt{41} = 0.024$ mg/L), the spiking recovery procedure appears to be unbiased with complete recovery a reasonable expectation. Control limits may therefore be set around the expectation of complete recovery with allowable deviations of $0 \pm 3 \times 0.1532$ or − 0.46 mg/L and 0.46 mg/L. The remaining 41 members of the data set are all within these limits.

A4.4 Had the spiking recovery procedure demonstrated a bias, the control limits would have been calculated from the estimate of the bias.

A4.5 In this example the data in column 6 may be used to obtain equivalent control limits in terms of percent recovery. With the omission of the three questionable results, the estimate of the standard deviation of the spiking recovery procedure is 11.782 % on a spike of 1.3 mg/L; 11.782 % of 1.3 mg/L is 0.1532 mg/L, which is the same estimate as obtained from column 5. However, the equivalency holds because identical spikes were employed in all recoveries. If variable spikes are used, then the estimate of the standard deviation and the ensuing control limits may have to be made in absolute units such as milligrams per litre rather than in percent recovery unless it is established that the characteristic percent recovery is similar for all spike levels.

TABLE A4.1 Estimating Analytical Procedure Variability by Using Spike Recoveries

1 Value for Spiked Sample	2 Value for Unspiked Sample	3 Calculated Recovery (1–2)	4 True Spike	5 Deviation From Expected (3–4)	6 % Recovery (3 / 4 × 100)
1.91	0.68	1.23	1.30	−0.07	94.615
1.78	0.57	1.21	1.30	−0.09	93.077
1.53	0.23	1.30	1.30	0	100
1.74	0.15	1.59	1.30	0.29	122.308
2.10	0.53	1.57	1.30	0.27	120.769
1.82	0.61	1.21	1.30	−0.09	93.077
2.07	0.54	1.53	1.30	0.23	117.692
1.39	0.14	1.25	1.30	−0.05	96.154
1.16	0.20	0.96	1.30	−0.34	73.846
1.55	0.19	1.36	1.30	0.06	104.615
2.02	0.41	1.61	1.30	0.31	123.846
1.58	0.36	1.22	1.30	−0.08	93.846
13.01	11.97	1.04	1.30	−0.26	80
1.46	0.17	1.29	1.30	−0.01	99.231
1.63	0.31	1.32	1.30	0.02	101.538
11.95	10.98	0.97	1.30	−0.33	74.615
1.68	0.27	1.41	1.30	0.11	108.462
1.83	0.47	1.36	1.30	0.06	104.615
1.62	0.43	1.19	1.30	−0.11	91.538
5.04	3.96	1.08	1.30	−0.22	83.077
2.53	1.22	1.31	1.30	0.01	100.769
2.69	1.09	1.60	1.30	0.3	123.077
1.50	0.25	1.25	1.30	−0.05	96.154
2.73	0.24	2.49	1.30	1.19	191.538
2.86	0.23	2.63	1.30	1.33	202.308
1.77	0.51	1.26	1.30	−0.04	96.923
1.88	0.55	1.33	1.30	0.03	102.308
0.90	0.57	0.33	1.30	−0.97	25.385
2.22	0.95	1.27	1.30	−0.03	97.692
1.99	0.85	1.14	1.30	0.16	87.692
1.54	0.26	1.28	1.30	−0.02	98.462
1.47	0.15	1.32	1.30	0.02	101.538
1.43	0.09	1.34	1.30	0.04	103.077
1.65	0.35	1.30	1.30	0	100
1.91	0.68	1.23	1.30	−0.07	94.615
2.06	0.93	1.13	1.30	−0.17	86.923
5.24	4.02	1.22	1.30	−0.08	93.846
1.58	0.27	1.31	1.30	0.01	100.769
1.63	0.28	1.35	1.30	0.05	103.846
1.52	0.23	1.29	1.30	−0.01	99.231
1.70	0.35	1.35	1.30	0.05	103.846
1.77	0.31	1.46	1.30	0.16	112.308
1.93	0.49	1.44	1.30	0.14	110.769
2.30	1.13	1.17	1.30	−0.13	90