Statistical Protocol for the NIOSH Validation Tests

KENNETH A. BUSH and DAVID G. TAYLOR

National Institute for Occupational Safety and Health, Robert A. Taft Laboratories, 4676 Columbia Parkway, Cincinnati, OH 45226

Early in 1974, the National Institute for Occupational Safety and Health (NIOSH) and the Occupational Safety and Health Administration (OSHA) announced a joint program to complete the existing workroom level standards promulgated by the U.S. Department of Labor in 1972 (29 CFR 1910.1000). At that time, a statistical protocol was developed which has since been used for laboratory validation of over 300 sampling and analytical methods for monitoring employee exposure to the toxic substances in the OSHA regulations. The validations were conducted by Stanford Research Institute (now SRI International) under contracts CDC-99-74-45 and 210-76-0123 with NIOSH. The contractor set up laboratory facilities and air generation-dilution systems to validate methods over a concentration range from one-half to two times the permissible exposure limits (PEL) for the toxic substances shown in 29 CFR 1910.1000, Tables Z-1, Z-2, and Z-3. The OSHA PEL's are occupational health standards for personal exposure limits and may be either an 8-hour time-weighted average (TWA) concentration or a ceiling standard specified for a short time interval (generally 30 minutes or less).

The purpose of the validation program was to assure that accurate personal sampling and analytical methods would be available for use by OSHA in monitoring for non-compliance to the OSHA permissible exposure limits (PEL's). The methods are available to others who may want to use them to determine worker exposure to the substances in the OSHA regulations.

When a standardized sampling/analytical method is used to measure the concentration of a workplace air contaminant, it is certain that there will be some error in the result. But the exact amount of error in a given result is uncertain because quantitative errors occur as if they were random variables, i.e. in a chance manner, even when the method is used correctly. However, for a method which is "in control", what is predictable is the long-term proportion of individual errors which do not exceed a selected limit of error. The

probability that a given error will be less than some selected limit could be calculated if certain statistical parameters of the method, were known, namely its coefficient of variation (CV) and (any) bias. (The CV is referred to as the relative standard deviation by chemists. It is the ratio of the standard deviation of replicate concentration measurements to the mean concentration provided by the method.) Usually, an approximately normal distribution of errors can be assumed to exist as a basis for calculating such probabilities. The CV is assumed to be constant over the four-fold range of concentrations used in a given method's validation tests.

In this paper, we define an accuracy standard in terms of its two statistical parameters. However, in order to evaluate the accuracy of a particular method in terms of its statistical parameters, we have the problem that estimates of the method's statistical parameters are themselves subject to random sampling variations because the estimates must be calculated from only a finite number of replicate samples. The high cost of generating and analyzing large numbers of replicate samples necessitated using only enough samples to assure that reasonably accurate estimates were obtained of the CV and bias parameters of a method. Therefore, we also give statistical decision criteria by which test data for a method can be evaluated to determine whether there is reasonable confidence that the method meets the accuracy standard.

Several assumptions were made prior to initiating the actual validation of a given method:

1. The analytical method had to be previously developed and tested for items such as sample collection efficiency, recovery, and sample stability.
2. Both the air sampling and analytical method were to be validated.
3. An independent method was needed to verify the laboratory generation atmospheres used to validate the method.
4. The accuracy requirement developed for the methods had to apply to a single sample analysis, and not require an average of the analyses of several samples, because OSHA compliance determinations may be made on the basis of a single sample.
5. The bias determined in the validation referred to the difference between average results of the test method and average results of the independent reference method. However, it was recognized that other sources of bias, e.g. some interferences, may increase the true bias of the method in some unique field situations.

Choudhary; Chemical Hazards in the Workplace
ACS Symposium Series; American Chemical Society: Washington, DC, 1981.

NIOSH Accuracy Criterion. Accuracy is determined by both the precision and bias of the sampling and analytical method. Bias was defined under item 5 above as the difference between average results by the test method and average results by an independent reference method. Precision refers to the distribution of sizes of differences between results for replicate samples and the mean for the test method at that concentration. The accuracy criterion and its implications with respect to the worst precision and bias which are allowable are discussed below. The goal, however, is to assure that, in the long run, single measurements by the method will come within ±25% of corresponding "true" air concentrations at least 95% of the time. This accuracy requirement applies to a concentration range of 0.5 to 2.0 times the environmental PEL.

In the case of normally distributed sampling and analysis errors (and no bias) the above requirement implies that the true coefficient of variation of the total error (i.e. net precision error of sampling and analysis), denoted by CV_T, should be no greater than 0.128 derived as follows: $CV_T = 0.25/1.96 = 0.128$. The number 0.128 is the largest acceptable true CV_T for which the net error would not exceed ±25% at the 95% confidence level. The number 1.96 is the appropiate Z-statistic (from tables of the standard normal distribution) at the same confidence level.

If bias exists, the largest acceptable CV_T would have to be smaller than 0.128 in order for there to be less than 5% "large errors" (i.e. errors exceeding ±25%). In such cases, there would not be a 50-50 division of positive and negative large errors - rather, large errors in the direction of the bias would occur more often than 2.5% of the time. Large errors in the other direction would occur correspondingly less often, to keep the total occurrence in both directions at 5%.

The solid curve in Figure 1 shows the relationship between the bias and largest acceptable level of the true precision parameter (denoted in Figure 1 as the "target level" of the CV_T of a method). Note that when the bias is zero, the largest acceptable true CV_T is 0.128. Formulae are given in Appendix I for computing the solid curve giving the CV_T target level and bias combinations which meet the NIOSH accuracy standard.

The dotted curve of Figure 1 gives corresponding maximum permissible estimates of CV_T (designated $\hat{CV}_T$), based on laboratory tests performed under the experimental design described below. The shaded area indicates the acceptable $\hat{CV}_T$ region for validation of a method. The concept of making allowance for the sampling error in the precision estimate itself will be developed more fully below under Statistical Analysis Protocol. Basically, in the case of an

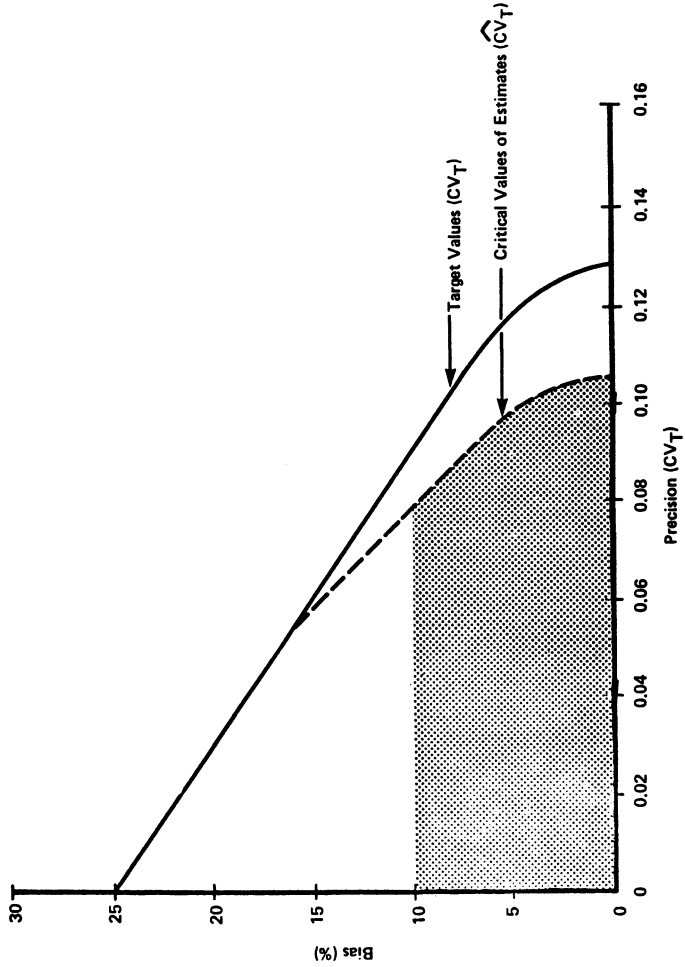

Figure 1. Combinations of CV_T and bias that meet the NIOSH accuracy standard

unbiased method, the estimate $\hat{CV}_T$ must be at or below 0.105 in order to be at least 95% confident that the true CV_T is at or below 0.128.

Statistical Experimental Design

Since the accuracy of an air concentration measurement is a function of both sampling and analysis, it is important to evaluate the method by testing both the sampling and analytical portions of the method. The validation program was designed to permit separate evaluation of the levels of error in these two parts of the method, as well as the total (net) error. All validation tests for a given method were carried out in a single laboratory, and although many of the methods had been used in the field previously, no field validations were undertaken.

Initially, the analytical method was tested to assure that it was acceptable for analyte recovery as well as for precision. The sampling medium was spiked with known amounts of the test chemical at three levels corresponding to one-half, one, and two times the occupational PEL for a given air volume. Six spiked samples for each level were analyzed. The success of this portion of the validation assured that the analytical precision was acceptable for the desired concentration range.

The second portion of the validation was to test the net precision due to both the sampling procedure and the analytic method used in sequence. This required the generation of known airborne concentrations of the toxic substance in a laboratory generator-dilution system. Three concentrations, at one-half, one, and two times the PEL, were prepared to test the sampling method. The generated concentrations were verified by a completely independent sampling and analytical method. For some substances this procedure was not possible and calculations based upon known flow and delivery rates, or on the experimentally determined collection efficiency, sample stability, and recovery were necessary to estimate the generated concentration. After selecting the recommended flow rate and sample volume (based upon the sampler capacity), the samples were collected from the laboratory generation-dilution system. Six samples at each of the three concentrations were collected using calibrated critical orifices. The data from these 18 samples, along with the 18 spiked sample results obtained in the analytical validation, were the basic statistical set of data used for the overall method validation. The data were used to determine the precision and bias of the method, which together determine its accuracy. The error of the personal sampling pump was not evaluated experimentally since sample flows in the laboratory tests were controlled by critical orifices in

Choudhary; Chemical Hazards in the Workplace
ACS Symposium Series; American Chemical Society: Washington, DC, 1981.

most cases. However, in the field, sampling pumps are used and their error was assumed to have a relative standard deviation of 0.05 (i.e. 5%) based on pump specifications.

Statistical Analysis Protocol

The purpose of the statistical analysis is to estimate the bias and the precision (measured by the CV_T of the total precision error of a subject method) and resolve the latter error into components CV_S due to the sampling method (less pump error), CV_A due to the analytical method (including error in the desorption efficiency factor), and CV_P (an assumed level of pump error). Appendix II gives the definitions and computational formulae for the statistical analysis.

Assuming normally distributed sampling and analysis errors (and no bias), the NIOSH accuracy standard is met if the true coefficient of variation of the total error, denoted by CV_T, is no greater than 0.128. However, estimates of CV_T (denoted by $\hat{CV}_T$), which were obtained in the laboratory validations, are themselves subject to appreciable random errors of estimation. Therefore, a "critical value" for the $\hat{CV}_T$ was needed (i.e. the value not to be exceeded by an experimental $\hat{CV}_T$ if the method is to be judged acceptable). The critical value of $\hat{CV}_T$ has to be lower than the maximum permissible true value (e.g. lower than $CV_T = 0.128$ when there is no bias). The maximum permissible value of the true CV_T will be referred to as its "target level". In order to have a confidence level of 95% that a subject method meets this required target level, on the basis of $\hat{CV}_T$ estimated from laboratory tests, an upper confidence limit for CV_T is calculated which must satisfy the following criterion: reject the method (i.e. decide it does not meet the accuracy standard) if the 95% upper confidence limit for CV_T exceeds the target level of CV_T. Otherwise, accept the method. This decision criterion was implemented in the form of the Decision Rule given below which is based on assumptions that errors are normally distributed and the method is unbiased. Biased methods are discussed further below.

For our validations, a $\hat{CV}_T$ is a pooled estimate calculated from the particular type of statistical data set (36 samples) described earlier in the Statistical Experimental Design section of this report. A statistical procedure is given in Hald(1) for determining an upper confidence limit for the coefficient of variation. This general theory had to be adapted appropriately for application to a pooled $\hat{CV}_T$ estimate. For this design, and under the stated assumptions, there is a one-to-one correspondence between values of $\hat{CV}_T$ and upper confidence limits for CV_T. Therefore, the confidence limit criterion given above is equivalent to another criterion based on the relationship of $\hat{CV}_T$ and its critical value. The

Choudhary; Chemical Hazards in the Workplace
ACS Symposium Series; American Chemical Society: Washington, DC, 1981.

Decision Rule is as follows:

> Decision Rule: The $\hat{CV}_T$ from lab tests would have to be less than the critical value 0.105 to be 95% confident that the true CV_T is at or below 0.128 (i.e., in order to be 95% confident that future errors by the same method would not exceed ±25% more than 5% of the time).

Figure 1 provides adjustments to critical values for $\hat{CV}_T$ when a method is biased. The dotted curve gives critical values of $\hat{CV}_T$ as a function of bias for a statistical significance test performed at the 5% probability level. Because uniform replicate determinations of the bias were not made in the validation tests, the bias is treated as a known constant rather than an estimated value. The experimental design could be modified to permit determination of the imprecision in the bias by providing for uniform replication of the independent method as well as the method under evaluation. Then the decision chart could be modified to include allowance for variability of replicate bias determinations. In cases where confidence limits can be calculated for the bias, the critical $\hat{CV}_T$ should be read from the dotted curve at a position corresponding to the 95% upper confidence limit for the bias. This is a conservative procedure.

The calculated points through which the curves of Figure 1 were drawn using a French curve are given below.

Bias (%)	Target CV_T(%)	Critical $\hat{CV}_T$(%)
0	12.8	10.5
2.5	12.5	10.3
5.0	11.8	9.8
10.0	9.1	7.9
15.0	6.1	5.8
16.8	5.0	5.0
20.0	3.0	(Unattainable)
25.0	0	(Unattainable)

Operating Characteristics of the Validation Test Program

As would be expected, in order to be able to have at least 95% confidence that the true CV_T does not exceed its target level, we must suffer the penalty of sometimes falsely accepting a "bad" method (i.e. one whose true CV_T is unsatisfactory). Such decision errors, referred to as "type-1 errors", occur randomly but have a controlled long-term frequency of less than 5% of the cases. (The 5% probability of type-1 error is by definition the complement of the confidence level.) The upper confidence limit on CV_T is below the target level when the method is judged acceptable under the Decision Rule.

Choudhary; Chemical Hazards in the Workplace
ACS Symposium Series; American Chemical Society: Washington, DC, 1981.

The validation test program can also have a "type-2 error", which is the mistake of deciding that a method is "bad" ($CV_T > 0.128$) when in fact it is "good" ($CV_T \leq 0.128$). The risk (probability) of making a type-2 decision error is not bounded (as is the case for the type-1 error). Rather, it depends on the true CV_T. In a previous report(2), it was shown that the probability of a type-2 error is large (0.88) for a "borderline" true CV_T (just below 0.128) but decreases to small probabilities of 0.10 for $CV_T = 0.091$, and 0.05 for $CV_T = 0.088$. Thus, more than 95% of methods whose CV_T's are below 0.088 (8.8%) will be accepted on the basis of their test results. "Good" methods whose true CV_T's are in the range 8.8% to 12.8% run a higher risk of not being approved; this risk could be lowered by using more than the now-prescribed 3 sets of 6 samples for the $\hat{CV}_T$ laboratory estimates in (each phase of) this program. However, the rate of improvement, in the precision of the laboratory estimates $\hat{CV}_T$, from using more samples would be small. For example, using 45 samples (15 per each of 3 groups) for each of the two phases instead of 18 (6 per group) only increases the "safe approval level" (0.05 probability of type-2 error) for CV_T from 0.088 (18 samples) to 0.099 (45 samples). The decision was made, therefore, to perform the smaller number (18) of tests for each of the two phases of the program.

Results of Validation Tests

Over 300 methods have been validated using the statistical protocol described above. Histograms have been prepared showing the distributions of precisions and biases obtained in the validation tests. Of 310 methods validated, only 31 (10%) had precision estimates ($\hat{CV}_T$'s) above 9% (See Figure 2). Apparently, only a small number of "good" methods have been tested whose CV_T's are in the borderline range where there is an appreciable chance of rejecting "good" methods. Since the pump error has a CV_P of 5% by itself, no values of $\hat{CV}_T$ fall below this level except for a few cases for which the method does not involve use of a personal sampling pump. It should be noted also that most of the methods have precisions clustering around 6-7% indicating the high quality of analytical methods tested.

The distribution of estimated biases for these methods is shown in Figure 3. Except for a bias of zero, the methods tend to be distributed evenly in the -10% to 10% bias region. The high proportion of zero-bias methods may be explained by the number of filter collection methods which have 100% collection efficiency; many of these methods use low-biased analysis techniques, particularly atomic absorption spectroscopy.

Choudhary; Chemical Hazards in the Workplace
ACS Symposium Series; American Chemical Society: Washington, DC, 1981.

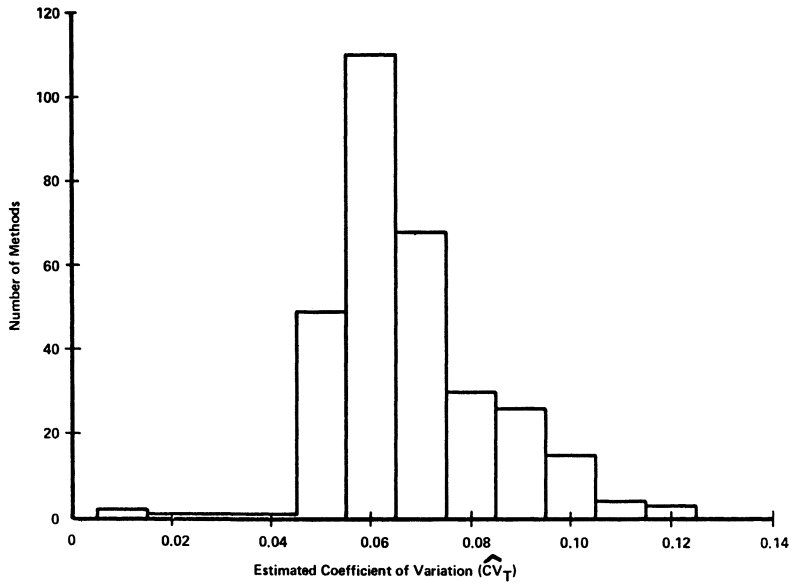

Figure 2. Histogram of CV_T (estimated coefficient of variation of net error attributable to sampling and analysis) for 310 methods

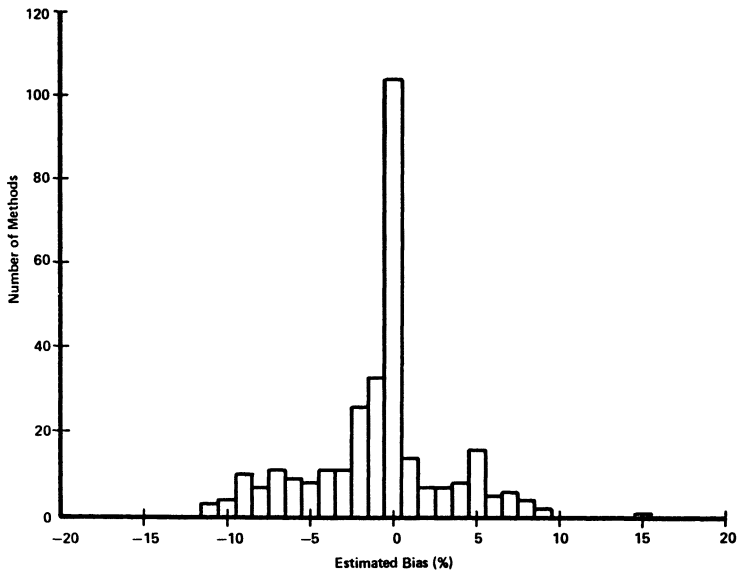

Figure 3. Estimated biases for 310 test methods

Summary

We have presented a statistical experimental design and a protocol to use in evaluating laboratory data to determine whether the sampling and analytical method tested meets a defined accuracy criterion. The accuracy is defined relative to a single measurement from the test method rather than for a mean of several replicate test results. Accuracy here is the difference between the test result and the "true" value, and thus, must combine the two sources of measurement error: 1) the random errors of the sampling and analysis (i.e. precision) represented by the total coefficient of variation (CV_T) of replicate measurements around their own mean and, 2) the error due to a real bias (systematic error) represented by the difference between average results by the subject collection-and-measurement method and average results from an independent method. The American Society for Testing and Materials, in their accuracy standard(3) states that accuracy does include both of these errors (Section 4.1). We have estimated both types of errors and referred results to a decision chart (Figure 1) to see if the test method does or does not meet the accuracy criterion.

Finally, we would like to point out that the statistical protocol for validation deals mainly with the last step in determining the validity of a monitoring method. The statistical protocol is not appropriate for application to a method that has not been completely developed. Tests for such items as sample collection efficiency, stability, and recovery; sampler capacity; and analytical range and calibration all should be evaluated prior to application of the statistical protocol in connection with laboratory validation testing.

Literature Cited

(1) Hald, A., "Statistical Theory with Engineering Applications", Chapter 11: part 11.8 and 11.9; Wiley, 1952.

(2) Busch, K. A., "Statistical Properties of the SRI Contract Protocol (CDC 99-74-45) for Estimation of Total Errors of Air Sampling/Analysis Procedures", memorandum to Deputy Director, Division of Laboratories and Criteria Development, Jan. 6, 1975.

(3) "Standard Recommended Practice for Use of the Terms Precision and Accuracy as Applied to Measurement of Property of a Material", E 177-71, in Annual Book of Standards, part 41, American Society for testing and Materials: Philadelphia, Pa., 1976.

APPENDIX I

TARGET VALUE OF CV_T FOR A BIASED METHOD

The maximum permissible CV_T (target value) for a biased method can be found by means of the formulae given below.

Let B = Bias ratio for the method

= (mean result by the method)÷(true concentration).

Standard normal deviates for left and right sides of the normal distribution corresponding to large errors (errors beyond ±25%) are given by:

$$Z_L = \frac{0.75-B}{B \cdot CV_T} \quad \text{and} \quad Z_R = \frac{1.25-B}{B \cdot CV_T}$$

For a given B, CV_T is the solution of the equation:

$$\int_{-\infty}^{Z_L} \frac{1}{\sqrt{2\pi}} e^{-(1/2)Z^2} dZ + \int_{Z_R}^{\infty} \frac{1}{\sqrt{2\pi}} e^{(-1/2)Z^2} dZ = 0.05$$

The equation must be solved iteratively. For any selected B, CV_T's are selected by trial and error in order to find the value of CV_T for which the sum of the integrals equals 0.05.

Example: $B = 1.1$, $Z_L = \frac{-0.35}{CV_T}$, $Z_R = \frac{0.15}{CV_T}$

For $CV_T = 0.09116$, $Z_L = -3.8394$, $Z_R = 1.6455$, and the sum of integrals is 0.0001 + 0.0499 = 0.0500. Thus a method with B = 1.1 (i.e. 10% bias) has $CV_T = 0.091$ as its target level.

Choudhary; Chemical Hazards in the Workplace
ACS Symposium Series; American Chemical Society: Washington, DC, 1981.

APPENDIX II

COMPUTATIONAL FORMULAE FOR STATISTICAL ANALYSIS

This appendix gives the formulae and definitions used in the protocol to statistically analyze laboratory data from validation tests.

Definitions and symbols are listed below:

Mean - arithmetic mean or average ($\bar{x}$), defined as the sum of the observations divided by the number of observations (n).

Standard Deviation - the positive square root of the variance, which in turn is defined as the sum of squares of the deviations of the observations from their mean ($\bar{x}$) divided by one less than the number of observations (n - 1).

$$\text{Std Dev} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

CV - coefficient of variation, or relative standard deviation, defined as the standard deviation divided by the mean.

$$CV = \frac{\text{Std Dev}}{\text{Mean}}$$

CV_1 - coefficient of variation (estimated value) for the six analytical samples at each of the 0.5, 1, and 2X OSHA PEL's for the recommended sample volume.

CV_2 - coefficient of variation (estimated value) for the six generated samples at each of the 0.5, 1, and 2X OSHA PEL's.

$\overline{CV}$ - pooled coefficient of variation: the value derived from the coefficients of variation (of a given type, e.g. CV_1 or CV_2) obtained from the analysis of 6 samples at each of the three test levels. The mathematical equation is expressed as:

Choudhary; Chemical Hazards in the Workplace
ACS Symposium Series; American Chemical Society: Washington, DC, 1981.

$$\overline{CV} = \sqrt{\frac{\sum_{i=1}^{3} f_i (CV_i)^2}{f}}$$

where:

f_i = degrees of freedom, equal to number of observations minus one ($n_i - 1$), at the i^{th} level.

CV_i = coefficient of variation (CV_1 or CV_2) of the observations at the i^{th} concentration level.

$$f = \sum_{i=1}^{3} f_i$$

i = index for the 3 concentration levels.

$\overline{CV}_1$ - pooled coefficient of variation calculated as above based on data for the 18 analytical (spiked) samples (3 groups of 6).

$\overline{CV}_{A+\overline{DE}}$ - derived correction to $\overline{CV}_1$ including precision error due to the use of the desorption efficiency factor, which is an average of 6 values.

$$\overline{CV}_{A+\overline{DE}} = \overline{CV}_1 \sqrt{7/6} = 1.0801\ \overline{CV}_1$$

$CV_{A+\overline{AMR}}$ - corrected $\overline{CV}_1$ analogous to use of a desorption efficiency factor noted above except that this notation is used where the factor is associated with analytical method recovery (AMR) other than for solid sorbents.

$$\overline{CV}_{A+\overline{AMR}} = 1.0801\ \overline{CV}_1$$

$\overline{CV}_2$ - pooled coefficient of variation based on the data for the 18 generated samples (3 groups of 6).

$\overline{CV}_S$ - coefficient of variation of the sample collection, not including the variability of the personal sampling pump. The value is dependent on the data from the 18 analytical and 18 generated samples.

$$\overline{CV}_S = \sqrt{(\overline{CV}_2)^2 - (\overline{CV}_1)^2}$$ (See "Note" below)

CV_P - coefficient of variation due to the pump error, assumed to be equal to 0.05.

$\hat{CV}_T$ - coefficient of variation of total procedure which consists of the composite variations in sampling and analysis, desorption efficiency, and the pump error.

$$\hat{CV}_T = \sqrt{(\overline{CV}_S)^2 + (\overline{CV}_{A+\overline{DE}})^2 + (CV_P)^2}$$

or:

$$\hat{CV}_T = \sqrt{(\overline{CV}_2)^2 - (\overline{CV}_1)^2 + 1.1667\ (\overline{CV}_1)^2 + (0.05)^2}$$

or:

$$\hat{CV}_T = \sqrt{(\overline{CV}_2)^2 + 0.1667\ \overline{CV}_1{}^2 + (0.05)^2} \quad \text{(See Note)}$$

NOTE: In case $\overline{CV}_2 < \overline{CV}_1$, take $\overline{CV}_S = 0$. Then replace $\overline{CV}_1$ by a pooled estimate ($\overline{CV}_1$*) based on $\overline{CV}_1$ and $\overline{CV}_2$,

$$\overline{CV}_1{}^* = \sqrt{\frac{f_1\ \overline{CV}_1{}^2 + f_2\ \overline{CV}_2{}^2}{f_1 + f_2}}$$

where f_1 and f_2 are the respective f-values used in the denominators of $CV_1{}^2$ and $CV_2{}^2$. Thus the equation to be used when $\overline{CV}_2 < \overline{CV}_1$ is:

$$\hat{CV}_T = \sqrt{1.1667(\overline{CV}_1{}^*)^2 + (0.05)^2}$$

GRUBB'S TEST for rejection of an observation is applied in order to determine if one of the observations should be rejected as being an outlier. The following equation was used for the test:

$$B_1{}' = \frac{x - \overline{x}}{s} \quad \text{or} \quad \frac{\overline{x} - x}{s}$$

where:

x = observation being tested (most distant from the mean)

$\overline{x}$ = mean of n observations

s = standard deviation based on n-1 degrees of freedom.

Choudhary; Chemical Hazards in the Workplace
ACS Symposium Series; American Chemical Society: Washington, DC, 1981.

For any 6 observations, a value can be rejected if $B_1' \geq 1.944$. The B_1' limit is based on a 1% significance level (i.e., a B_1' value calculated from the data can be expected to exceed 1.944 only 1% of the time if the observation is a legitimate one conforming to the underlying theory). For validation testing reject no more than two values in a set of 18 results and the two may not be in any one group of 6 replicates.

BARTLETT'S TEST for homogeneity of CV's is applied in order to test the feasibility of "pooling the coefficients of variation" for any set of 18 generated samples (i.e., 6 at each of the 0.5, 1, and 2X OSHA standard levels). The following equation for the Chi-square, with 2 degrees of freedom, was used:

$$\text{Chi-square} = \frac{f \ln (\overline{CV}_2)^2 - \sum_{i=1}^{3} f_i \ln (\overline{CV}_{2i})^2}{1 + \frac{1}{3(3-1)}\left[\left(\sum_{i=1}^{3} \frac{1}{f_i}\right) - \frac{1}{f}\right]}$$

where:

$\overline{CV}_2$ = pooled coefficient of variation of 18 generated samples

$\overline{CV}_{2i}$ = coefficient of variation of 6 generated samples at the i^{th} level

f_i = degrees of freedom associated with $(CV_{2i})^2$ and equal to number of observations at the i^{th} level minus one.

$$f = \sum_{i=1}^{3} f_1$$

In order to pass Bartlett's test at the 1% significance level, chi-square must be less than or equal to 9.21 (chi-square has 2 degrees of freedom).

RECEIVED October 21, 1980.