



A11103 693412

NIST
PUBLICATIONS

NIST Special Publication 829

Use of NIST Standard Reference Materials for Decisions on Performance of Analytical Chemical Methods and Laboratories

CAC Quality Assurance Task Group:

D. Becker, R. Christensen,
L. Currie, B. Diamondstone, K. Eberhardt,
T. Gills, H. Hertz, G. Klouda, J. Moody,
R. Parris, R. Schaffer, E. Steel, J. Taylor,
R. Watters, and R. Zeisler

QC
100
U57
#829
1992
C.2

U.S. Department of Commerce
Technology Administration
National Institute of Standards and Technology

1998

[illegible]

Use of NIST Standard Reference Materials for Decisions on Performance of Analytical Chemical Methods and Laboratories

CAC Quality Assurance Task Group:

D. Becker, R. Christensen, L. Currie,
B. Diamondstone, K. Eberhardt, T. Gills,
H. Hertz, G. Klouda, J. Moody, R. Parris,
R. Schaffer, E. Steel, J. Taylor, R. Watters,
and R. Zeisler

Chemical Science and Technology Laboratory
National Institute of Standards and Technology
Gaithersburg, MD 20899

January 1992



U.S. Department of Commerce
Robert A. Mosbacher, Secretary
Technology Administration
Robert M. White, Under Secretary for Technology
National Institute of Standards and Technology
John W. Lyons, Director

National Institute of Standards
and Technology
Special Publication 829
Natl. Inst. Stand. Technol.
Spec. Publ. 829
30 pages (Jan. 1992)
CODEN: NSPUE2

U.S. Government Printing Office
Washington: 1992

For sale by the Superintendent
of Documents
U.S. Government Printing Office
Washington, DC 20402

USE OF NIST STANDARD REFERENCE MATERIALS FOR DECISIONS ON PERFORMANCE OF ANALYTICAL CHEMICAL METHODS AND LABORATORIES

Table of Contents

A.	INTRODUCTION	1
B.	APPROPRIATE SRM	1
C.	STATISTICAL CONTROL	1
D.	DETECTION OF BIAS	2
D1.	Planning the experiment: bias detection limit vs. number of replicates	2
D2.	Testing the hypothesis of "no bias": experimental results	6
D3.	The treatment of SRM uncertainty	7
E.	APPLICATIONS	11
E1.	Establishing statistical control of the analytical measurement process	12
E2.	Tests for analytical measurement bias	14
E3.	Tests for analytical measurement acceptability	17
E4.	Tests for single measurements	20
F.	REFERENCES	22
G.	APPENDIX	25
G1.	Percentiles of the t distribution	25
G2.	List of Symbols	26

USE OF NIST STANDARD REFERENCE MATERIALS FOR DECISIONS ON PERFORMANCE OF ANALYTICAL CHEMICAL METHODS AND LABORATORIES

A. INTRODUCTION

NIST Standard Reference Materials (SRMs) are used extensively for the evaluation of analytical methods and laboratory performance. The general principles of SRM use and the statistical tools useful for interpreting such measurements are discussed in NBS Special Publication 260-100 [1]. The present document describes specific guidelines and applications. Statistical guidance is essential when developing a protocol for a specific measurement program and for interpreting measurement data. The statistical methods outlined in this document and in reference [1] are more completely discussed with their related applications and assumptions in reference [2].

The following discussion is based on the use of an appropriate SRM to evaluate analytical methods and laboratory performance.

B. APPROPRIATE SRM

SRMs used for the applications that follow should meet the following requirements to the extent possible:

- 1) Reasonable matrix match with the samples customarily analyzed,
- 2) Reasonable match of analyte concentration(s),
- 3) Uncertainty of certified concentrations should be small with respect to the requisite uncertainty for the intended use.

It is virtually impossible for SRMs to exactly match the compositions of laboratory samples, and the SRM uncertainty may not be negligible in some applications. Accordingly, professional judgment and analytical expertise are needed in the selection of the most appropriate SRM. In most cases, some compromises will be inevitable. Notwithstanding these limitations, the use of SRMs is considered to be one of the best available approaches for decisions on the accuracy of measurement data. Specific directions for use, such as the amount of sample, the specimen treatment, and other analysis protocols are given on the SRM certificate. Unless these directions are followed explicitly, performance judgments based on the certified concentrations of the SRM may not be valid.

C. STATISTICAL CONTROL

The chemical measurement process (CMP) to be evaluated or used must be in a state of statistical control at the time any definitive measurements are made. That is, the measurement process must be stable and capable of producing a limiting mean (μ) and a fixed standard deviation (σ). While this can never be rigorously demonstrated, there should be no reasonable doubt that control has been achieved [3]. Furthermore, the standard deviation should be reasonably small with respect to any level of bias that is of concern. It is difficult to detect a bias smaller than the standard deviation, without imposing major demands on system long-term stability and requiring large numbers of

replicates. Also, it is impossible to achieve a bias detection limit smaller than twice the uncertainty of the standard reference material used (see sec. D3).

D. DETECTION OF BIAS*

If the uncertainty in the certified concentration of an SRM (for a given analyte) is negligible compared to the level of bias to be detected, the bias of a CMP may be estimated as the difference between the sample mean obtained by analyzing the SRM using the CMP and the SRM's certified concentration. Knowledge of the standard deviation of the CMP permits one to: (a) determine the relation between the number of replicates and the minimum detectable bias -- necessary for designing (planning) the experiment; and (b) test for bias and/or estimate a confidence interval for bias, given an experimental (mean) result. See reference [2] for a general discussion of these concepts.

D1. Planning the experiment: bias detection limit (Δ_D) vs. number of replicates

The objective of the first (planning) phase is to assure that the statistical (t) test for bias has adequate "power" to detect the level of bias considered important. That is, if the test for bias is made at the 5% significance level ($\alpha = 0.05$), we wish to have a 95% chance of detecting an absolute difference, Δ_D , between a measured mean and a "true" value (β = risk of false negative, and $1-\beta = 0.95$ = power of the test). It can be shown [4] that

$$\Delta_D \approx (t_{1-\alpha/2} + t_{1-\beta}) \sigma / \sqrt{n} \quad (1)$$

where $t_{1-\alpha/2}$ is the 2-sided Student's t , $t_{1-\beta}$ is the 1-sided Student's t , σ is the standard deviation, and n is the number of independent replicates. Equation (1) yields direct values for Δ_D or σ , given n ; but an iterative solution is generally required to calculate n , given Δ_D and σ , since the t 's depend on n . Alternatively, the adjustment described in the first footnote to table 1 can be used in lieu of iteration.

For large n or known σ , $t \rightarrow z$, the normal deviate, so $\Delta_D \rightarrow (1.960 + 1.645) \sigma / \sqrt{n}$ for $\alpha = \beta = 0.05$. In this case,

$$n \geq (3.605 \sigma / \Delta_D)^2 \approx 13(\sigma / \Delta_D)^2 = 13/d^2 \quad (2)$$

where $d = \Delta_D / \sigma$, the ratio of the bias to be detected to the measurement standard deviation.

Thus, at least 13 replicates are required to detect bias equal to the standard deviation. Stated differently, the standard error of the mean must be smaller than the bias to be detected by at least a factor of 3.6.

*See the Appendix for a list of symbols and a table of the t distribution.

D1a. Tabulated numbers of independent replicates (n)

Table 1, based on eq (1), has been prepared to facilitate estimation of the minimum number of replicate measurements necessary to detect a bias of a prescribed magnitude, for several values of $d = \Delta_D/\sigma$.

Table 1. Minimum Number of Replicates Required to Detect a Prescribed Difference^{1,2,3}

$d = \frac{\Delta_D}{\sigma}$	$\alpha = 0.10$ $1-\alpha/2 = 0.95$			$\alpha = 0.05$ $(1-\alpha/2 = 0.975)$			$\alpha = 0.01$ $(1-\alpha/2 = 0.995)$		
	$1 - \beta_0$			$1 - \beta_0$			$1 - \beta_0$		
	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
.5	35	44	64	43	52	74	60	72	97
.6	24	31	44	30	37	52	42	50	67
.7	18	23	33	22	27	38	31	37	50
.8	14	17	25	17	21	29	24	28	38
.9	11	14	20	13	17	23	19	22	30
1.0	9	11	16	11	13	19	15	18	25
1.2	6	8	11	8	10	13	11	13	17
1.4	5	6	9	6	7	10	8	10	13
1.6	4	5	7	5	6	8	6	7	10
1.8	3	4	5	4	5	6	5	6	8
2.0	3	3	4	3	4	5	4	5	7
2.5	2	2	3	2	3	3	3	3	4
3.0	1	2	2	2	2	3	2	2	3

¹If the standard deviation must be estimated from the data set, an approximate result is given by adding 2 to the tabulated numbers for $\alpha = 0.10$ or 0.05 or adding 4 to the tabulated numbers for $\alpha = 0.01$.

²The numbers of replicates in the table have been rounded upward. This means that $\beta \leq \beta_0$. Such rounding can lead to significant "integer distortion," especially when n is small, e.g., for $\alpha = \beta_0 = 0.05$ and $d = 1.8$, n (calculated from eq (2)) has been rounded up from 4.01 to 5.

³If the uncertainty of the SRM (U) is not negligible, replace d with $d' = (\Delta_D - 2U)/\sigma$. (See sec. D3.)

Table 1 may be used in the following manner:

- 1) Select acceptable α and β risks for an analytical decision.
- 2) Calculate $d = \Delta_D/\sigma$, the ratio of the bias to be detected to the standard deviation of the CMP.
- 3) Read the minimum number of required measurements from the table.

Note that the number in step 3 above should be increased by 2 to 4 if the standard deviation has not already been evaluated with a reasonable number of degrees of freedom, say ≥ 30 , but must be estimated from the data set used for bias detection. (The first footnote to table 1 indicates the appropriate increments to n .)

The values for Δ_D and σ may be absolute values or relative values (e.g., relative error and relative standard deviation), provided both are on the same basis.

Another use of the table is in the estimation of the risk of erroneous decisions on bias detection, based on a limited number of measurements. As an example, consider the case in which $d = 1.0$ and the feasible number of measurements is limited to $n = 10$. The closest combination of α and β risks found in the table is $\alpha = \beta = 0.10$. If the precision of the CMP could be improved by 20%, so that $d = 1.2$, the risks based on 10 replicates would be decreased to $\alpha = \beta = 0.05$.

From inspection of eq {2} and table 1, it is clear that the minimum number of replicates, n , varies as $1/d^2$. Taking the case for σ -known and $\alpha = \beta = 0.05$, we see that duplicate measurements suffice if $\Delta_D > 2.5 \sigma$, but as noted above, 13 measurements are required when $\Delta_D = \sigma$.

EXAMPLES

SRM 2704, Buffalo River Sediment, is used to test a method for the determination of Si. The Si certified concentration is 29.08 ± 0.13 wt% ($x_0 \pm U$). (Note that the SRM uncertainty U is ignored in these examples, and that α and β are each taken equal to 0.05.)

Planning the experiment (Δ_D , given n , and vice versa)

If the CMP to be evaluated has an imprecision (σ) of 2.5 wt% Si, what is the minimum detectable bias (Δ_D) for $\alpha = \beta = 0.05$ for a given number of replicates (n)?

1) σ -estimated, $n = 5$ ($df=5-1=4$):

$$\begin{aligned} \Delta_D &\cong \frac{(t_{0.975} + t_{0.95})\sigma}{\sqrt{n}} && \text{(Eq {1})} \\ &= \frac{(2.776 + 2.132)2.5}{\sqrt{5}} = 5.49 \text{ wt\% Si} \end{aligned}$$

2) σ -known, any n :

$$\Delta_D = \frac{(z_{0.975} + z_{0.95})\sigma}{\sqrt{n}} \quad (\text{Eq } \{1\})$$

$$= \frac{(1.960 + 1.645)2.5}{\sqrt{n}} = \frac{9.01}{\sqrt{n}}$$

Comparative results:

Minimum detectable bias, Δ_D
(wt% Si)

	<u>$n = 5$</u>	<u>$n = 25$</u>
σ -estimated	5.49	1.89
σ -known	4.03	1.80

3) Approximation, using table 1:

For $n = 25$ (σ -known), interpolation gives $d \approx 0.73$. Thus,

$$\Delta_D = \sigma d \approx (2.5) (0.73) = 1.8 \text{ wt\%}.$$

For $n = 25$ (σ -estimated), $n' = n-2$ must be used. Interpolating for 23 replicates gives $d \approx 0.77$. Thus,

$$\Delta_D \approx (2.5) (0.77) = 1.9 \text{ wt\%}.$$

How many replicates are required to detect a bias (Δ_D) of 5% of the certified concentration using a spectrometric method which has a $\sigma = 2.5$ wt%? ($\Delta_D = 0.05 \times 29.08 = 1.454$ wt%)

$$1) \quad \sigma\text{-estimated: } n \geq \left[\frac{(t_{0.975} + t_{0.95})\sigma}{\Delta_D} \right]^2 \quad (\text{Eq } \{1\} \text{ transformed})$$

$$\geq 41 \text{ (by iteration)}$$

Note that convergence is extremely rapid, since n must be an integer. In fact, the correct answer can generally be obtained by adding 1 or 2 replicates to the number calculated for the " σ -known" case. Iteration is necessary because Student's t depends on the number of replicates, n , through the degrees of freedom ($df = n - 1$).

$$2) \quad \sigma\text{-known:} \quad n \approx 13(\sigma/\Delta_D)^2 \quad (\text{Eq (2)})$$

$$\approx 13(2.5/1.454)^2 \approx 38.4$$

Rounding up, n should be 39, since replicates are discrete.

- 3) Approximate results may be obtained "by inspection" using table 1, and the value for $d = \Delta_D/\sigma = \frac{1.454}{2.5} = 0.582$. n (σ -known) thus lies between 52 ($d = 0.50$) and 37 ($d = 0.60$). Crude interpolation yields $n = 40$ (σ -known) and $n = 42$ (σ -estimated from the experiment).

D2. Testing the hypothesis of "no bias": experimental results

The objective of the second phase is to apply the t -test to the null hypothesis ("no bias"), given an experimental result: \bar{x} , s , n . The critical value (Δ_C) for testing for bias is

$$\Delta_C = t_{1-\alpha/2} s / \sqrt{n}. \quad (3)$$

The estimate of bias ($\hat{\Delta}$) is the difference between the observed mean concentration (\bar{x}) and the certified concentration of the analyte in the SRM (x_0).

$$\hat{\Delta} = \bar{x} - x_0 \quad (4)$$

If the absolute value of the estimated bias does not exceed Δ_C , the null hypothesis is not rejected; i.e., bias is not detected.* This does not mean that the CMP is unbiased, but that whatever bias might be present is "acceptable," provided that n is large enough to ensure that the test has adequate power (see above).

A complementary treatment of the experimental outcome is to compute a confidence interval (CI) for the bias. If the interval spans zero, any bias that is present is statistically insignificant $\frac{1}{n}$ another way of phrasing the t -test [5]. Thus,

$$CI = \hat{\Delta} \pm t_{1-\alpha/2} s / \sqrt{n}. \quad (5)$$

In a properly designed experiment, the confidence limits for an unbiased CMP are considerably smaller (in absolute value) than any bias of practical importance.

*The reader may have noticed that the critical value (Δ_C) tends to be smaller than the detection limit, Δ_D . This occurs because the detection limit is calculated to meet the stringent requirement that the estimated bias ($\hat{\Delta}$) has high probability ($1-\beta$) of exceeding the critical value (Δ_C) whenever the true bias is at the detection limit (Δ_D). True bias somewhat below the detection limit may also be detected (i.e., yield data with $\hat{\Delta}$ greater than the critical value) but will do so with correspondingly lower probability (lower than $1-\beta$).

EXAMPLES

Testing the hypothesis: experimental results

Suppose an experimental result of 27.32 wt% Si with a standard deviation of 2.64 wt% is obtained by a CMP for SRM 2704:

$$\bar{x} = 27.32 \text{ wt\%}$$

$$s = 2.64 \text{ wt\%}$$

$$\begin{aligned}\hat{\Delta} &= \bar{x} - x_0 \\ &= 27.32 - 29.08 = -1.76 \text{ wt\%}\end{aligned}\quad (\text{Eq (4)})$$

The critical value is

$$\begin{aligned}\Delta_C &= t_{0.975} s / \sqrt{n} \\ &(\text{if } \sigma \text{ is known, use } z_{0.975} \sigma / \sqrt{n})\end{aligned}\quad (\text{Eq (3)})$$

If there were 5 replicates, is bias detected?

$$\text{For } n = 5, t_{0.975} = 2.776: \Delta_C = 2.776(2.64/\sqrt{5}) = 3.28 \text{ wt\%}$$

$$|\hat{\Delta}| < \Delta_C, \text{ so bias is not detected at the 0.05 significance level.}$$

Bias Uncertainty Interval:

A confidence interval for the bias of CMP is given by:

$$\begin{aligned}CI &= \hat{\Delta} \pm t_{0.975} s / \sqrt{n} \\ &= -1.76 \pm 2.776(2.64/\sqrt{5}) \\ &= \text{from } -5.04 \text{ to } 1.52 \text{ wt\% (includes 0)}\end{aligned}\quad (\text{Eq (5)})$$

If n is increased to 25, Δ_C is reduced to 1.09 wt% Si (eq (3)). In this case, $|\hat{\Delta}| > \Delta_C$, so bias is detected. Its uncertainty interval comprises -2.85 to -0.67 wt% Si (eq (5)).

D3. The treatment of SRM uncertainty

Thus far we have assumed negligible uncertainty (U) for the certified concentration of the SRM. If that is not the case, we must take into account the magnitude of U in testing for bias and constructing confidence intervals for bias. Unfortunately, this cannot be done in a rigorous fashion unless the estimated value (x_0) for the SRM is truly a random variable $\frac{1}{n}$ i.e., a quantity derived strictly from random error $\frac{1}{n}$ and the SRM is recertified each time we wish to make the test. Generally, neither of these conditions is fulfilled: (1) the SRM is certified once, not each time a test for bias is made; (2) U frequently involves systematic

components in addition to random measurement error, and the method used for estimating and combining such components is not always the same, or even known to the user. (The one realistic case in which the SRM estimate may be treated as a random variable is when the primary source of error is within-sample heterogeneity for the SRM issued to the user. In this case, the actual concentration varies randomly each time the SRM is sampled.)

Two procedures may be followed: (1) treat the SRM interval as bounds for fixed (systematic) error [6]; (2) treat the SRM uncertainty interval as a random error confidence interval, and "propagate" the corresponding variance [7]. The first, which is discussed below, is clearly the more conservative approach and the one we recommend. The second approach is not discussed here.

To illustrate, let us assume that certificate information is given in the form

$x_0 \pm U$, where x_0 is the certified concentration, and U represents the assigned (symmetric) uncertainty bounds.

Note that x_0 is not necessarily the true value of the analyte concentration, but rather the best estimate of the true value based on measurement. Also, note that U need not be symmetric. This presentation will not, however, treat asymmetric limits, nor non-normal random error.

Taking the lower and upper bounds for the SRM uncertainty to be symmetric ($\pm U$), we now treat U as a fixed offset that increases both the detection limit for bias, and the uncertainty interval for experimentally estimated bias. The original expressions given in sections D1 and D2 are modified as follows.

$$\text{Bias Detection Limit: } \Delta_D \approx (t_{1-\alpha/2} + t_{1-\beta}) \sigma/\sqrt{n} + 2U \quad (6)$$

$$\text{Critical Value: } \Delta_C = t_{1-\alpha/2} s/\sqrt{n} + U \quad (7)$$

$$\begin{array}{ll} \text{Bias Uncertainty} & \\ \text{Interval:}^* & UI = \hat{\Delta} \pm (t_{1-\alpha/2} s/\sqrt{n} + U) \end{array} \quad (8)$$

Illustrations which take SRM uncertainty into account are given below. Even without numerical examples, however, it is useful to consider the limiting cases: When U is small compared to the standard error (σ/\sqrt{n}), it may be ignored and the above expressions revert to the earlier ones. When U is large compared to the standard error, both the critical value and bias uncertainty limit approach U ; and the bias detection limit approaches $2U$. For further information on the detection of bias, and the effect of systematic error on detection limits, see reference [8].

*The uncertainty interval (UI) is introduced here as a generalization of a statistical confidence interval (CI). This is necessary because the concept of a rigorous confidence level, $1-\alpha$, is inapplicable in the presence of non-statistical systematic error bounds. If the systematic error bounds are negligible or treated as random, UI and CI are identical.

EXAMPLES

The certified concentration for Si in SRM 2704, Buffalo River Sediment, is 29.08 ± 0.13 wt% ($x_0 \pm U$).

We shall assume that the SRM is used to evaluate the bias of a gravimetric method having a standard deviation for silicon of $\sigma = 0.20$ wt%. Thus the uncertainty interval for the SRM cannot be ignored. (As before, we shall take $\alpha = \beta = 0.05$.)

Planning the experiment

Given the imprecision (σ) of 0.20 wt% for the CMP to be evaluated, what is the minimum detectable bias (Δ_D) for a given number of replicates (n)?

- 1) σ -estimated, $n = 5$:

$$\begin{aligned}\Delta_D &\approx (t_{0.975} + t_{0.95}) \sigma / \sqrt{n} + 2U & (\text{Eq (6)}) \\ &= (2.776 + 2.132)(0.20/\sqrt{5}) + 2(0.13) = 0.699 \text{ wt\% Si}\end{aligned}$$

- 2) σ -known, any n :

$$\begin{aligned}\Delta_D &\approx (z_{0.975} + z_{0.95}) \sigma / \sqrt{n} + 2U & (\text{Eq (6)}) \\ &= (1.960 + 1.645)(0.20/\sqrt{n}) + 2(0.13) = 0.721/\sqrt{n} + 0.26\end{aligned}$$

Comparative results:

Minimum detectable bias, Δ_D
(wt% Si)

	<u>$n = 5$</u>	<u>$n = 25$</u>
σ -estimated	0.699	0.411
σ -known	0.582	0.404

- 3) Approximation, using table 1*

For $n = 25$ (σ -known), interpolation gives $d' \approx 0.73$. Thus,

$$\Delta_D = \sigma d' + 2U \approx (0.20)(0.73) + 2(0.13) = 0.406 \text{ wt\%}$$

For $n = 25$ (σ -estimated), $n' = n-2$ must be used. Interpolating for 23 replicates gives $d' \approx 0.77$. Thus,

*Note that allowance for systematic error bounds requires that d be replaced by $d' = (\Delta_D - 2U)/\sigma$ for use with table 1. This transformation follows directly from the defining eqs (1) and (6).

$$\Delta_D = \sigma d' + 2U \approx (0.20)(0.77) + 2(0.13) = 0.414\%$$

Thus, when as many as 25 replicates are available for estimating σ , the bias detection limit is nearly as small as for the case σ -known.

How many replicates are required to detect a bias of 0.22 wt%?

Bias Detection Limit:

$$\begin{aligned}\Delta_D &\approx (t_{0.975} + t_{0.95}) \sigma / \sqrt{n} + 2U \\ &\geq 2U = 0.26 \text{ wt}\%\end{aligned}\tag{Eq (6)}$$

The smallest achievable bias detection limit (Δ_D) is $2U$. Therefore, a bias detection limit of 0.22 wt% cannot be achieved regardless of the number of replicates. Corresponding to the minimum value for Δ_D of $2U = .26\%$, the minimum critical value, Δ_C , is $U = 0.13\%$ (see footnote on page 6).

How many replicates are required to detect a bias equivalent to 1.5% of the certified concentration, using the gravimetric method having a standard deviation (σ) of 0.20 wt% Si?

$$\Delta_D = 29.08 (0.015) = 0.4362 \text{ wt}\% \text{ Si}$$

1) σ -known:

$$\begin{aligned}n &\geq \left(\frac{(z_{0.975} + z_{0.95})\sigma}{\Delta_D - 2U} \right)^2 \\ &= \left(\frac{(1.960 + 1.645)0.20}{0.4362 - 2(0.13)} \right)^2 = 16.74\end{aligned}\tag{Eq (6) transformed}$$

Since replicates are discrete, the minimum value for n becomes 17.

2) σ -estimated:

$$\begin{aligned}n &\geq \left(\frac{(t_{0.975} + t_{0.95})\sigma}{\Delta_D - 2U} \right)^2 \\ &= \left(\frac{(t_{0.975} + t_{0.95})0.2}{0.4362 - 2(0.13)} \right)^2 = 19 \text{ (by iteration)}\end{aligned}\tag{Eq (6) transformed}$$

3) Approximation, using table 1:

$$\text{First calculate } d' = \frac{\Delta_D - 2U}{\sigma} = \frac{0.4362 - 2(0.13)}{0.20} = 0.881$$

n (σ -known) thus lies between 17 ($d'=0.9$) and 21 ($d'=0.8$). Crude interpolation yields $n = 18$ (σ -known) and $n = 20$ (σ -estimated).

Testing the hypothesis: experimental results

Experimental results examined for bias

Suppose that a more precise method of analysis were used to obtain the following results:

$$\bar{x} = 29.40 \text{ wt\%}$$

$$s = 0.17 \text{ wt\%}$$

$$\hat{\Delta} = \bar{x} - x_0 \quad (\text{Eq (4)})$$

$$\hat{\Delta} = 29.40 - 29.08 = 0.32 \text{ wt\% Si}$$

If there were 5 replicates, is bias detected?

$$n = 5: \Delta_C = t_{0.975} s / \sqrt{n} + U \quad (\text{Eq (7)})$$

$$= 2.776(0.17/\sqrt{5}) + 0.13 = 0.341 \text{ wt\% Si}$$

$$|\hat{\Delta}| < \Delta_C \text{ so bias is not detected at the 0.05 significance level.}$$

Bias Uncertainty Interval:

$$UI = \hat{\Delta} \pm (t_{0.975} s / \sqrt{n} + U) \quad (\text{Eq (8)})$$

$$= 0.32 \pm (2.776(0.17/\sqrt{5}) + 0.13)$$

$$= \text{from } -0.021 \text{ to } 0.661 \text{ wt\% Si}$$

If n is increased to 25, Δ_C is reduced to 0.200 wt% Si (eq (7)). In this case, $|\hat{\Delta}| > \Delta_C$, so bias is detected. Its uncertainty interval comprises 0.120 to 0.520 wt% Si (eq (8)).

E. APPLICATIONS

The most frequent applications of SRMs for the evaluation of measurement processes relate to the following questions:

- 1) Are the analytical results produced by a laboratory under statistical control? (E1)
- 2) Are the analytical results biased? (E2)
 - 2a) Is the method biased? (E2a)
 - 2b) When the method is known to be unbiased, are the results from a particular laboratory biased? (E2b)
 - 2c) When the method bias is not known, are the results from a particular laboratory biased? (E2c)
- 3) Are the analytical results acceptable, even if they exhibit some bias? (E3)
 - 3a) Is the mean result of a set of replicate measurements acceptable? (E3a)
 - 3b) Does an acceptable percentage of results fall within specified limits of a measurement program? (E3b)
- 4) Are the analytical results biased or unacceptable, based on tests with only a single measurement? (E4)

E1. Establishing statistical control of the analytical measurement process

It is recommended that SRMs be used in combination with control charts [1,3] for systematic monitoring of a measuring system for attainment and maintenance of statistical control. General guidance for this purpose is contained in sections 3.3 and 3.4 of reference [1]. When it is not feasible for a laboratory to use SRMs for this purpose on a regular basis (due to cost considerations, for example), a laboratory may use its own or other control samples on a routine basis, and on a less frequent basis, measure an SRM to verify the reliability of the control data obtained with other materials [3].

EXAMPLE

Over a period of 8 years, SRM 909, Human Serum, has been used at NIST as a control material to monitor the performance of a definitive isotope dilution method for measuring cholesterol in samples of human serum. Data from 15 sets (four measurements per set) have been plotted on control charts, figures 1 and 2, using methods described in reference [3].

The first control chart (fig. 1) is used to test the measurement process for stability of the mean, and is a plot of the difference of the mean of four measurements from the certified concentration for cholesterol in SRM 909. The upper and lower control limits (UCL, LCL) are "3-sigma" control limits calculated from the standard deviation of the differences of the means of the first 12 sets from the certified concentration. The upper

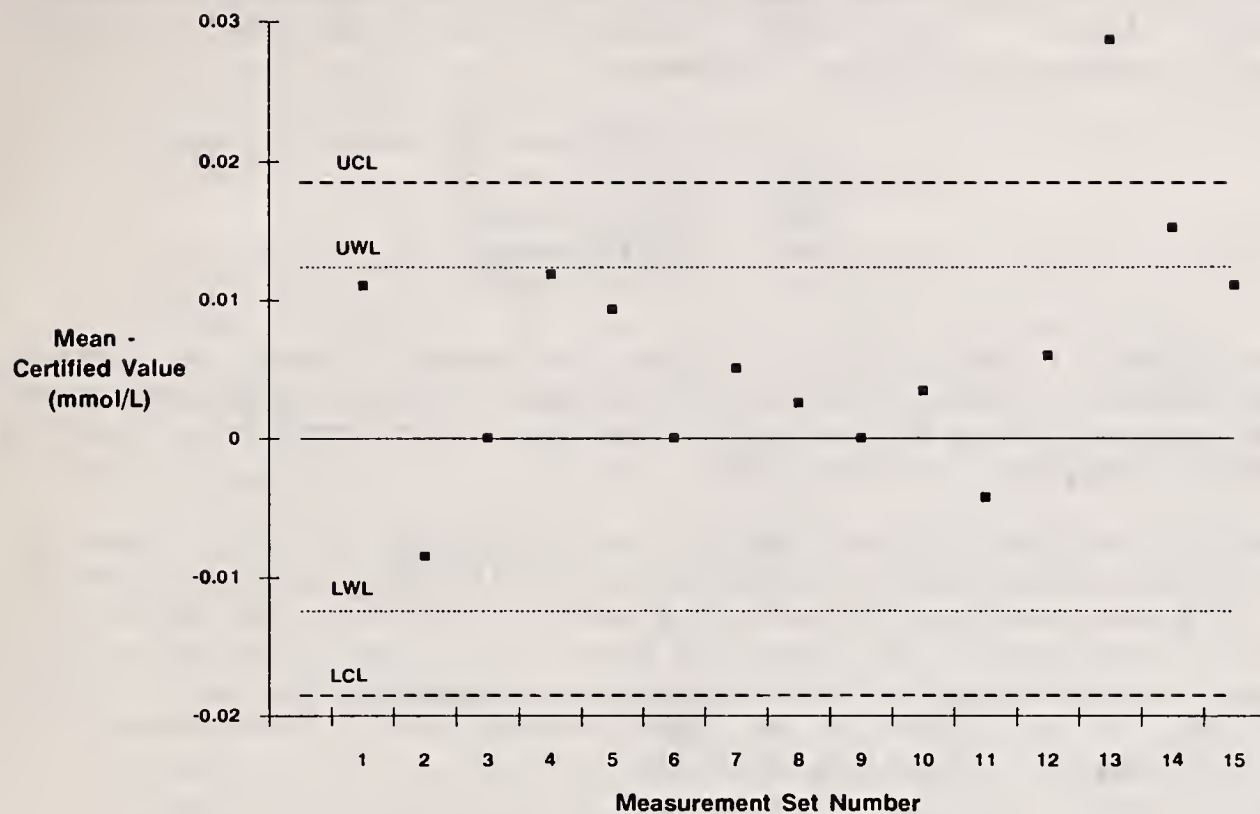


Figure 1. Control Charts to Test for Stability of the Mean

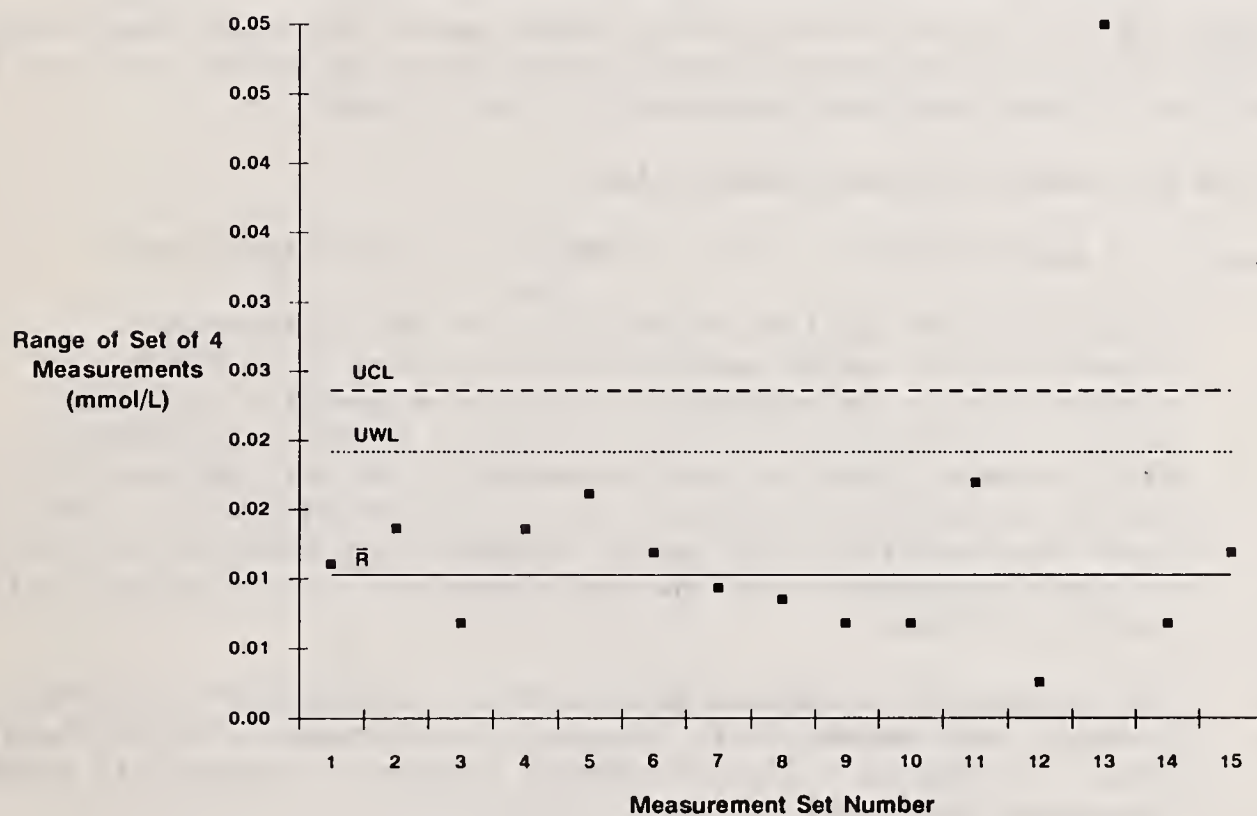


Figure 2. Control Chart to Test Precision of CMP

and lower warning limits (UWL, LWL) are the corresponding "2-sigma" control limits. The standard deviation of the means about the certified concentration is 0.0062. Therefore:

$$\begin{aligned}\sigma &= 0.0062 \text{ mmol/L} \\ \text{UCL} &= +0.0186 \text{ mmol/L} \\ \text{UWL} &= +0.0124 \text{ mmol/L} \\ \text{LWL} &= -0.0124 \text{ mmol/L} \\ \text{LCL} &= -0.0186 \text{ mmol/L}\end{aligned}$$

These control limits may be used to make judgments about the subsequent measurements. The result from Set 13 shows that the measurement process was not in control at that point, but the results from Sets 14 and 15 show that control has been recovered.

The second example is a range control chart (fig. 2). It is used to indicate whether the precision of the CMP is under control. Control limits were assigned, based on the mean of the ranges of the first 12 sets. The mean of the ranges (\bar{R}) is 0.0103. This is multiplied by factors appropriate for data with four measurements per set to arrive at an upper control limit and an upper warning limit. (See reference [2], p. 18-3 and reference [3], p. 139.)

$$\begin{aligned}\bar{R} &= 0.0103 \\ \text{Factor} = 2.282 & \quad \text{UCL} = 0.0235 \text{ mmol/L} \\ \text{Factor} = 1.855 & \quad \text{UWL} = 0.0191 \text{ mmol/L}\end{aligned}$$

Comparing the data in the subsequent measurement sets with these control limits shows that the precision of the CMP is out of control for Set 13, and that control has been recovered for Sets 14 and 15.

E2. Tests for analytical measurement bias

E2a. Is a method biased? -- Use of SRMs in collaborative tests

Application E2a applies to decisions on the performance characteristics of an analytical method when it is being standardized or is proposed for use for a specific analytical purpose. This case is commonly known as method validation [8] and it is commonly based on the evaluation of results obtained by a number of different laboratories in a collaborative test. SRMs are ideal test materials for use in collaborative tests of methods, due to their homogeneity and the confidence that can be placed in their certified values.

The concept of validation of methods by collaborative testing is probably well-known to all readers, so an example will not be given here. Procedures for collaborative testing are adequately described elsewhere [10,11,12].

E2b. Does a laboratory using an unbiased method produce unbiased data?

Application E2b applies to decisions that every laboratory should address whenever it uses methods for the first time or in a new application; namely, the demonstration of its ability to use a method that has been previously validated [9].

The procedure recommended consists of making a set of measurements and comparing the mean value with the certified value of the reference material. If the measured value does not differ significantly from the certified value, it may be concluded that the data are unbiased. Since the method used has been previously established to be unbiased at a given level of significance, any bias that is discovered can be attributed to laboratory performance. The laboratory then needs to investigate and correct the sources of its bias, if the method is to be used. Tests should be made at appropriate analyte levels within the range of measurement.

EXAMPLE

Aluminum was determined in SRM 1646, Estuarine Sediment, using Inductively Coupled Plasma Emission Spectrometry. Eight independent measurements were made, which resulted in a mean value of 5.86 wt% Al and a standard deviation of 0.30 wt%. The certified concentration and its uncertainty for Al in SRM 1646 are 6.25 ± 0.20 wt%.

$$\bar{x} = 5.86 \text{ wt\%}$$

$$s = 0.30 \text{ wt\%}$$

$$x_0 = 6.25 \text{ wt\%}$$

$$U = 0.20 \text{ wt\%}$$

$$\begin{aligned}\hat{\Delta} &= \bar{x} - x_0 && (\text{Eq (4)}) \\ &= 5.86 - 6.25 = -0.39 \text{ wt\% Al}\end{aligned}$$

$$\begin{aligned}\Delta_c &= t_{0.975} s/\sqrt{n} + U && (\text{Eq (7)}) \\ &= 2.365 (0.30/\sqrt{8}) + 0.20 = 0.45 \text{ wt\% Al}\end{aligned}$$

$|\hat{\Delta}| < \Delta_c$ so bias is not detected at the 0.05 significance level.

Bias Uncertainty Interval:

$$\begin{aligned} UI &= \hat{\Delta} \pm (t_{0.975} s / \sqrt{n} + U) & (\text{Eq (8)}) \\ &= -0.39 \pm (2.365 (0.30 / \sqrt{8}) + 0.20) \\ &= \text{from } -0.84 \text{ to } +0.06 \text{ wt\% Al} \end{aligned}$$

E2c. Does a laboratory using its own (unvalidated) method produce unbiased data?

This case applies to situations in which a laboratory utilizes methodology that has not been validated by others, and desires to know whether the data produced are unbiased. There are two possible sources of bias in the analytical results: (1) bias inherent in the method, and (2) bias resulting from the laboratory's use of the method.

The procedure described in Application E2a may be followed. If the observed difference between the mean of the experimental data and the certified concentration of the SRM is not significant, one may conclude, within the significance level of the statistical test, that the combination of both the method and the laboratory produces unbiased data. If bias is detected, one is uncertain whether this bias is due to source (1), (2), or a combination of the above causes. A research investigation will ordinarily be required to answer these questions. One may devise appropriate tests to systematically investigate contributions to bias from sources such as calibration problems, blank corrections, contamination, and losses [3]. Likewise, one may investigate the various steps in the method for their contributions to bias. Comparison of the test results with those obtained using a reference or definitive method is another way to evaluate bias [3], and does not involve use of an SRM.

EXAMPLE

Two gas chromatographic methods, which differed only in the internal standard used, were evaluated for the determination of selected polycyclic aromatic hydrocarbons (PAH) in traffic tunnel particulate material. The following means and standard deviations were obtained from six analyses of SRM 1650, Diesel Particulate Material, by each method, for the listed PAH.

SRM 1650 - Diesel Particulate Material

	Concentration, $\mu\text{g/g}$				Certified Concentration		
	Method A		Method B				
	\bar{x}	s	\bar{x}	s	x_0	\pm	U
Fluoranthene	56.6	7.2	65.2	7.3	51	± 4	
Pyrene	53.4	8.4	61.6	9.2	48	± 4	
Benz[a]anthracene	5.1	2.4	5.8	2.7	6.5	± 1.1	

For each PAH result, the absolute difference between the mean and the SRM certified concentration ($\hat{\Delta}$) was calculated. Then for $\alpha = 0.05$, the critical value (Δ_c) was calculated using eq (7) ($n = 6$, $t_{0.975} = 2.571$).

	Method A		Method B	
	$\hat{\Delta}$	Δ_c	$\hat{\Delta}$	Δ_c
Fluoranthene	5.6	11.6	14.2	11.7
Pyrene	5.4	12.8	13.6	13.7
Benz[a]anthracene	-1.4	3.6	-0.7	3.9

The determination of fluoranthene using Method B is biased at the significance level of the test. Because the methods differ only in the internal standards used, the choice of internal standard in Method B is considered to be inappropriate. For the other results, the absolute difference between the mean and the SRM certified concentration is less than the critical value, so bias is not detected. However, in the case of pyrene by Method B, the result is on the borderline, so further investigation may be warranted.

E3. Tests for analytical measurement acceptability

Application E3 applies to decisions related to a laboratory's own assessment of its analytical measurement capability and to the selection and validation of laboratories to be used in a specific measurement program. Application E3a deals with bias in a measurement process; therefore, the population (or limiting) mean of a laboratory's results should fall within the specified limits. In Application E3b, we are concerned with individual results from a laboratory's measurement process; therefore, most of the population, i.e., a specified percentage, of a laboratory's results should fall within the specified limits.

The difference between this situation and Application E2 is the concept of "acceptability." The user of the data decides what limits of error in the data are acceptable, based on practical considerations. Cost and benefits are prime considerations when deciding what limits are acceptable. These

limits are generally larger than the uncertainties of the SRM and of the method used and may often be considerably larger. This extra limit of error (A) is added to the uncertainty bounds of the SRM so that the overall acceptable range is $x_0 \pm (U+A)$. Equation (7) is then modified to yield the critical value for acceptability:

$$\Delta_c = t_{1-\alpha/2} s/\sqrt{n} + (U+A) \quad (9)$$

E3a. Does a laboratory using an unbiased or reference method produce a population mean result with acceptably small bias?

Application E3a deals with decisions on the acceptability of the population mean of a laboratory's results which were produced using an unbiased or reference method. The recommended procedure to be used is as follows:

- 1) Put specification limits around the certified concentration of the SRM to indicate the limits of bias considered to be acceptable. These limits should include the uncertainty assigned to the certified concentration. Calculate the critical value using eq (9).
- 2) Compare the difference between the analytical result and the certified value with the critical value. If $|\hat{\Delta}| < \Delta_c$, the laboratory is considered to be producing acceptable data.

Note that specification limits may be based on arbitrary decisions or on the statistics of group performance in a collaborative test. For example, they may represent the limits for values that 99% of the laboratories are expected to produce when using a method correctly.

EXAMPLE

A laboratory used SRM 1173, a Ni-Cr-Mo-V steel, as a quality control material to check the acceptability of results for verifying that an unknown steel sample was of a similar alloy type. The element determined was carbon, and the predetermined allowance for bias was $\pm 5\%$ of the true value for the SRM in addition to the uncertainty in the certified concentration. The certified concentration for carbon in SRM 1173 is 0.423 ± 0.004 wt%. The limits of acceptability were ± 0.021 wt%. The mean value of four measurements was 0.400 wt% with a standard deviation of 0.003 wt%.

$$\bar{x} = 0.400 \text{ wt\%}$$

$$s = 0.003 \text{ wt\%}$$

$$x_0 = 0.423 \text{ wt\%}$$

$$U = 0.004 \text{ wt\%}$$

$$A = 0.021 \text{ wt\%}$$

$$n = 4$$

$$\hat{\Delta} = \bar{x} - x_0 \quad (\text{Eq (4)})$$

$$= 0.400 - 0.423 = -0.023 \text{ wt\% C}$$

$$\Delta_c = t_{1-\alpha/2} s/\sqrt{n} + (U+A) \quad (\text{Eq (9)})$$

$$= 3.182(0.003/\sqrt{4}) + (0.004 + 0.021) = 0.030 \text{ wt\% C}$$

$|\hat{\Delta}| < \Delta_c$ so unacceptable results are not detected at the 0.05 significance level (performance is deemed acceptable).

E3b. Does an acceptable percentage of a laboratory's results fall within specified limits?

Application E3b applies to decisions on the acceptability of results from an unbiased method or a test method in an ongoing measurement program, for example. Since the standard deviation of the measurements is estimated from a limited sampling of the population of measurements, a tolerance interval, which allows for the coverage of a specified percentage of this population at a certain probability level, is computed.* The recommended procedure is:

- 1) Note the specification limits around the certified concentration of the SRM which have been established for the measurement program.
- 2) Compute the tolerance interval (TI) for the measurement process as the 2-sided tolerance interval for the population of measurement results where $\gamma = 0.90$, $P = 0.90$:

$$TI = \bar{x} \pm Ks \quad (10)$$

*The tolerance interval approach presented presumes single observations which are normally distributed. Tables are available for results which are means of replicates [13]. Alternative methods for evaluating CMP performance may be appropriate, such as: (a) "prediction intervals" (expected coverage tolerance intervals) [14], and (b) uncertainties comprising bounds (possibly asymmetric) for systematic and/or random error based on "expert judgment." If normality cannot be assumed, a distribution-free approach may be applied, but at least 50 observations are required [15].

In the above equation, \bar{x} and s denote the observed mean and standard deviation based on n independent samples (measurements), and K is the factor for the statistical tolerance interval corresponding to the above choices for γ and P , where γ is the probability that at least a proportion P of the population of results will be included in the interval. Values for K are obtained from statistical tables, such as in table A-6 of reference [2].

- 3) If the limits given by eq (10) lie within the specification limits, consider the CMP (for the given laboratory method) to be performing "acceptably." Note that this comparison takes into account possible bias and its uncertainty as well as random measurement error.

EXAMPLE

Assume that 10 measurements of methane in air (SRM 1658a), using a specified, validated GC-FID method yield a mean of 1.038 μmole methane per mole air and an observed standard deviation of 0.052. In a table for 2-sided normal tolerance intervals for $n = 10$, $\gamma = 0.90$, and $P = 0.90$, we find that $K = 2.535$. The tolerance interval estimated for this CMP is thus:

$$\bar{x} = 1.038 \mu\text{mol/mol}$$

$$s = 0.052 \mu\text{mol/mol}$$

$$TI = \bar{x} \pm Ks \quad (\text{Eq (10)})$$

$$= 1.038 \pm 2.535(0.052)$$

$$= \text{from } 0.906 \text{ to } 1.170 \mu\text{mol/mol}$$

If the specification limits around the certified concentration range from 0.900 to 1.100 $\mu\text{mol/mol}$, the performance of this CMP (laboratory method) must be judged "unacceptable," since the tolerance interval does not lie totally within the specification limits.

E4. Tests for single measurements

Application E4 applies to decisions regarding method validation or laboratory performance when the ongoing measurement protocol results in a single datum. In this case the standard deviation, s , must be established through experience with the measurement program which is often recorded using a control chart. For decisions concerning bias, eq (6) is used with the appropriate t value chosen for the number of degrees of freedom used to estimate s . The value for n is 1. For decisions concerning acceptability, eq (9) is used with the same provisions for n and df .

EXAMPLE

To illustrate, we return to the control chart example (see E1.), involving the measurement of cholesterol in SRM 909, Human Serum. Over a period of 8 years, 12 measurements of cholesterol in 1-gram samples of the control material were made. (For this example, we treat the mean value for a set of four determinations, as described in E1, as a single measurement result.) The standard deviation of a single measurement was 0.0062 mmol/L, and the mean of the 12 measurements differed from the certified concentration by an insignificant amount. The uncertainty of the certified concentration for 1 gram of material is ± 0.014 mmol/L. A thirteenth measurement differed from the certified concentration by 0.029 mmol/L. Does this measurement indicate bias?

$$s = 0.0062 \text{ mmol/L}$$

$$n = 1$$

$$t_{0.975} = 2.201 \text{ for 11 degrees of freedom}$$

$$U = 0.014 \text{ mmol/L}$$

$$\begin{aligned}\hat{\Delta} &= \bar{x} - x_0 \\ &= 0.029 \text{ mmol/L}\end{aligned}\tag{Eq (4)}$$

$$\begin{aligned}\Delta_c &= t_{0.975} s / \sqrt{n} + U \\ &= 2.201(0.0062) + 0.014 = 0.0276 \text{ mmol/L}\end{aligned}\tag{Eq (7)}$$

$|\hat{\Delta}| > \Delta_c$ so an unacceptable result is detected at the 0.05 significance level, and the thirteenth measurement is biased.

Decisions on measurement data such as those described above apply only to the measurement system and measurement situations tested. Any extension of the decisions to any other systems or situations will need to be justified. Because of the uncertainty of generalization, it is recommended that measurements made for validation of methods and qualification of laboratories should simulate the expected analytical conditions as closely as possible. When a variety of analytical conditions are to be expected (analyte levels and sample matrices), the entire range of expected conditions should be tested. This subject is discussed further in section 6.4 of reference [1].

F. REFERENCES

- [1] Taylor, J. K., Standard Reference Materials: Handbook for SRM Users; Natl. Bur. Stand. (U.S.) Spec. Publ. 260-100; U.S. Government Printing Office: Washington, DC, 1985.
- [2] Natrella, M. G., "Comparing materials or products with respect to average performance," In Experimental Statistics; Natl. Bur. Stand. (U.S.) Handbook 91; U.S. Government Printing Office: Washington, DC, 1963; Chapter 3, pp. 3-1 to 3-42. (See also the collection of statistical tables at the end of the Handbook.)
- [3] Taylor, J. K., Quality Assurance of Chemical Measurements; Lewis Publishers: Chelsea, MI, 1987; Chapter 14.
- [4] Currie, L. A., "Analyte and bias detection limits; approximations to the non-central t distribution," Natl. Inst. Stand. and Technol. Internal Report; Gaithersburg, MD, September 1990 (In preparation).
- [5] Natrella, M. G., "The relation between confidence intervals and tests of significance," In Statistical Concepts and Procedures, Precision Measurement and Calibration; Ku, H. H., Ed.; Natl. Bur. Stand. (U.S.) Spec. Publ. 300, Vol. 1; U.S. Government Printing Office: Washington, DC, 1969; pp. 388-391.
- [6] Eisenhart, C., "Realistic evaluation of precision and accuracy of instrument calibration systems," J. Res. Natl. Bur. Stand. (U.S.) 1963, 67C, 161.
- [7] Collé, R.; Karp, P., "Measurement uncertainties: Report of an international working group meeting," J. Res. Natl. Bur. Stand. (U.S.) 1987, 92, 243.
- [8] Currie, L. A., "Detection: Overview of historical, societal, and technical issues," In Detection in Analytical Chemistry: Importance, Theory, and Practice; Currie, L. A., Ed.; ACS Symposium Series 361; American Chemical Society: Washington, DC, 1988; Chapter 1.
- [9] Taylor, J. K., "Validation of analytical methods," Anal. Chem. 1983, 55, 600A-608A.
- [10] Taylor, J. K., "Role of collaborative and cooperative studies in the evaluation of analytical methods," J. Assoc. Off. Anal. Chem. 1986, 69, 398-400.
- [11] "Guidelines for collaborative study procedure to validate characteristics of a method of analysis," J. Assoc. Off. Anal. Chem. 1989, 72, 694-704.
- [12] ASTM E691, "Practice for conducting an interlaboratory test program to determine the precision of test methods," American Society for Testing and Materials: Philadelphia, PA 19103.

- [13] Eberhardt, K. E.; Mee, R. W.; Reeve, C. P., "Computing factors for exact two-sided tolerance limits for a normal distribution," Communications in Statistics 1989, 18, 397-413.
- [14] Whitmore, G. A., "Prediction limits for a univariate normal distribution," The American Statistician 1986, 40, 141-143.
- [15] Natrella, M. G., "Tables A-30 and A-31: Distribution-free tolerance limits," In Experimental Statistics; Natl. Bur. Stand. (U.S.) Handbook 91; U. S. Government Printing Office: Washington, DC, 1963; pp. T-75 and T-76.

G. APPENDIX

G1. Percentiles of the t distribution

(From reference [2], table A-4)

df	$t_{.60}$	$t_{.70}$	$t_{.80}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$
1	.325	.727	1.376	3.078	6.314	12.706	31.821	63.657
2	.289	.617	1.061	1.886	2.920	4.303	6.965	9.925
3	.277	.584	.978	1.638	2.353	3.182	4.541	5.841
4	.271	.569	.941	1.533	2.132	2.776	3.747	4.604
5	.267	.559	.920	1.476	2.015	2.571	3.365	4.032
6	.265	.553	.906	1.440	1.943	2.447	3.143	3.707
7	.263	.549	.896	1.415	1.895	2.365	2.998	3.499
8	.262	.546	.889	1.397	1.860	2.306	2.896	3.355
9	.261	.543	.883	1.383	1.833	2.262	2.821	3.250
10	.260	.542	.879	1.372	1.812	2.228	2.764	3.169
11	.260	.540	.876	1.363	1.796	2.201	2.718	3.106
12	.259	.539	.873	1.356	1.782	2.179	2.681	3.055
13	.259	.538	.870	1.350	1.771	2.160	2.650	3.012
14	.258	.537	.868	1.345	1.761	2.145	2.624	2.977
15	.258	.536	.866	1.341	1.753	2.131	2.602	2.947
16	.258	.535	.865	1.337	1.746	2.120	2.583	2.921
17	.257	.534	.863	1.333	1.740	2.110	2.567	2.898
18	.257	.534	.862	1.330	1.734	2.101	2.552	2.878
19	.257	.533	.861	1.328	1.729	2.093	2.539	2.861
20	.257	.533	.860	1.325	1.725	2.086	2.528	2.845
21	.257	.532	.859	1.323	1.721	2.080	2.518	2.831
22	.256	.532	.858	1.321	1.717	2.074	2.508	2.819
23	.256	.532	.858	1.319	1.714	2.069	2.500	2.807
24	.256	.531	.857	1.318	1.711	2.064	2.492	2.797
25	.256	.531	.856	1.316	1.708	2.060	2.485	2.787
26	.256	.531	.856	1.315	1.706	2.056	2.479	2.779
27	.256	.531	.855	1.314	1.703	2.052	2.473	2.771
28	.256	.530	.855	1.313	1.701	2.048	2.467	2.763
29	.256	.530	.854	1.311	1.699	2.045	2.462	2.756
30	.256	.530	.854	1.310	1.697	2.042	2.457	2.750
40	.255	.529	.851	1.303	1.684	2.021	2.423	2.704
60	.254	.527	.848	1.296	1.671	2.000	2.390	2.660
120	.254	.526	.845	1.289	1.658	1.980	2.358	2.617
∞	.253	.524	.842	1.282	1.645	1.960	2.326	2.576

Taken from M.G. Natrella, Experimental Statistics, NBS Handbook 91, U.S. Government Printing Office: Washington, DC, 1963. (Table A-4).

G2. List of Symbols

α	probability of incorrectly rejecting the tested (null) hypothesis. (for example, the probability of concluding a method to be biased when it is not biased)
A	"Acceptable" error limit from eq {9}
β	probability of incorrectly accepting the tested (null) hypothesis (for example, the probability of concluding a method not to be biased when it is biased)
β_0	limiting value for β , where $\beta_0 \geq \beta$. β_0 has been used in this paper because the values of the minimum number of replicates in table 1 were rounded upward to the nearest whole number. See footnote 2, page 3 for explanation.
γ	probability that at least a proportion, P, of the distribution will be included within the tolerance interval (TI)
CI	confidence interval
CL	$1-\alpha$, confidence level
CMP	Chemical Measurement Process
Δ_c	critical value from eqs {3} and {7} NOTE: Equations {3} and {7} are equal when U is 0.
Δ_D	bias detection limit from eqs {1} and {6} NOTE: Equations {1} and {6} are equal when U is 0.
$\hat{\Delta}$	estimate of bias, difference between observed mean and SRM certified value
d	Δ_D/σ = ratio of minimum detectable bias to the standard deviation of the CMP
d'	$(\Delta_D - 2U)/\sigma$ from eqs {1} and {6}
df	degrees of freedom
K	factor for two-sided tolerance limits (see table A-6 of reference [2])
LWL	Lower Warning Limit for control chart
LCL	Lower Control Limit for control chart
μ	population mean or limiting mean of a CMP
n	number of observations or analyses

P	minimum proportion of the distribution that will be included within the tolerance interval (TI) with a γ
\bar{R}	probability mean of ranges
σ	population standard deviation or limiting standard deviation of a CMP
s	estimated standard deviation
SRM	NIST Standard Reference Material
t	Student's "t" distribution
t_x	where $x = 0.90, 0.95, 0.975, 0.99, \dots$ = value from the table of percentiles of the t distribution in Appendix G1
TI	tolerance interval
\bar{x}	estimated mean
x_0	SRM certified value
U	SRM uncertainty
UCL	Upper Control Limit for control chart
UWL	Upper Warning Limit for control chart
UI	uncertainty interval, a generalization of CI, see footnote, page 8
z	the normal deviate (deviation from the population mean in units of σ)
z_p	where $p = 0.90, 0.95, 0.975, 0.99, \dots$ = percentile of the normal distribution. Use the values from Appendix G1 in the row in which $df = \infty$.

NIST-114A (REV. 3-89)		U.S. DEPARTMENT OF COMMERCE NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY		1. PUBLICATION OR REPORT NUMBER NIST/SP =829	
BIBLIOGRAPHIC DATA SHEET				2. PERFORMING ORGANIZATION REPORT NUMBER	
				3. PUBLICATION DATE January 1992	
4. TITLE AND SUBTITLE Use of NIST Standard Reference Materials for Decisions on Performance of Analytical Chemical Methods and Laboratories					
5. AUTHOR(S) CAC Quality Assurance Task Group: D. Becker, R. Christensen, L. Currie, B. Diamondstone, K. Eberhardt, T. Gills, H. Hertz, G. Klouda, J. Moody, R. Parris, R. Schaffer, E. Steel, J. Taylor, R. Watters, and R. Zeisler					
6. PERFORMING ORGANIZATION (IF JOINT OR OTHER THAN NIST, SEE INSTRUCTIONS) U.S. DEPARTMENT OF COMMERCE NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY GAITHERSBURG, MD 20899				7. CONTRACT/GRANT NUMBER	
				8. TYPE OF REPORT AND PERIOD COVERED Final	
9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS (STREET, CITY, STATE, ZIP) NIST CAC 222/A309 Gaithersburg, MD 20899					
10. SUPPLEMENTARY NOTES					
<input type="checkbox"/> DOCUMENT DESCRIBES A COMPUTER PROGRAM; SF-185, FIPS SOFTWARE SUMMARY, IS ATTACHED.					
11. ABSTRACT (A 200-WORD OR LESS FACTUAL SUMMARY OF MOST SIGNIFICANT INFORMATION. IF DOCUMENT INCLUDES A SIGNIFICANT BIBLIOGRAPHY OR LITERATURE SURVEY, MENTION IT HERE.) NIST Standard Reference Materials (SRMs) are used extensively for the evaluation of analytical methods and laboratory performance. It is virtually impossible for SRMs to exactly match the compositions of laboratory samples, and the SRM uncertainty may not be negligible in some applications. As a result, professional judgment and analytical expertise are needed in the selection of the most appropriate SRM. In most cases, some compromises will be inevitable. Notwithstanding these limitations, the use of SRMs is considered to be one of the best available approaches for decisions on the accuracy of measurement data. This document describes specific guidelines and applications for using SRMs to help design measurement protocols and to interpret the accuracy of measurement data.					
12. KEY WORDS (6 TO 12 ENTRIES; ALPHABETICAL ORDER; CAPITALIZE ONLY PROPER NAMES; AND SEPARATE KEY WORDS BY SEMICOLONS) Analytical chemistry; chemical measurement process; critical value; minimum detectable bias; SRMs; statistical control					
13. AVAILABILITY				14. NUMBER OF PRINTED PAGES	
<input checked="" type="checkbox"/> UNLIMITED FOR OFFICIAL DISTRIBUTION. DO NOT RELEASE TO NATIONAL TECHNICAL INFORMATION SERVICE (NTIS).				30	
<input checked="" type="checkbox"/> ORDER FROM SUPERINTENDENT OF DOCUMENTS, U.S. GOVERNMENT PRINTING OFFICE, WASHINGTON, DC 20402.				15. PRICE	
<input checked="" type="checkbox"/> ORDER FROM NATIONAL TECHNICAL INFORMATION SERVICE (NTIS), SPRINGFIELD, VA 22161.					

NIST Technical Publications

Periodical

Journal of Research of the National Institute of Standards and Technology—Reports NIST research and development in those disciplines of the physical and engineering sciences in which the Institute is active. These include physics, chemistry, engineering, mathematics, and computer sciences.

Papers cover a broad range of subjects, with major emphasis on measurement methodology and the basic technology underlying standardization. Also included from time to time are survey articles on topics closely related to the Institute's technical and scientific programs. Issued six times a year.

Nonperiodicals

Monographs—Major contributions to the technical literature on various subjects related to the Institute's scientific and technical activities.

Handbooks—Recommended codes of engineering and industrial practice (including safety codes) developed in cooperation with interested industries, professional organizations, and regulatory bodies.

Special Publications—Include proceedings of conferences sponsored by NIST, NIST annual reports, and other special publications appropriate to this grouping such as wall charts, pocket cards, and bibliographies.

Applied Mathematics Series—Mathematical tables, manuals, and studies of special interest to physicists, engineers, chemists, biologists, mathematicians, computer programmers, and others engaged in scientific and technical work.

National Standard Reference Data Series—Provides quantitative data on the physical and chemical properties of materials, compiled from the world's literature and critically evaluated. Developed under a worldwide program coordinated by NIST under the authority of the National Standard Data Act (Public Law 90-396). NOTE: The Journal of Physical and Chemical Reference Data (JPCRD) is published bimonthly for NIST by the American Chemical Society (ACS) and the American Institute of Physics (AIP). Subscriptions, reprints, and supplements are available from ACS, 1155 Sixteenth St., NW., Washington, DC 20056.

Building Science Series—Disseminates technical information developed at the Institute on building materials, components, systems, and whole structures. The series presents research results, test methods, and performance criteria related to the structural and environmental functions and the durability and safety characteristics of building elements and systems.

Technical Notes—Studies or reports which are complete in themselves but restrictive in their treatment of a subject. Analogous to monographs but not so comprehensive in scope or definitive in treatment of the subject area. Often serve as a vehicle for final reports of work performed at NIST under the sponsorship of other government agencies.

Voluntary Product Standards—Developed under procedures published by the Department of Commerce in Part 10, Title 15, of the Code of Federal Regulations. The standards establish nationally recognized requirements for products, and provide all concerned interests with a basis for common understanding of the characteristics of the products. NIST administers this program as a supplement to the activities of the private sector standardizing organizations.

Consumer Information Series—Practical information, based on NIST research and experience, covering areas of interest to the consumer. Easily understandable language and illustrations provide useful background knowledge for shopping in today's technological marketplace.

Order the above NIST publications from: Superintendent of Documents, Government Printing Office, Washington, DC 20402.

Order the following NIST publications—FIPS and NISTIRs—from the National Technical Information Service, Springfield, VA 22161.

Federal Information Processing Standards Publications (FIPS PUB)—Publications in this series collectively constitute the Federal Information Processing Standards Register. The Register serves as the official source of information in the Federal Government regarding standards issued by NIST pursuant to the Federal Property and Administrative Services Act of 1949 as amended, Public Law 89-306 (79 Stat. 1127), and as implemented by Executive Order 11717 (38 FR 12315, dated May 11, 1973) and Part 6 of Title 15 CFR (Code of Federal Regulations).

NIST Interagency Reports (NISTIR)—A special series of interim or final reports on work performed by NIST for outside sponsors (both government and non-government). In general, initial distribution is handled by the sponsor; public distribution is by the National Technical Information Service, Springfield, VA 22161, in paper copy or microfiche form.

U.S. Department of Commerce
National Institute of Standards and Technology
Gaithersburg, MD 20899

Official Business
Penalty for Private Use \$300