Designation: D 6091 – 03

An American National Standard

Standard Practice for 99 %/95 % Interlaboratory Detection Estimate (IDE) for Analytical Methods with Negligible Calibration Error[1]

This standard is issued under the fixed designation D 6091; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 This practice establishes a standard for computing a 99 %/95 % Interlaboratory Detection Estimate (IDE) and provides guidance concerning the appropriate use and application.

1.2 The IDE is computed to be the lowest concentration at which there is 90 % confidence that a single measurement from a laboratory selected from the population of qualified laboratories represented in an interlaboratory study will have a true detection probability of at least 95 % and a true nondetection probability of at least 99 % (when measuring a blank sample).

1.3 The fundamental assumption of the collaborative study is that the media tested, the concentrations tested, and the protocol followed in the study provide a representative and fair evaluation of the scope and applicability of the test method as written. When properly applied, the IDE procedure ensures that the 99 %/95 % IDE has the following properties:

1.3.1 *Routinely Achievable IDE Value*—Most laboratories are able to attain the IDE detection performance in routine analyses, using a standard measurement system, at reasonable cost. This property is needed for a detection limit to be practically feasible. Representative laboratories must be included in the data to calculate the IDE.

1.3.2 *Routine Sources of Error Accounted for*—The IDE should realistically include sources of bias and variation which are common to the measurement process. These sources include, but are not limited to: intrinsic instrument noise, some typical amount of carryover error, plus differences in laboratories, analysts, sample preparation, and instruments.

1.3.3 *Avoidable Sources of Error Excluded*—The IDE should realistically exclude avoidable sources of bias and variation, that is, those which can reasonably be avoided in routine field measurements. Avoidable sources would include, but are not limited to: modifications to the sample, measurement procedure, or measurement equipment of the validated method, and gross and easily discernible transcription errors (provided there was a way to detect and either correct or eliminate them).

1.3.4 *Low Probability of False Detection*—The IDE is a true concentration consistent with a measured concentration threshold (critical measured value) that will provide a high probability, 99 %, of true nondetection (a low probability of false detection, $\alpha = 1$ %). Thus, when measuring a blank sample, the probability of not detecting the analyte would be 99 %. To be useful, this must be demonstrated for the particular matrix being used, and not just for reagent water.

1.3.5 *Low Probability of False Nondetection*—The IDE should be a true concentration at which there is a high probability, at least 95 %, of true detection (a low probability of false nondetection, $\beta = 5$ %, at the IDE), with a simultaneous low probability of false detection (see 1.3.4). Thus, when measuring a sample at the IDE, the probability of detection would be at least 95 %. To be useful, this must be demonstrated for the particular matrix being used, and not just for reagent water.

NOTE 1—The referenced probabilities, α and β, are key parameters for risk-based assessment of a detection limit.

1.4 The IDE applies to measurement methods for which calibration error is minor relative to other sources, such as when the dominant source of variation is one of the following (with comment):

1.4.1 *Sample Preparation*, and calibration standards do not have to go through sample preparation.

1.4.2 *Differences in Analysts*, and analysts have little opportunity to affect calibration results (such as with automated calibration).

1.4.3 *Differences in Laboratories*, for whatever reasons, perhaps difficult to identify and eliminate.

1.4.4 *Differences in Instruments* (measurement equipment), which could take the form of differences in manufacturer, model, hardware, electronics, sampling rate, chemical processing rate, integration time, software algorithms, internal signal processing and thresholds, effective sample volume, and contamination level.

1.5 *Alternative Data Quality Objectives*—Other values forα, β, confidence, etc. may be chosen for calculating an IDE; however, this procedure addresses only the 99 %/95 % IDE.

2. Referenced Documents

2.1 *ASTM Standards:*

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

Current edition approved Aug. 10, 2003. Published September 2003. Originally approved in 1997. Last previous edition approved in 1997 as D 6091 – 97.

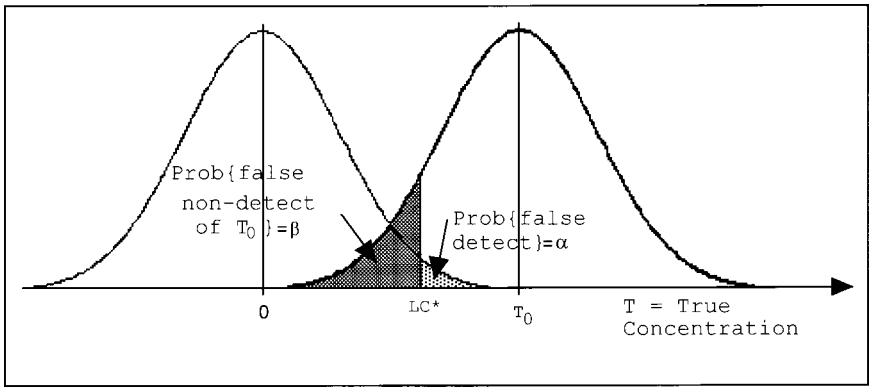

FIG. 1 Simplest Case of Reliable Detection

D 2777 Practice for the Determination of Precision and Bias of Applicable Test Methods of Committee D19 on Water[2]

3. Terminology

3.1 *Definitions:*

3.1.1 *99 %/95 % Interlaboratory Detection Estimate (99 %/95 % IDE, also denoted LD for Limit of Detection in accordance with Currie* **(1)**[3]—The lowest concentration at which there is 90 % confidence that a single measurement from a laboratory selected from the population of qualified laboratories represented in an interlaboratory study will have a true detection probability of at least 95 % and a true nondetection probability of at least 99 %.

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *Censored Measurement*—A measurement that is not reported numerically nor is reported missing but as a nondetect or a less-than, for example, "less than 0.1 ppb." The former means that an algorithm in the measurement system determined that the measurement should not be reported numerically for one of two reasons: either it was considered not sufficiently precise or accurate, or the identification of the analyte was suspect. A reported less-than may have the same meaning, but it also implies (perhaps erroneously) that any concentration greater than or equal to the accompanying value (for example, 0.1 ppb) can be measured and will be reported numerically.

3.2.2 *Detection Limit (DL) or Limit of Detection (LD)*—A numerical value, expressed in physical units or proportion, intended to represent the lowest level of reliable detection (a level which can be discriminated from zero with high probability while simultaneously allowing high probability of nondetection when blank samples are measured.

NOTE 2—In some cases, the discrimination may be from a value other than zero, such as a background level. Note also that a DL also depends on other characteristics of the measurement and detection process, such as described in 1.3.2. The IDE is an example of a DL.

3.2.3 *Probability of False Detection*—The false positive probability, denoted α, that a single measurement of a blank sample will result in a detection. (See Fig. 1.) This probability is often referred to as the Type 1 error probability and depends on the analyte, measurement system, analytical method, matrix, analyst, and measurement (recovery) threshold (measurement critical value) used to decide whether detection has occurred. This definition can be generalized to refer to unwanted detection from a single measurement of a sample at any nonzero concentration of the analyte rather than a blank sample, provided that the nonzero concentration is less than the detection limit or IDE.

3.2.4 *Probability of False Nondetection*—The false negative probability, denoted β or β (*T*), that a single measurement of a sample containing a nonzero concentration, *T*, of an analyte of interest will result in a nondetection. This is the complement of the probability of true detection. (See Fig. 1.) This probability function is often referred to as the Type 2 error probability function, and it depends explicitly on the concentration (*T*). It depends implicitly on the analyte, measurement system, analytical method, matrix, analyst, and critical value for detection.

3.2.5 *Probability of True Detection*—The probability, denoted 1-β or 1-β (*T*), that a single measurement of a sample containing a nonzero concentration, *T*, of an analyte of interest will result in a detection. (See Fig. 1.) This probability is often referred to as statistical power or the power of detection, and it depends explicitly on the concentration (*T*). It depends implicitly on the analyte, measurement system, analytical method, matrix, analyst, and critical value for detection.

3.2.6 *Probability of True Nondetection*—The true negative probability, denoted 1-α, that a single measurement of a blank sample will result in a nondetection. This is the complement of the probability of false detection. (See Fig. 1.) This probability also depends on the analyte, measurement system, analytical method, matrix, analyst, and response threshold. The probability of true nondetection can be similarly generalized: it can apply to a single measurement of a sample at any nonzero concentration less than the detection limit or IDE.

3.2.7 *100(1-γ) %—Confidence Statistical Tolerance Limit for 100(1-δ) % of a Population (also known as a One-Sided Statistical Tolerance Interval)*—A statistically determined limit that will, with 100(1-γ) % confidence, exceed (or fall below) 100(1-δ) % of the population (the 100(1-δ) % quantile). See Hahn and Meeker **(2)** for further explanation and tables of values.

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

[3] The boldface numbers in parentheses refer to the list of references at the end of this standard.

4. Summary of Practice

4.1 Every ASTM D-19 test method is evaluated to determine precision and bias by conducting a collaborative study in accordance with Practice D 2777. That study, or a similar collaborative study, can also be used to evaluate the lowest concentration level of reliable detection for a test method, referred to herein as the Interlaboratory Detection Estimate. Such a study must include concentrations suitable for modeling the uncertainty of mean recovery of interlaboratory measurement (preferably without extrapolation). It must also be planned and conducted to allow the known, routine sources of measurement variability to be observed at typical levels of influence. After it is conducted, outlying laboratories and individual measurements should be eliminated using an accepted, scientifically based procedure for outlier removal, such as found in Practice D 2777. The IDE computations must be based on retained data from at least six independent laboratories at each concentration level.

4.2 Retained data are analyzed to identify and fit one of three proposed interlaboratory standard deviation (ILSD) models which describe the relationship between the interlaboratory standard deviation of measurements and the true concentration. The identification process involves evaluating the models in order, from simplest to most complex: constant, straight-line, or exponential (all with respect to true concentration, *T*). Evaluation includes statistical significance and residual analysis.

4.3 The chosen model is used to predict interlaboratory measurement standard deviation at any true concentration within the study concentration range. If interlaboratory standard deviation is not constant, the predictions are used to generate weights for fitting the mean recovery relationship (the straight-line relationship between measured concentration and true concentration), using weighted least squares (otherwise, ordinary least squares is used). The mean recovery curve is evaluated for statistical significance and lack of fit and using residual analysis. An ILSD model prediction is also used to estimate the interlaboratory standard deviation of measurements of blanks. This estimate is used to compute *YC*, a measurement critical value for detection (see 6.4.1). The *YC* is the value that with approximately 90 % confidence will not be exceeded by 99 % of all measurements of blanks made by qualified laboratories as represented in the study. The *LC* computed from *YC* is the true concentration with expected measurement equal to *YC* (see 6.4.2). The model is also used to predict interlaboratory standard deviation at nonzero concentrations. The IDE is directly or iteratively computed to be the true concentration that with approximately 90 % confidence will produce measurements that will exceed *YC* at least 95 % of the time and simultaneously not exceed more than 1 % of the time when blank samples are measured.

5. Significance and Use

5.1 Appropriate application of this practice should result in an IDE achievable by most laboratories properly using the test method studied. This IDE provides the basis for any prospective use of the test method by qualified laboratories for reliable detection of low-level concentrations of the same analyte as the one studied in this practice and same media (matrix).

5.2 The IDE values may be used to compare the detection power of different methods for analysis of the same analyte in the same matrix.

5.3 The IDE provides high probability (approximately 95 %) that result values of the method studied which exceed the IDE represent presence of analyte in the sample and high probability (approximately 99 %) that blank samples will not result in a detection.

5.4 The IDE procedure should be used to establish the interlaboratory detection capability for any application of a method where interlaboratory detection is important to data use. The intent of IDE is not to set reporting limits.

6. Procedure

6.1 The procedure described as follows has stages described in the following sections: IDE Study Plan, Design and Protocol (6.2); Conduct the IDE Study, Screen the Data, and Choose a Model (6.3); and Compute the IDE (6.4). A flowchart of the procedure is shown in Fig. 2.

6.2 *IDE Study Plan, Design, and Protocol*:

6.2.1 *Choose Analyte, Matrix, and Method*—At least one analyte of interest is selected, typically one for which there is interest in trace levels of concentration, such as toxic materials that are controlled and regulated. For each analyte, an approximate maximum true concentration is selected based on the following considerations:

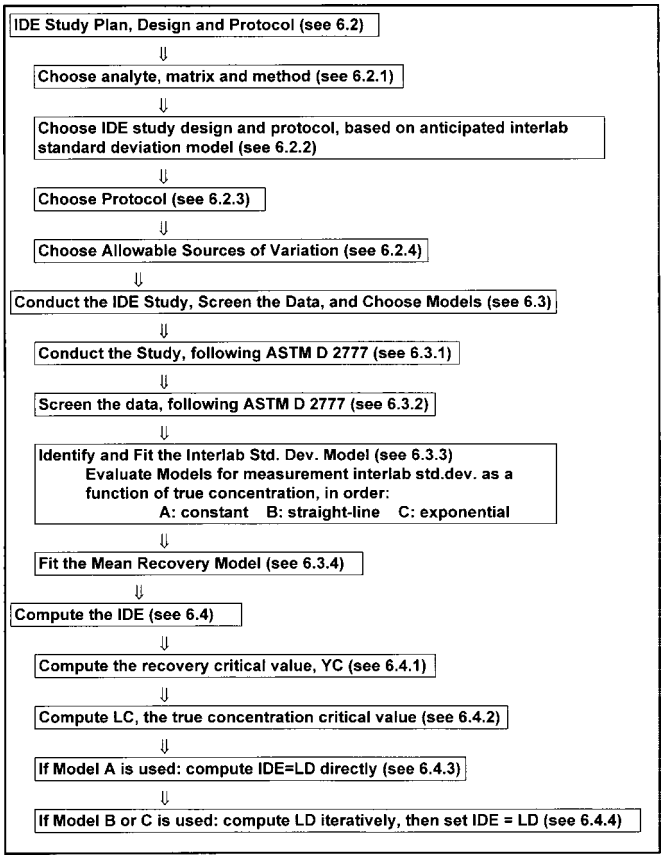

FIG. 2 Flowchart of IDE Procedure

6.2.1.1 The anticipated IDE should be exceeded by a factor of 2 or more,

6.2.1.2 A single model (ideally a straight-line model in true concentration, *T*) should describe mean recovery from zero to that maximum concentration,

6.2.1.3 A single model in true concentration should describe interlaboratory measurement standard deviation from zero to that maximum concentration, and

6.2.1.4 The range must be sufficient to enable statistically significant coefficients to be estimated for the ILSD model and mean recovery model. One or more matrices of interest are also selected, and an accepted standard analytical method for those analytes is selected for study. If there is no possibility of matrix interference, then it may only be necessary to determine a list of acceptable matrices which can be used instead of selecting a specific matrix. For example, for a particular analyte, concentration range, and method it may be supposed that reagent waters from different laboratories are indistinguishable, but for another analyte or another concentration range that assumption may not hold.

6.2.2 *Choose IDE Study Design and Protocol*, based (if possible) on anticipated interlaboratory standard deviation (ILSD) model. Section 7 of Practice D 2777 can be followed for the study design and protocol. The anticipated form of the ILSD model (the relationship between interlaboratory measurement standard deviation and true concentration) can help in choosing an IDE study design. Three models are proposed herein for the interlaboratory measurement standard deviation with respect to true concentration: constant, straight-line (increasing), and exponential (increasing). Chemistry, physics, empirical evidence, or informed judgment may make one model more likely than others. However, it may not be possible to anticipate the relationship between standard deviation and true concentration.

6.2.2.1 Select an IDE study design that has enough distinct concentrations to assess statistical lack of fit of the models (see Draper and Smith **(3))**. Recommended designs are: (*a*) The semi-geometric design at five or more true concentrations, $\{T_1, T_2$, and so forth$\}$, such as: $\{0, IDE_0/4, IDE_0/2, IDE_0, 2 \times IDE_0, 4 \times IDE_0\}$, where IDE_0 is an initial estimate of the IDE (such as $10 \times s'$, where s' is the interlaboratory measurement standard deviation at a trace-level, nonzero concentration), (*b*) equi-spaced design: $\{0, IDE_0/2, IDE_0, (3/2) \times IDE_0, 2 \times IDE_0, (5/2) \times IDE_0\}$, and (*c*) any other design with at least five concentrations, provided that the design includes blanks, at least one concentration approximately equal to $2 \times IDE_0$, and at least one nonzero concentration below IDE_0.

6.2.2.2 The study concentration levels must either be: known (true concentration levels), or knowable, after the fact. A concentration is considered known if reference standards can be purchased or constructed and knowable if an accurate determination can be made (for example, the median value from many laboratories, or results from a recognized laboratory, such as NIST, using a high-accuracy method).

6.2.3 *Choose Protocol*—The protocol should follow Section 7 of Practice D 2777. It should include design run order and details on when the system is to be purged, have extra blanks run, and so forth. It should take into consideration possible problems with carryover, study cost (in time and money), and time constants of measurement system drift or sample degradation.

6.2.3.1 For purposes of the collaborative study, the study supervisor should provide instructions to participating laboratories to disable (if possible) any internal measurement system thresholds (such as an instrument detection limit or peak-area threshold) that are used to determine whether a numerical measurement is to be reported as a nondetect or less-than, or as a number (censoring). If censoring is unavoidable, the laboratory censoring threshold must be reported with its study data. However, qualitative criteria used by the method to identify and discriminate analytes are separate criteria and must be satisfied according to the method.

6.2.4 *Choose Allowable Sources of Variation*—It is assumed that collectively the many sources of variation will contribute to cause interlaboratory measurements at any true concentration to be normally distributed. Representative between-laboratory variation can only be seen if the number of laboratories providing usable data is maximized. Ordinary within-laboratory variation must be allowed to affect the measurement process as happens in routine measurement. Ideally, there would be many laboratories, and each measurement at each laboratory would be an unsuspecting blind measurement made by a different analyst using a different (qualified) measurement system on a different day, in random order.

6.2.4.1 As emphasized in Practice D 2777, maximizing the number of participating laboratories is often the most important thing that can be done to guarantee a successful study, and there are several reasons why the number of participating laboratories will somewhat exceed the number of laboratories providing a full set of usable data. A minimum of ten participating laboratories is recommended.

6.2.4.2 If possible, the study should be conducted completely blind, particularly if the method is labor-intensive, as opposed to a highly automated method. That is, not only should the analysts not be aware of the true concentrations of the samples they are measuring, but they should not even be aware of the fact that they are measuring special, study samples. This is to minimize the extra care distortion of data so common in analytical studies.

6.2.4.3 For each laboratory, the maximum number of qualified analysts possible should be involved in the study since there are variations which may be allowed by the method, may be practiced by different analysts, and will be seen in routine analyses.

6.2.4.4 For each laboratory, the maximum number of qualified measurement systems should be used since there are model-to-model and instrument-to-instrument differences in equipment and maintenance, as will be seen in routine analyses.

6.2.4.5 For each laboratory, the IDE study should be scheduled to span the maximum possible number of days consistent with holding time constraints since day-to-day changes in analytical laboratory environmental conditions, contamination, solvent purity, and other factors can affect measurements, and will be seen in routine analyses.

6.3 *Conduct the IDE Study, Screen the Data, and Choose a Model*:

6.3.1 The IDE study should be conducted in accordance with Section 9 of Practice D 2777. Blank correction should not be performed by the laboratories, unless the method requires this subtraction in order to perform the test. Each laboratory should supply method blank data along with the uncorrected measurement values, and the study supervisor can determine whether the reported measurements should be corrected.

6.3.2 The IDE study data should be screened in accordance with the initial subsections relating to removing data, Section 10 of Practice D 2777. Skip to 6.5 if, for any concentration, more than 10 % of the retained measurements are nondetects or less-thans.

6.3.3 *Identify and Fit the ILSD Model*—The ILSD model should be identified, and its coefficients should be estimated by using the following procedure. See Caulcutt and Boddy **((4))** for more discussion of standard deviation modeling and weighted least squares (WLS) in analytical chemistry. This model is an attempt to characterize the unknown (or partly known) function between interlaboratory measurement standard deviation and true concentration, $\sigma = G(T)$. It is used for two purposes: to provide weights for the WLS regression to fit the mean recovery model and to provide the interlaboratory standard deviation estimates crucial to determining critical values and the IDE.

6.3.3.1 Three ILSD models are proposed. The identification process considers (fits and evaluates) each model in turn, from simplest to most complex, until a suitable model is found. Prior knowledge can be combined with empirical results to influence the selection of a model if a suitable refereed publication can be cited. See Carroll and Ruppert **((5))** for further discussion of standard deviation modeling. The model order is as follows:

Model A (Constant ILSD Model):

$$s = g + \text{error} \tag{1}$$

where: g is a fitted constant. Standard deviation does not change with concentration, resulting in a relative standard deviation that declines with increasing T.

Model B (Straight-line ILSD Model):

$$s = g + h \times T + \text{error} \tag{2}$$

where: g and h are fitted constants. Standard deviation increases linearly with concentration, resulting in an asymptotically constant relative standard deviation as T increases.

Model C (Exponential ILSD Model):

$$s = g \times \exp\{h \times T\} + \text{error} \quad \text{or} \tag{3}$$

$$s = g \times \exp\{h \times T\} \times \text{error} \tag{4}$$

where: g and h are fitted constants. Interlaboratory standard deviation increases exponentially with concentration, resulting in a relative standard deviation that may initially decline as T increases but eventually increases as T increases. Error can be additive or multiplicative.

(a) (*a*) In all cases, it is assumed that $g > 0$. A value of $g < 0$ has no practical interpretation and may indicate that a different ILSD model should be used. Furthermore, it is assumed that g is not underestimated due to censored data among measurements of blanks or other low-concentration samples. (Censoring is addressed in 6.2.3.1, 6.3.2, and 6.5.)

(b) (*b*) If $h < 0$, it must not be statistically significant, and Model A should be evaluated.

6.3.3.2 *ILSD Model Identification and Fitting Procedure*:

(*1*) Merge all retained IDE study data (after possible elimination of some data in accordance with 6.3.2).

(*2*) For each true concentration, T_k, compute the adjusted interlaboratory sample standard deviation, s_k, an estimate of the true underlying interlaboratory measurement standard deviation, σ_k. The adjusted interlaboratory sample standard deviation is the sample standard deviation s_k, multiplied by the bias-correction factor, a'_n found in Table 1. In this Practice, all references to computed and fitted values of the interlaborator sample standard deviation refer to adjusted values. Note that a simplifying approximation can be used if the number of retained replicates is the same for each spike level; unadjusted sample standard deviations can be sued, and the final IDE can be multiplied by the adjustment factor (see the example). The larger the number of replicates, the better the approximation.

(*3*) Plot s_k versus T_k.

(*4*) Using ordinary least squares (OLS) (see Caulcutt and Boddy **(4)**, regress s_k on T_k, temporarily assuming that a straight-line model is valid. This provides coefficients, g and h, in the relationship:

$$s_k = g + h \times T_k + \text{error} \tag{5}$$

(*e*) Evaluate the reasonableness of Model A (the constant ILSD model) by doing two things. Note the *p*-value associated with slope estimate h, from the OLS regression. If it is less than 5 %, there is statistically significant slope, and Model A should be rejected; proceed to the next step. Secondly, examine the plot produced in step (*c*), or a plot of the residuals from the OLS fit. If obvious systematic curvature is present (for example, quadratic or exponential-like behavior), Model A should be rejected; proceed to step (*h*). If Model A is not rejected, skip to 6.3.4.

(*f*) Model A is rejected, due to statistically significant slope. Compute residuals:

$$r_k = s_k - (g + h \times T_k) \tag{6}$$

TABLE 1 Bias-Correction Adjustment Factors for Sample Standard Deviations Based on n Measurements (at at particular concentration)[A]

n	2	3	4	5	6	7	8	9	10
a'_n	1.253	1.128	1.085	1.064	1.051	1.042	1.036	1.031	1.028

[A] For each true concentration T_k, the adjusted value $s_k = a'_n s'_k$ should be modeled in place of sample standard deviation s'_k. For $n > 10$, use the formula $a'_n = 1 + [4(n - 1)]^{-1}$. See Johnson and Kotz **(7)**.

Plot r_k versus T_k.

(*g*) Evaluate the reasonableness of Model B (the straight-line ILSD model). Examine the plot produced in step (*f*). If obvious systematic curvature is present (for example, quadratic or exponential-like behavior), with a minimum that appears to be within the concentration range, Model B should be rejected; proceed to step (*h*). If Model B is not rejected, skip to 6.3.4.

(*h*) To evaluate the reasonableness of Model C (the exponential ILSD model), the model must first be fit. There are two approaches. The simplest approach is to do OLS regression on the log of the interlaboratory sample standard deviations:

$$\ln s_k = \ln g + h \times T_k + \text{error} \tag{7}$$

This corresponds to the multiplicative error assumption, which is generally a good assumption. The fit will provide h directly and $g' = \ln g$ which is converted, $g = \exp\{g'\}$. Alternatively, the fit can be done using nonlinear least squares (NLLS), by Newton-Raphson iteration or another method. This approach corresponds to the less-plausible additive error assumption. In either case, the fit should satisfy two types of evaluation. First, the *p*-value for h should be less than 5 %. Secondly, a plot of the residuals, in log form, should be constructed. Plot r_k versus T_k, where:

$$r_k = \ln s_k - (\ln g + h \times T_k) \tag{8}$$

The plot should show no systematic behavior (for example, curvature). If the fit satisfies both types of evaluation, proceed to 6.3.4. Otherwise, a different and possibly more complex model will have to be used. One possibility is the Rocke and Lorenzato **(6)** model, which has:

$$s \approx (g + h \times T^2)^{1/2} \tag{9}$$

This model has nearly constant (slightly increasing) ILSD for low true concentrations, changing to standard deviation nearly proportional to concentration for higher concentration levels. It can be fit and evaluated using NLLS or maximum likelihood. If there are enough true concentrations, a model with more coefficients could be considered, such as quadratic (strictly increasing with increasing concentration), or even cubic.

6.3.4 *Fit the Mean Recovery Model*—The mean recovery model is a simple straight line:

$$\textit{M}\text{odel } R\text{: } Y = a + b \times T + \text{error} \tag{10}$$

The fitting procedure depends on the model selection from 6.3.3. If Model A was selected for ILSD, then OLS can be used to fit Model R for mean recovery (see Caulcutt and Boddy **(4)**). If a nonconstant ILSD model was selected, such as Model B or C, then WLS should be used to fit mean recovery. This approximately provides the minimum variance unbiased linear estimate of the coefficients a and b. The WLS procedure appears in 6.3.4.1.

6.3.4.1 *Weighted Least Squares Procedure, Using the Interlaboratory Standard Deviation Model*:

(a) (*a*) Using the ILSD model and coefficient estimates from 6.3.3, compute predicted interlaboratory standard deviation, $\hat{s}_k$ for each true concentration, T_k:

$$\textit{M}\text{odel } B\text{: } \hat{s}_k = g + h \times T_k \tag{11}$$

$$\textit{M}\text{odel } C\text{: } \hat{s}_k = g \times \exp\{h \times T_k\} \tag{12}$$

(b) (*b*) Compute weights for WLS:

$$w_k = (\hat{s}_k)^{-2} \tag{13}$$

(c) Note that if this is done using computer software, the default setting for weights may be different. For example, instead of supplying the values, $(\hat{s}_k)^{-2}$ as weights, the software may require the user to supply values $(\hat{s}_k)$ or $(\hat{s}_k)^2$ as weights that are internally transformed by the software.

(d) (*c*) Carry out WLS computations analogous to OLS computations. See Table 2 or Caulcutt and Boddy **(4)**. The result will be coefficient estimates, a and b, for the mean recovery model, Model R.

(e) (*d*) There are three approximate approaches to WLS commonly practiced but that are not acceptable for this application. One approach uses the reciprocal squared sample standard deviations as weights. In this context, since a standard deviation model is explicitly evaluated and selected, the predicted value for $s_k$is probably more precise than a sample value. The predicted value should be used in place of the sample standard deviation for weight computation. A second approach omits the blank measurements, divides through the rest of the measurements by the true concentrations, and does OLS using the independent variable $1/T$ in the model:

$$Y/T = a \times (1/T) + b + \text{error} \tag{14}$$

(f) This is not acceptable because it leads to loss of data and because the weights so generated implicitly assume that interlaboratory standard deviation is strictly proportional to true concentration. The IDE concept and computation rests on positive, quantifiable interlaboratory standard deviation for measurements of blanks, and a proportional relationship cannot hold for arbitrarily small concentrations. The third approach exploits the same approximate but untrue proportional relationship to obtain mathematically simpler WLS formulas.

(g) (*e*) After fitting, the mean recovery model should be evaluated for reasonableness and lack of fit. This should be

TABLE 2 Computations to Estimate Straight-Line Model Coefficients By Means of Least Squares—Ordinary and Weighted

Ordinary Least Squares, OLS	Weighted Least Squares, WLS
$\bar{T} = \frac{1}{n_i}\sum_{i=1}^{n} T_i$,	$\bar{T}_w = \sum_{i=1}^{n} w_i T_i \Big/ \sum_{i=1}^{n} w_i$
$\bar{y} = \bar{T} = \frac{1}{n_i}\sum_{i=1}^{n} T_i$	$\bar{y}_w = \sum_{i=1}^{n} w_i y_i \Big/ \sum_{i=1}^{n} w_i$
$S_{TT} = \sum_{i=1}^{n} (T_i - \bar{T})^2$	$S_{wTT} = \sum_{i=1}^{n} w_i (T_i - \bar{T})^2$
$S_{TY} = \sum_{i=1}^{n} (T_i - \bar{T})(y_i - \bar{y})$	$S_{wTY} = \sum_{i=1}^{n} w_i (T_i - \bar{T})(y_i - \bar{y})$
Slope = $b = S_{TY}/S_{TT}$ Intercept = $a = \bar{y} - b\bar{T}$	Slope = $b = S_{wTY}/S_{wTT}$ Intercept = $a = \bar{y}_w - b\bar{T}_w$

done by ensuring the following: (*1*) The fit is statistically significant (overall *p*-value <5 %); (*2*) The lack of fit *p*-value (if available; see Caulcutt and Boddy **(4)** or Draper and Smith **(3))** is not statistically significant (lack of fit p-value > 5 %); (*3*) A plot of the residuals should show no obvious systematic curvature (for example, quadratic or exponential-like behavior). If the mean recovery model fails the evaluation, then the study supervisor will have to determine if only a subset of the data should be analyzed (perhaps the model fails for the higher concentration(s)), or if more data are needed.

6.4 *Compute the IDE*—The IDE is computed using the ILSD model to estimate interlaboratory standard deviation at true concentration = 0 and at the IDE, and using the mean recovery model to transform measured concentrations to true concentrations and vice versa. The computation has three stages, where the following are computed in succession: *YC* = recovery critical value, *LC* = true concentration critical value, and *LD* = IDE. Additionally, one can compute *YD* = the expected measurement at the IDE.

6.4.1 Compute the recovery critical value:

$$YC = k1 \times \hat{s} + a \tag{15}$$

where:

- $k1$ = one-sided, 90 % confidence upper statistical tolerance limit (also known as the one-sided statistical tolerance interval) for the 99 % quantile of the normal distribution), based on n observations (see Table 2),
- n = total number of measurements retained in the IDE study after 6.3.2,
- $\hat{s}(0)$ = G (0), the predicted interlaboratory standard deviation of the measurement of a blank sample. G (0) = g for ILSD Model B or C, but for Model A, G (0) should be set to the root mean squared error (RMSE) from the recovery model fit,
- a = estimated mean recovery intercept, and
- YC = measurement value that with (approximately) 90 % confidence will be exceeded no more than 1 % of the time when a blank sample is measured.

6.4.2 Compute *LC*, the true concentration critical value, by inverting the mean recovery formula with value *YC*:

$$LC = R^{-1}(YC) = (YC - a)/b \tag{16}$$

where:

- R^{-1} = inverse prediction function that transforms a measured concentration into the true concentration, based on mean recovery, modeled by a straight line, and
- LC = true concentration that has an expected recovery that with (approximately) 90 % confidence, will be exceeded no more than 1 % of the time when a blank sample is measured.

6.4.3 If the constant ILSD model (Model A) was used, compute:

$$LD = LC + k2 \times \hat{s}(0)/b \tag{17}$$

where: $k2$ = one-sided, 90 % confidence upper statistical tolerance limit for the 95 % quantile of the normal distribution, based on *n* observations (see Table 3).

6.4.4 If either the straight-line ILSD model (Model B) or the exponential ILSD model (Model C) was used, compute *LD* by recursively solving:

$$LD_{i+1} = R^{-1}(k1 \times \hat{s}(0) + k2 \times G(LD_i) + a)$$

$$= [k1 \times \hat{s}(0) + k2 \times G(LD_i)]/b \tag{18}$$

where: $G(LD_i)$ is the predicted interlaboratory standard deviation at true concentration LD_i. Therefore, the recursive *LD* formulas are as follows:

$$\text{Model } B\text{: } LD_{i+1} = [k1 \times \hat{s}(0) + k2 \times (g + h \times LD_i)]/b \tag{19}$$

$$\text{Model } C\text{: } LD_{i+1} = [k1 \times \hat{s}(0) + k2 \times (g \times \exp\{h \times LD_i\})]/b \tag{20}$$

If a different, nonconstant ILSD model, such as the Rocke and Lorenzato model **(6)** is used, the recursive formula for *LD* would take the general form of (Eq 18).

6.4.4.1 A reasonable initial estimate for *LD* is:

$$LD_0 = 2 \times LC, \text{ or} \tag{21}$$

$$LD_0 = LC + k2 \times \hat{s}(0)/b \tag{22}$$

For each iteration, the current estimate of *LD* is plugged into the right-hand side of the recursive formula, producing a new estimate for *LD*. Iterations should continue until the relative difference between successive *LD* estimates is < 1 %. The *LD* is the true concentration about which with (approximately) 90 % confidence, a single sample measurement will produce a reported measurement that 95 % of the time will exceed *YC*.

6.4.5 The result is *IDE* = *LD*. The *IDE* is the true concentration at which the measurement of a single sample will

TABLE 3 90 %-Confidence Upper, One-sided Statistical Tolerance Limit Factors for Computing the 99 %/95 % IDE[A]

Number of Observations Retained, *n*	99 % Quantile, *k*1	95 % Quantile, *k*2
5	4.67	3.40
10	3.53	2.57
15	3.21	2.33
20	3.05	2.21
25	2.95	2.13
30	2.88	2.08
35	2.83	2.04
40	2.79	2.01
45	2.76	1.99
50	2.74	1.97
55	2.71	1.95
60	2.69	1.93
65	2.68	1.92
70	2.66	1.91
75	2.65	1.90
80	2.64	1.89
90	2.62	1.87
100	2.60	1.86
150	2.55	1.82
200	2.51	1.79

[A] Computed using STINT software (93/12/3 version), by Prof. W. Meeker and J. Chow of Iowa State University.

exceed *YC* 95 % of the time (resulting in a detection), and simultaneously, the measurement of a single blank sample will exceed *YC* only 1 % of the time, both with approximately 90 % confidence:

$$YD = a + b \times LD \quad (23)$$

is the expected measurement value for a sample at true concentration *LD*.

6.5 *Nontrivial Amount of Censored Data*, >10 % for at least one true concentration of data reported as nondetects or less-thans. Despite the attempt in 6.2.3.1 to reduce or eliminate reported nondetects or less-thans, they may still occur at a level that disrupts the analysis of the data presented in 6.3 and 6.4. If this happens, the study supervisor should contact laboratories with such measurements to see whether the uncensored data can be extracted from data archives. If this is not a sufficient remedy, serious consideration should be given to augmenting the IDE study with measurements of samples at new and different concentrations (generally, higher). A third and final option is to follow the procedure in 6.5.1 through 6.5.4. It should be noted, however, that the procedure provides no assurance of the probability of false positives, and the IDE so computed should always be identified with such a qualifier.

6.5.1 Use the Rocke and Lorenzato Model **(6)** as the ILSD model and fit it using NLLS with only data for concentrations that did not have more than 10 % nondetects or less-thans.

6.5.2 Use the same data as in 6.5.1 to fit Model R (the straight-line mean recovery model) using WLS.

6.5.3 If less than half of all blank sample results are reported as nondetect or less-thans, proceed with 6.4.2 through 6.4.5, using the models.

6.5.4 If half or more of all blank sample results are reported as nondetect or less-thans, use linear interpolation among low-concentration samples to estimate the true concentration that would have a detection probability of 50 %. This is the effective *LC*. For example, if nondetect or less-than was reported for 70 % of blank samples and for 20 % of samples with $T = 3$ ppb, then:

$$LC = 3 \times (70 - 50)/(70 - 20) = 1.2 \text{ ppb} \quad (24)$$

Proceed with 6.4.4 through 6.4.5.

7. Data Analysis

7.1 The data analysis for eliminating data is given in Section 10 of Practice D 2777.

7.2 The data analysis involved in computing an IDE is shown by example in Section 10.

8. Report

8.1 The analysis report should be structured as in Annex A1.

8.2 The report should be given a second-party review to verify that:

8.2.1 The data transcription and reporting have been correctly performed,

8.2.2 The analysis of the data has been correctly performed, and

8.2.3 The results of the analysis have been appropriately used, including possible rejection of assumptions necessary to compute an IDE.

8.3 A statement of the review and the results should accompany the report. Reviewer(s) should be qualified in one or more of the following areas: (*1*) applied statistics and (*2*) analytical chemistry.

9. Rationale

9.1 The basic rationale for the 99 %/95 % IDE is contained in Currie **((1))**, and is shown in Fig. 1. For a selected test method, this figure shows single-laboratory variation in measurements of both blank samples and samples at true concentration = T_0, assuming perfect recovery. The variation shown is according to the normal distribution with known mean (zero bias) and known interlaboratory standard deviation. The critical value, *LC**, is used to determine detection. It can be moved to decrease α = the probability of a false detection at the price of increasing β = the probability of a false nondetection, or vice versa. Given an acceptable value for α, a value for *LC** can be found. Given, also, an acceptable value for β, a suitable value for T_0 can be found. T_0 is then a single-laboratory detection limit at which reliable detection can occur by definition of acceptable α and β. Following this IDE procedure, this concept (*LC* and *LD*) can be extended to a method's interlaboratory detection capability estimation.

9.2 There are several real-world complications to Fig. 1. See Maddalone et al **(7)** and see Gibbons **(8)**. Some of these complications are listed with their remedies:

9.2.1 Recovery is not perfect; the relationship between measured values and true concentrations cannot be assumed to be trivial. There is bias between true and measured values. It can and should be modeled, typically by a straight line.

9.2.2 Variation is introduced by different laboratories, analysts, models and pieces of equipment, environmental factors, latitude in a test method, contamination, carry-over influence, and other factors. It is intractable to model these individually, but their collective contributions towards measurement interlaboratory standard deviation can be observed if it is part of how a study is designed and conducted.

9.2.3 The interlaboratory standard deviation of measurements (quantified by the standard deviation of the normal distribution) is unknown. Standard deviations must be estimated with finite sample sizes, and statistical tolerance limits must be used to obtain high confidence of an estimate of a distribution quantile.

9.2.4 Interlaboratory standard deviation of measurements may change with true concentration, possibly due to the physical principle of the test method. Short of severely restricting the concentration range for a study, this requires an empirical ILSD model to enable prediction of the interlaboratory standard deviation of measurements at different true concentrations.

9.3 A more realistic picture of analytical measurement is shown in Fig. 3 (though 11.2.2 cannot be shown).

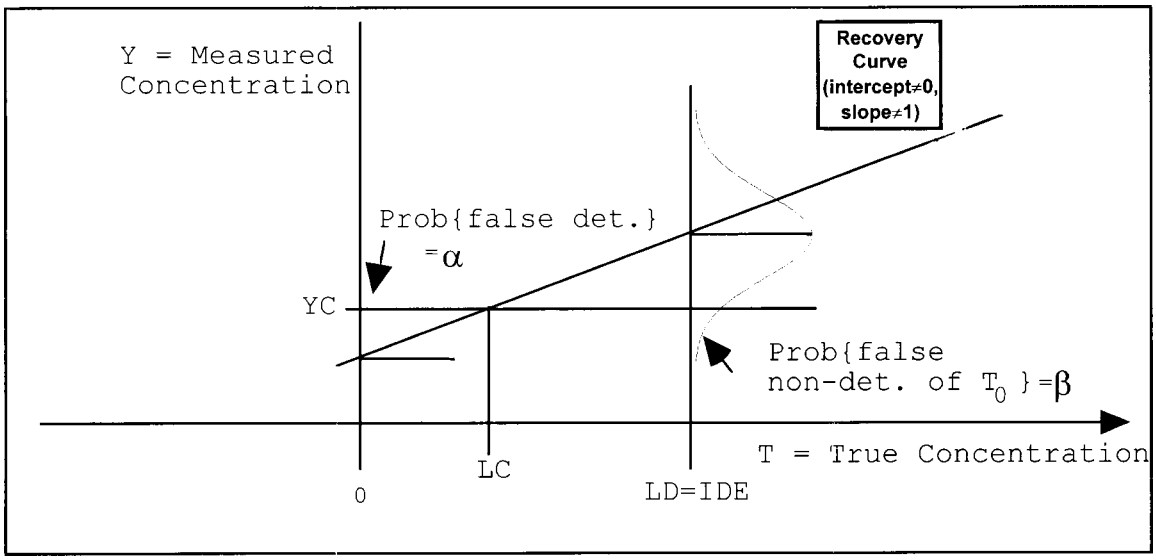

FIG. 3 Reliable Detection at the IDE (Realistic Case)

10. Example (Straight-Line ILSD Model)

10.1 *Identify and Fit the ILSD Model*—Ten laboratories participated in a (synthesized) IDE study where single measurements were made at each of five concentrations, including blanks: T_k = {0, 0.25, 0.50, 1, 2} ppb. Considerations of 6.2, 6.3.1, and 6.3.2 are not described in Section 10. The procedure described in 6.3.3 is followed, using the adjusted-IDE approximation of 6.3.3.2, assuming that no data were eliminated in accordance with 6.3.2.

10.1.1 The reported measurements are shown in Table 4. These values are also shown in Fig. 4. The straight-line recovery model appears to be plausible, and the data appear to have interlaboratory measurement standard deviation that increases with concentration. Note that for this example, high blank measurements and an unusually high recovery slope were used for the purposes of illustration (to distinguish measured values from true values). In practice, the recovery curve intercept and slope would typically be closer to 0 and 1, respectively.

TABLE 4 Reported Measurements and Computed Statistics from IDE Study

True Concentration, T_k, ppb	Reported Measurement, Y_i, one per Laboratory, ppb	Sample Standard Deviation	Predicted Standard Deviation	Weights for WLS
0.0	1.41, 3.94, 2.22, 3.48, 1.96, 0.92, 2.17, 2.36, 4.50, 3.26	1.137	1.089	0.843
0.25	4.10, 3.51, 4.07, 4.34, 4.54, 2.76, 2.03, 4.13, 6.06, 6.47	1.336	1.328	0.567
0.50	3.97, 7.34, 6.41, 6.25, 6.38, 7.64, 4.67, 6.74, 4.38, 6.48	1.255	1.568	0.407
1.0	7.54, 7.68, 8.38, 7.14, 3.12, 10.97, 11.15, 10.44, 9.73, 7.27	2.406	2.046	0.239
2.0	8.20, 13.97, 12.88, 18.31, 16.47, 16.06, 12.56, 14.21, 13.96, 17.37	2.900	3.003	0.111

10.1.2 Interlaboratory sample standard deviations at each true concentration are computed, and are shown in Table 4.

10.1.3 A plot of interlaboratory sample standard deviation versus true concentration is shown in Fig. 5. There is increasing qualitative evidence of an increase in standard deviation with increasing concentration.

10.1.4 A straight-line regression (OLS) is conducted of the interlaboratory sample standard deviations, s_k, versus T_k. The results are shown in Table 5, and the fit is shown in Fig. 5. The estimates are intercept g = 1.0891 and slope h = 0.95682.

10.1.5 The p-value associated with the slope estimate, h, is 1.28 % < 5 %, so Model A, the constant ILSD model, is rejected.

10.1.6 The residuals from the straight-line interlaboratory standard deviation fit are computed as follows and are displayed in Fig. 6:

$$r_k = s_k - (1.089 + 0.957 \times T_k) \tag{25}$$

10.1.7 There is no evidence of systematic curvature, so the analysis proceeds in accordance with 6.3.4.

10.2 *Fit the Mean Recovery Model*—Since the interlaboratory standard deviation has been shown to be nonconstant with respect to true concentration, WLS is used to fit the mean recovery model, and the fitted ILSD model is used explicitly to estimate the ILSD at arbitrary true concentrations. The procedure described in 6.3.4 is followed.

10.2.1 The estimate of intercept, g, and the estimate of slope, h, in the straight-line ILSD model, are used to predict the ILSD at each true concentration, T_k. These predicted values, $\hat{s}_k$, are shown in Table 4 and are assumed to be closer to the true ILSDs, σ_k, than are the sample ILSDs, s_k.

10.2.2 Weights are computed, based on the predicted ILSDs:

$$w_k = (\hat{s}_k)^{-2} \tag{26}$$

They are shown in Table 4.

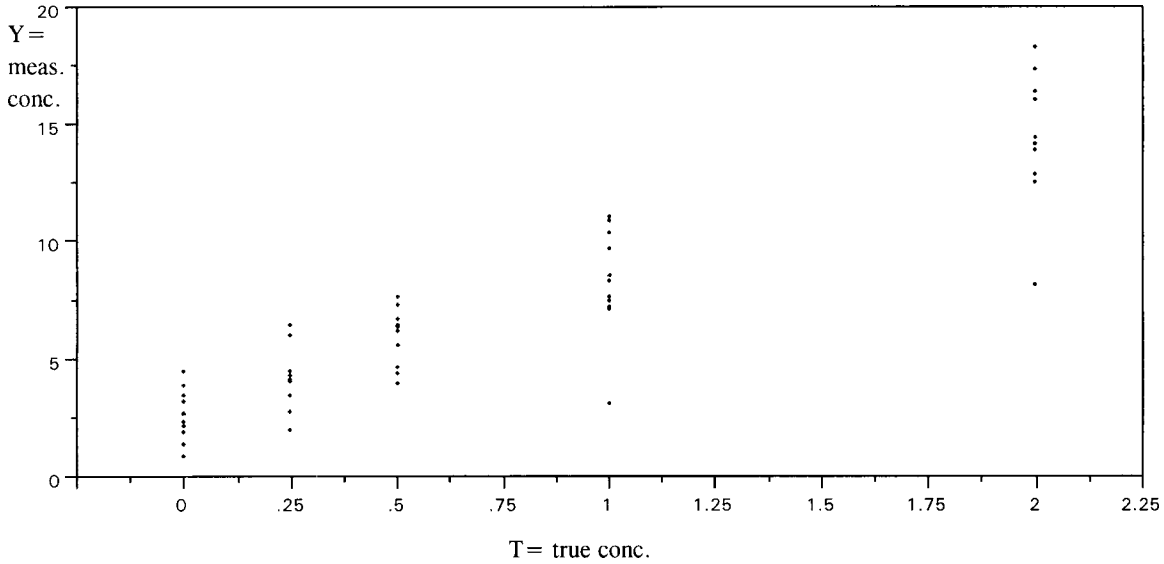

FIG. 4 Reported Measurements versus True Concentration, One Measurement per Laboratory at Each Concentration (ppb)

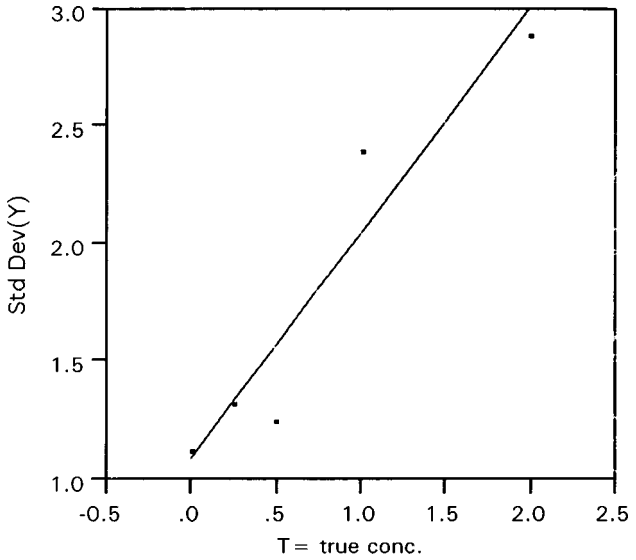

FIG. 5 Sample Standard Deviation (*Y*) versus True Concentration (s_k versus T_k)

TABLE 5 Results of Straight-Line Fit of s_k versus T_k by OLS

Linear Fit				
Standard Deviation (*Y*) = 1.0891 + 0.95682 T				
Standard Deviation (*Y*) = g + h T				
Summary of Fit				
RSquare	0.904996			
RSquare Adj	0.873329			
Analysis of Variance				
Source	df	Sum of Squares	Mean Square	F Ratio
Model	1	2.2887587	2.28876	28.5778
Error	3	0.2402664	0.08009	Prob>F
C total	4	2.5290251		0.0128
Parameter Estimates				
Term	Estimate	Standard Error	T-Ratio	Prob> T
g (intercept)	1.0891019	0.184493	5.90	0.0097
h (slope)	0.9568195	0.178985	5.35	0.0128

[A] Key results are underlined.

10.2.3 The WLS is carried out to estimate the coefficients, *a* and *b*, of the straight-line mean recovery relationship:

$$\text{Model } R\text{: } Y = a + b \times T + \text{error} \quad (27)$$

The results of WLS are shown in Fig. 7 and in Table 6.

10.2.4 The fit is evaluated as follows: (*1*) The overall *p*-value is <0.0001 < 5 %; (*2*) The lack of fit *p*-value is 0.8537 > 5 %; (*3*) Fig. 8 shows a plot of the residuals versus true concentration and shows no evidence of systematic curvature. Therefore, the straight-line mean recovery fit is acceptable.

10.3 *Compute the IDE*—Having obtained acceptable fits of a ILSD model and a mean recovery model, the IDE can be computed. The procedure described in 6.4 is followed.

10.3.1 The recovery critical value is computed and is shown in Fig. 7:

$$YC = k1 \times \hat{s}(0) + a = 2.74 \times 1.089 + 2.73 = 5.71 \quad (28)$$

where:

k1 = 2.74 = the one-sided statistical tolerance limit for 90 % confidence of the 99 % quantile, based on the normal distribution assumption and n = 50 observations,

$\hat{s}(0)$ = g = 1.089 is the predicted ILSD at T = 0 (blank samples), and

a = 2.73 = intercept from the mean recovery curve (recall that this is set much higher than 0 for this example to clearly distinguish measured values from true values in the plots and tables; see 10.1.1).

10.3.2 The true concentration critical value is computed and is shown in Fig. 7:

$$LC = (YC - a)/b = (5.71 - 2.73)/5.87 = 0.51 \text{ ppb} \quad (29)$$

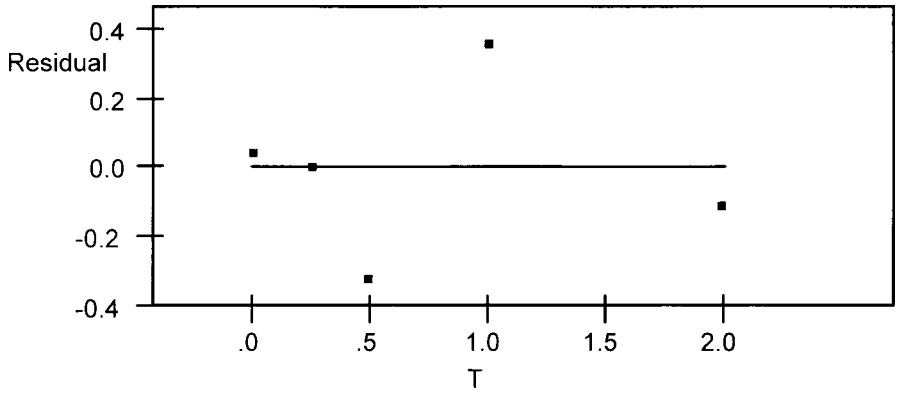

FIG. 6 Residuals from Straight-Line Model of Interlaboratory Measurement Standard Deviation versus True Concentration

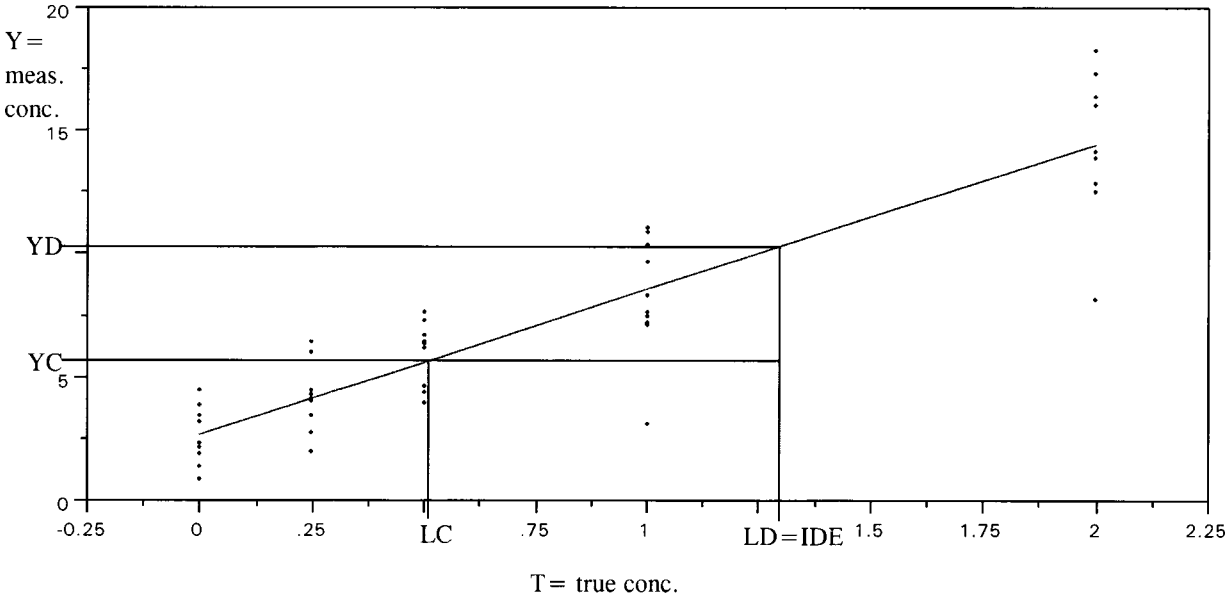

FIG. 7 Weighted Least Squares Fit of Mean Recovery Relationship, with IDE and Critical Limits

where: $b = 5.87$ = slope of the recovery curve (recall that this is set much higher than 1 for this example to clearly distinguish measured values from true values in the plots and tables (see section 12.1.1).

10.3.3 The IDE (also called the *LD*, in the tradition of Currie **(1)** is computed recursively. An initial value is set as follows:

$$LD_0 = LC + k2 \times \hat{s}(0)/b$$
$$= 0.51 + 1.97 \times 1.089/5.87$$
$$= 0.874 \quad (30)$$

where: $k2 = 1.97$ = one-sided statistical tolerance interval for 90 % confidence of the 95 % quantile, based on the normal distribution assumption and $n = 50$ observations: Then the recursive function is solved, iteratively, as follows:

$$LD_1 = R^{-1}(k1 \times \hat{s}(0) + k2 \times G(LD_0) + a)$$
$$= LC + k2 \times (g + h \times LD_0)/b$$
$$= 0.511 + 1.97 \times (1.089 + 0.957 \times 0.874)/5.87$$
$$= 1.154 \quad (31)$$

$$LD_2 = 0.511 + 1.97 \times (1.089 + 0.957 \times 1.154)/5.87 = 1.245 \quad (32)$$

etc., until convergence is achieved at about the eighth iteration, $LD_7 \approx LD_8 = 1.287$. Therefore, $IDE = LD \times$ (adjustment factor from Table 1) $= 1.287 \times 1.028 \approx 1.3$ ppb, as is shown in Fig. 7. Note that $LD > 2 \times LC = 1.02$.

10.4 Based on this study, there is (approximately) 90 % confidence that the analyte can be detected at least 95 % of the time at 1.3 ppb, and simultaneously that blank samples will result in nondetect no more than 1 % of the time.

NOTE 3—In this example the calculated IDE is less than most calculated standard deviation values in Table 4. This is because the data used for the example reflect high blank values and an unusually high recovery slope. This example serves to illustrate the utility of the practice even when such anomalous results are reported.

10.4.1 Also shown in Fig. 7 is the expected measurement value at the IDe concentration:

TABLE 6 Numerical Results of WLS to Fit the Straight-line Mean Recovery Relationship Between Measured Concentration and True Concentration

Response: *Y*-Linear Fit					
Y = 2.729549 + 5.8711952 T					
Y = a + b T					
Summary of Fit					
RSquare		0.794662			
RSquare Adj		0.790384			
Root Mean Square Error		0.982227			
Lack of Fit					
Source	df	Sum of Squares	Mean Square	F Ratio	
Lack of fit	3	0.789330	0.26311	0.2601	
Pure error	45	45.519596	1.01155	Prob>F	
Total error	48	46.308925		0.8537	
Parameter Estimates					
Term	Estimate	Standard Error	T Ratio	Prob> *T*	
a (intercept)	2.729549	0.264938	10.30	<0.0001	
b (slope)		5.8711952	0.430774	13.63	<0.0001
Analysis of Variance					
Source	df	Sum of Squares	Mean Square	F Ratio	
Model	1	179.21612	179.216	185.7606	
Error	48	46.30893	0.965	Prob>F	
C total	49	225.52504		<0.0001	

$$YD = R(LD) = a + b \times LD = 2.73 + 5.87 \times 1.287 = 10.3 \quad (33)$$

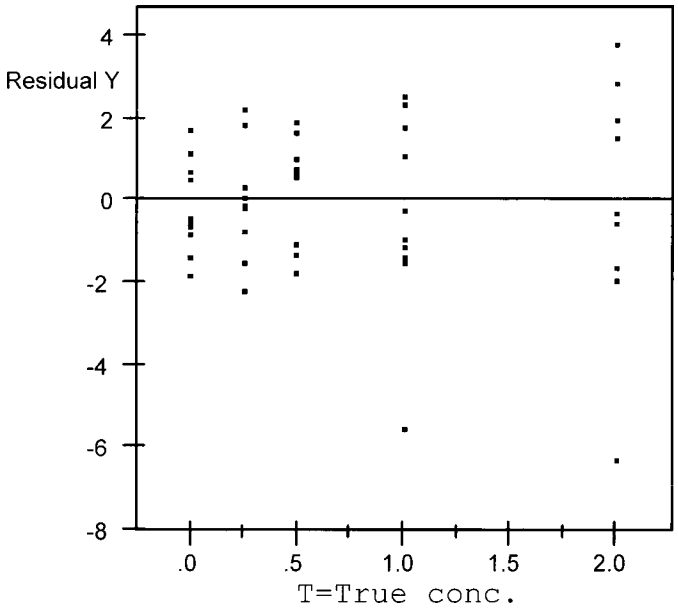

FIG. 8 Plot of Residuals from WLS Fit of Straight-line Mean Recovery Relationship versus True Concentration

11. Keywords

11.1 critical limit; detection; detection limit; false detection; false nondetection; false positive; matrix effects; statistical tolerance limit; true detection; true nondetection

ANNEX

(Mandatory Information)

A1. ANNOTATED OUTLINE FOR ANALYSIS REPORTS

A1.1 This outline presents the information to be included in the reports of analysis performed in accordance with this practice.

A1.2 *Single-Laboratory IDE Report*:

A1.2.1 Identification of laboratory, identification of analytical method, analyte(s), matrix (or matrices), sample properties (for example, volume).

A1.3 Any anomalies in the study, including QA/QC sample results.

A1.4 99 %/95 % Interlaboratory Detection Estimate Report.

A1.4.1 Data screening results, individual values and laboratories omitted from further analysis, and missing values.

A1.4.2 The ILSD model selected.

A1.4.3 Coefficient estimates for the ILSD model and mean recovery model.

REFERENCES

(**1**) L. A. Currie, ed., *Detection in Analytical Chemistry—Importance, Theory, and Practice*, American Chemical Society, Chapter 1, 1988, p. 10.

(**2**) Hahn, G., and Meeker, W., *Statistical Intervals—A Guide for Practitioners*, Wiley, New York, 1991, pp. 34–36, 41–52.

(**3**) Draper, N., and Smith, J., *Applied Regression Analysis, 2nd Edition*, Wiley, New York, 1981, pp. 33–40.

(**4**) Caulcutt, R., and Boddy, R., *Statistics for Analytical Chemists*, Chapman and Hall, New York, 1983, pp. 79–85, 100–110, 114–123.

(**5**) Carroll, R. J., and Ruppert, D., *Transformations and Weighting in Regression*, Chapman and Hall, 1988, Chapter 3, pp. 63–114.

(**6**) Rocke, D. M., and Lorenzato, S., "A Two-Component Model for Measurement Error in Analytical Chemistry," *Technometrics*, Vol 37, No. 2, May 1995, pp. 176–184.

(**7**) Maddalone, R., Rice, J., et al, "Defining Detection and Quantitation Limits," *Water Environment and Technology*, Vol 5, No. 3, January 1993, pp. 41–44.

(**8**) Gibbons, R., *Statistical Methods for Groundwater Monitoring*, Wiley, New York, 1994, pp. 95–121.