A Simple Definition of Detection Limit

Charles H. PROCTOR

The detection limit of an analytical method tells how low a concentration can be said to be measured. Users, laboratories, and equipment manufacturers prefer low ones. Industries under environmental regulation insist on honest ones. However, the detection limits of environmental regulations generally refer to relatively complex detection experiments (assays) or surveys that differ from the simple measurement operation that gave rise to the detection limit concept in analytical chemistry. The first contribution of this article is to clarify the limited scope of the simple definition and then define the population quantity that is being estimated. Such specificity allows alternative estimates to be evaluated more on statistical grounds of bias and variance rather than on whether they push the resulting value up or down. The fundamental statistical problem is to fit a variance function so that the derived relative standard deviation (RSD) function of concentration is smooth and monotone decreasing and thus can be inverted. The solution is to accept the best-fitting analytical version of the RSD function (log-log and hybrid models are described), fit it, invert it to get the point estimate, and provide a jackknife standard error.

Key Words: Analytical chemistry; Imputing censored observations; Interlaboratory study (ILS); Jackknife standard error; Multivariate calibration; Quantitation limit; Relative standard deviation (RSD).

1. INTRODUCTION

After a brief review of the history of the detection limit concept that distinguishes the assay experiment setting from the interlaboratory study (ILS), this article focuses on the ILS. The statistical model for ILS data provides reproducibility standard deviations for each material from which relative standard deviations (RSDs) are obtained. These lead to the RSD function that permits a detection limit to be defined. However, in order to fit the population RSD function one must choose some analytical function as there are too few data points to use less-structured functions. Two candidates are offered, a log-log and a hybrid, and the corresponding population detection limits are then defined as the goals of estimation.

The suggested estimators follow the same calculations as the definitions. Each of the two models provides fitted values to match the empirical RSDs and the closeness of fit

Charles H. Proctor is Professor Emeritus, Department of Statistics, North Carolina State University (E-mail: *chproctor@msn.com*).

Journal of Agricultural, Biological, and Environmental Statistics, Volume 13, Number 1, Pages 99–120
DOI: 10.1198/108571108X273476

for those materials having concentrations near the detection limit can be compared. In addition, a Tukey jackknife calculation of standard error is described that also allows the estimators to be evaluated. Three examples of ILS data plus two single laboratory experiments illustrate these considerations.

2. BACKGROUND

When Kaiser (1970a, 1970b) proposed a definition of the detection limit of an analytic method, he applied it to the instrument response of the method rather than to the measured concentration. The definition itself was that concentration for which relative measurement error standard deviation is 1/3. Since most analytical chemistry measurement involves converting instrument responses (e.g., light intensities) to concentrations, chemists are generally at home with instrument responses. However, the calibration function that maps instrument responses to concentrations is often nonlinear and if the method was recently developed it may not be as yet agreed upon. Additionally, the measured concentration ("net concentration") may be taken to be an amount above background and determining background can be problematic. Thus, although determining the detection limit first as an instrument response and second transforming it by the calibration function to a concentration may seem simple, it is not. The solution adopted herein is to include all aspects of calibration and determining background, in the measurement protocol so that the (net) concentrations of interest are the output of measurement. Under these conditions the definition begins to be simple.

Kaiser's (1970b) choice of three was based on the Tchebycheff inequality which shows that any distribution shape cannot have more than 1/9 of the probability beyond three standard deviations from the mean. It is a distribution-free result and led to a valuable criterion of merit, the "detection limit," for a measurement method. Another measure of merit is "quantitation limit" that is also simply defined by replacing the 1/3 by 1/10. However, numerous other criterion quantities (e.g., "decision limit") introduced by Currie (1968) and others (e.g., "LLD," "IDL," "MDL," and "LOD") and reviewed by Currie (1988), are not simple.

Their definitions are based on a detection experiment, also called an assay; see, for example, Davidian, Carroll, and Smith (1988) and Cox (2005). A detection experiment yields instrument responses from standard reference materials as well as from the unknown material. The data are used to test the statistical null hypothesis of zero (or only background) concentration in the unknown with prescribed Type I and Type II error probabilities. The detection limit is defined as the concentration at which, with probability equal to $1 - \beta$, the recorded instrument signals exceed the critical point of a one-sided α-level test. When responses follow the normal distribution with known standard deviation and the test is for $\beta = 0.10$ and $\alpha = 0.05$, then this detection limit is at 2.927 standard deviations so the definitions ($3 \approx 2.927$) appear to coincide. This approximate equality can be misleading since in the first case the inference is estimation and the result is a number that, as characterized by Lehmann (1991), "one hopes will be close to" the parameter, in the second ("detection experiment") case the inference is hypothesis testing and the result is a number

that locates the division between the "accept" region and the "reject" or, as characterized again by Lehmann (1986), the "critical region" of sample space.

The assay detection limit will depend on the size of the experiment (number of reference materials, number of duplicates at the reference materials and number of duplicates of the unknown) as well as on the distribution of (possibly transformed) instrument responses, on the calibration curve and on the correction, if any, for background. It is necessary to specify β and α. Sometimes a further tolerance probability is added. The provenance of the unknown material, which has been collected by some sampling operation, can be important, particularly in environmental applications. All of these considerations complicate such definitions.

Upon reviewing uses of detection limits in environmental monitoring by the EPA (Oct. 2004) and by Gibbons and Coleman (2001) one is struck by the rich variety of calculations proposed and the relative shortage of formal statistical inferential machinery, especially by the lack of a well-defined parameter space. That is, the definitions of detection limit in the regulatory settings are as estimates, not parameters. A term used in the EPA review is "approach," which refers to a scheme for calculation and rules for making use of the result in enforcing regulations and settling disputes. In order to convert data into regulatory actions, statistical theory is admittedly not necessary, but could be helpful in reducing the amount of trial and error expended in finding workable "approaches."

In sum, detection limit as defined herein is a measure of merit for an analytic method (specifically a concentrations measurement) governed by a detailed protocol. The analytic method accepts material as a laboratory sample and returns the measured (estimated) concentration as the test result. All details of converting instrument responses into concentrations are handled by the protocol as are the procedures for collecting the material into the laboratory sample.

To find such a detection limit requires collecting ratios of empirical standard deviations divided by concentrations (RSDs) that pass from above 1/3 to below as the concentrations increase away from zero. Two important experimental situations that provide such ratios are interlaboratory studies (ILSs) and single laboratory calibration experiments; we will examine data from some of those to illustrate estimation procedures. The final set of data is a multivariate calibration, which offers particularly complicated problems, as reviewed by Olivieri et al. (2006) for determining a limit of detection based on hypothesis testing, but the simple definition seems applicable.

3. STATISTICAL MODEL AND PARAMETER DEFINITION

The present work was motivated by an attempt to write a standard for calculating a detection limit that would take advantage of the statistical style of inference, but it becomes necessary to clearly distinguish between population quantities and sample quantities. In order to do this one needs many letters, Greek and Latin, and the notation becomes extensive. The reader is forewarned of the rather heavy notation and Appendix A, a guide to notation, should help.

For concreteness we begin with a particular experiment, the interlaboratory study (ILS),

also known as a collaborative study or round robin. The ASTM standard E 691 (2006) describes the conduct and data analysis for such a study. A number of laboratories agree to perform a (often a recently developed) measurement method on a collection of materials prepared to be homogeneous and distributed as laboratory samples. There are commonly eight or more laboratories and five or more materials, while each laboratory makes two or three duplicate measurements on each material. The laboratories are treated as a random sample from those qualified and the materials are chosen to range over those likely to be faced by users of the method. The purpose of an ILS is to verify the practicality, and determine the bias and precision, of the measurement method for each of the materials so that users may have confidence in the results from the method when performed by a typical laboratory.

The data from an ILS are balanced, with L laboratories, M materials, and D duplicates so the total number of data points is $n = LMD$. The model equation is

$$y_{ijk} = \mu_j + B_{ij} + \varepsilon_{ijk}, \text{ for } j = 1 \text{ to } M, \ i = 1 \text{ to } L, \text{ and } k = 1 \text{ to } D, \tag{3.1}$$

where y_{ijk} is the concentration measured on the jth material by the ith laboratory at the kth duplicate. The μ_j are fixed concentration constants, while the B_{ij} and ε_{ijk} quantities are taken to be random over laboratories and over duplicates with (heterogeneous) variances σ_{jL}^2 and σ_{jr}^2. The variance of the ε_{ijk} is known as repeatability variance, while the sum $\sigma_{jL}^2 + \sigma_{jr}^2$ is known as reproducibility variance and written σ_{jR}^2. The population RSD ratio for material j is σ_{jR}/μ_j which, as a population RSD quantity, will be written η_j.

When RSD η_j is plotted on the ordinate and concentration μ_j on the abscissa the scatter may suggest a smooth decreasing function. The RSD ratios for concentrations close to zero will be large ($>1/3$), due to division by a near-zero concentration, and the RSD in its normal operating range will be smaller ($<1/3$), but deciding on some specific smooth function can be problematic.

The variety of materials may disrupt the regularity of the results. In a worst case the materials may not bracket the detection limit and consequently the detection limit will not be defined. Measurement of low-level concentrations is much affected by the matrix, which is the substances included with the laboratory sample, other than the analyte of interest, and the materials may differ in matrix, as well as in concentration. Thus, as concentration and matrix change one may also find abrupt changes in measurement standard deviation.

In any case there are M materials and thus only M points in the population plot. The smoothing method is left undefined but the population RSD curve must be monotone decreasing in order that inverting it when $\eta = 1/3$ gives a unique concentration, γ_{DL}, which defines the population detection limit. Admittedly this definition suffers from the vagueness of the smoothing operation but it seems more realistic to keep the population definition for γ_{DL} free from any particular function.

Formally then, one supposes there exists a function $f_\eta(.)$ of concentration μ such that

$$\sum_{j=1}^{M} \left(\eta_j - f_\eta(\mu_j)\right)^2 \text{ is a minimum over choices of smooth decreasing functions } f_\eta. \tag{3.2}$$

Population detection limit is defined as $\gamma_{\mathrm{DL}} = f_\eta^{-1}(1/3)$. Ratios other than 1/3 can be

inserted to define related indices such as quantitation limit, or the "determination limit" of Currie (1968), which is $\gamma_{\mathrm{QL}} = f_{\eta}^{-1}(1/10)$.

4. ESTIMATION STRATEGIES

Two candidates for practical use as RSD function are a log-log function used by Horwitz (1982) and the hybrid function suggested by Kanzelmeyer as cited in Gibbons and Coleman (2001, p. 87). The log-log function in basic form is

$$\log(\eta) = \alpha + \beta \log(\mu), \tag{4.1}$$

where η is RSD and μ is concentration. The Horwitz curve has been faulted by Hamaker (1987) as well as by Rocke and Lorenzato (1995) because it implies that at zero concentration measurement error standard deviation must be zero, which is empirically not the case. The definition of the $\mathrm{RSD}_{\mathrm{LL}}$ function to be used here will thus be the following,

$$\begin{aligned} \log(\eta) &= \log\left(\mathrm{RSD}_{\mathrm{LL}}(\mu; \alpha, \beta, \gamma_0, \gamma_{\min})\right) \\ &= \alpha + \beta \log(\mu) \text{ for } \gamma_0 < \mu \leq \gamma_{\min}, \text{ and} \\ &= \alpha + (1+\beta)\log(\gamma_0) - \log(\mu) \text{ for } 0 < \mu \leq \gamma_0. \end{aligned} \tag{4.2}$$

This extended version effectively adjoins a linear segment to the standard deviation function for concentrations less than some γ_0 near zero so they have the constant standard deviation usually observed there. It means that the $\mathrm{RSD}_{\mathrm{LL}}$ function becomes a hyperbola near zero which merges smoothly with the power curve beyond the γ_0 concentration. The upper bound $\gamma_{\min}$ is undefined so far, but will be estimated from the data as the lowest concentration showing a positive first difference. Inverting $\mathrm{RSD}_{\mathrm{LL}}$ at 1/3 leads to $\gamma_{\mathrm{DL,LL}}$, the population LL detection limit parameter. Although $\gamma_{\mathrm{DL,LL}}$ is not exactly equal to γ_{DL} they are nearly equal so long as the $\mathrm{RSD}_{\mathrm{LL}}$ function fits the data reasonably well and this can be assessed in actual cases.

The alternative function, $\mathrm{RSD}_{\mathrm{Hy}}$, was developed from Kanzelmeyer's observation that many sets of concentrations data showed constant standard deviations near zero and then increasing ones further from zero. He suggested that the model for reproducibility standard deviation, σ_R, and thereby for RSDs should be

$$\sigma_R = \sqrt{\phi + \gamma\, c^2}, \text{ from which } \mathrm{RSD}_{\mathrm{Hy}}(c; \phi, \gamma) = \sqrt{\phi c^{-2} + \gamma}\,. \tag{4.3}$$

Here c is concentration and the φ and γ (without subscript) are the only two parameters, although it may be found useful to restrict the fitting by the same $\gamma_{\min}$ bound as for the log-log function. When $\mathrm{RSD}_{\mathrm{Hy}}$ is set equal to 1/3 and inverted to give a concentration, the result is $\gamma_{\mathrm{DL,Hy}}$, a population detection limit.

5. ESTIMATION DETAILS

Routine data analysis for an ILS calls for calculating, for each material, a reproducibility standard deviation and this is the key quantity for determining detection limit. The

calculation requires material averages, $\bar{y}_i = \sum_{j,k} y_{ijk}/LD$, and the laboratory averages, $\bar{y}_{ij} = \sum_{k=1}^{D} y_{ijk}/D$, and then the estimates of repeatability and reproducibility variances (in accord with ASTM E 691, 2006) are the ANOVA ones,

$$s_{jr}^2 = \sum_{i=1}^{L}\sum_{k=1}^{D}\left(y_{ijk} - \bar{y}_{ij}\right)^2 / L(D-1) \text{ to estimate repeatability variance,}$$

$$s_{jL}^2 = \max\left(0, \frac{\sum_{i=1}^{L}\left(\bar{y}_{ij} - \bar{y}_j\right)^2}{L-1} - \frac{s_{jr}^2}{D}\right) \text{ to estimate laboratories variance, and}$$

$$s_{jR}^2 = s_{jr}^2 + s_{jL}^2 \text{ to estimate total or reproducibility variance for the } j\text{th material.} \quad (5.1)$$

Consistent RSD estimates are given by $\mathrm{RSD}_j = s_{jR}/\bar{y}_j$. In some cases the materials have been prepared as standard reference materials (SRMs) and have a certified value, say c_j, which can, and normally should, be substituted for the "consensus" value $\bar{y}_j$.

Estimates of α and β of (4.1) are obtained by regressing logs of RSD estimates on logs of concentrations. Ordinary least squares estimates, a and b, can be based on $u_j = \log(\mathrm{RSD}_j)$ and $v_j = \log(\bar{y}_j)$ [or $v_j = \log(c_j)$ if reference concentrations are available] as

$$b = \sum_{J=1}^{M}(u_j - \bar{u})(v_j - \bar{v}) / \sum_{j=1}^{M}(v_j - \bar{v})^2, \text{ and}$$

$$a = \bar{u} - b\bar{v}, \text{ where } \bar{u} \text{ and } \bar{v} \text{ are means.} \quad (5.2)$$

One exponentiates the empirical prediction equation, where now c (without subscript) is measured concentration to get

$$\mathrm{RSD} = \exp(a)c^b. \quad (5.3)$$

The estimated detection limit is thus $c_{\mathrm{DL,LL}} = (\exp(-a)/3)^{1/b}$. These prosaic computational details are presented to prove the simplicity of the calculation, which is well within the capabilities of a hand-held calculator or a simple statistical package.

Notice that no specific distribution is assumed for the y_{ijk} although in practice measurement error distributions are close to normality. Thus, no distribution is expected for the RSD_j but again normality is not too far from reasonable. In certain ILS the range of materials may be considerably higher than the detection limit and one may find the RSDs increasing with concentration, although the log-log function follows a monotone decrease. In such cases the function should be fit only over the range 0 out to some cutoff concentration, $\gamma_{\min}$, at which the minimum RSD appears to fall. The definition (4.2) includes this fourth parameter.

If one of the materials (say $j = 1$) is a blank and its reproducibility standard deviation, s_{1R}, can be calculated for the blank, then a serviceable estimate of γ_0, the concentration near zero with the same measurement errors as for the blank, is found as

$$c_0 = \left(\exp(a)s_{1R}\right)^{1/(1+b)}. \quad (5.4)$$

As mentioned earlier, the upper limit $\gamma_{\min}$ is estimated as $c_{\min}$ equal to the lowest concentration at which a positive first difference appears. Once having the estimates c_0,

$c_{\min}$, a, and b the consistent estimator $c_{\mathrm{DL,LL}}$ becomes

$$\begin{aligned} c_{\mathrm{DL,LL}} &= 3\exp(a)c_0^{1+b} \text{ for c} \le \text{c}_0 \\ &= (3\exp(a))^{-1/b} \text{ for } c_0 \le c \le c_{\min}. \end{aligned} \tag{5.5}$$

In many cases b will be near -1 since this occurs whenever s_{jR} is fairly constant for the materials near zero concentration and when many near-zero materials were included. This leads to $c_{\mathrm{DL,LL}} = 3s_{1R}$. Such a simple estimate will have a coefficient of sampling variation determined by that of s_{1R}. If there is only one duplicate, then s_{1R} is just a sample standard deviation on L observations and will have a coefficient of sampling variation $\mathrm{CV}(c_{\mathrm{DL,LL}}) \approx 1/\sqrt{2L}$. Continuing with this crude level of approximation suggests that $c_{\mathrm{DL,LL}}$ of (5.5) is an average of a number of standard deviations. After materials beyond a judged $\gamma_{\min}$ are culled and detection limit is estimated by the log-log fit on say M' materials then the approximation

$$\mathrm{CV}(c_{\mathrm{DL,LL}}) = \frac{1}{\sqrt{2LM'}} \tag{5.6}$$

may not be far from correct. In the following examples, the CV will be better estimated by a Tukey (1958) jackknife calculation and it can then be seen how realistic is this approximate formula.

The hybrid model (4.3) is fit iteratively to observed RSD_j by minimizing the sum of squared deviations. Estimates obtained from the fitting will be denoted h_2 for φ and g_2 for γ so that the estimate of γ_{DL} becomes

$$c_{\mathrm{DL,Hy}} = \sqrt{\frac{h_2}{\frac{1}{9} - g_2}}. \tag{5.7}$$

Starting values are set small at, for example, $h_{20} = 0.001$ and $g_{20} = 0.001$. For each RSD_j its deviation is calculated by (4.3) as

$$\mathrm{dev}_j = \mathrm{RSD}_j - \left(h_{20}c_j^{-2} + g_{20}\right)^{0.5}. \tag{5.8}$$

Two partial derivatives are calculated as

$$\begin{aligned} f_{1j} &= -0.5\left(h_{20}c_j^{-2} + g_{20}\right)^{-0.5} c_j^{-2} \quad \text{and} \\ f_{2j} &= -0.5\left(h_{20}c_j^{-2} + g_{20}\right)^{-0.5}. \end{aligned} \tag{5.9}$$

The multiple regression with dev as dependent variable and f_1 and f_2 as independent variables, but with no intercept, finds two regression coefficients to add, respectively, to g_{20} and to h_{20} to produce improved estimates g_{21} and h_{21}. The process is continued until convergence and more computational guidance is found in Gallant (1987). The SAS code is given in Appendix C.

This model (4.3) for the measurement standard deviation function also results, as shown by Rocke and Lorenzato (1995), from supposing a particular distribution for the individual measurements. Maximizing this likelihood is the correct estimation procedure when data

follow the model. However, for routine use, the computer programs of Appendix C will be adequately accurate.

The relative advantages of the two RSD function models (Horwitz or log-log versus the Kanzelmeyer or hybrid curve) will be discovered only when one matches particular sets of empirical RSD patterns to the two somewhat different theoretical RSD curves. As the examples will show, the log-log fit will be confined to just the lower concentrations once $\gamma_{\min}$ is located, while the hybrid model will usually model standard deviations further along the concentrations range and may serve to give a better quantitation limit. On the other hand, the log-log fit, once the linear segment is attached, will almost always fit more closely and realistically over the range where the detection limit is to be found. As will be shown in the following, the fit of the hybrid model can sometimes be improved by omitting the blank material.

Both models are relatively straightforward when ordinary least squares is used and more elaborate versions with weighted least squares hardly seem worthwhile. Both models wisely propose an analytical function rather than any less parametric curve because this is a situation where there are relatively few data points. Unfortunately, there is no end of possible analytical functions so there is no end of candidate population detection limits, but fortunately any that fit at all well will yield essentially equal detection limits.

6. FURTHER COMPUTATIONAL CONSIDERATIONS

6.1 NUMERICAL CONVENTIONS

If the materials were prepared as reference materials, then those known concentrations will be used for calculating the RSD_j. If one of the materials was prepared to have zero concentration, then for computational purposes the concentration will be taken as 0.0001 or some other number of zeroes that just breaks it away from zero and signals that the material is a blank. If the material is "found" and the concentration is not known exactly, then the consensus value, the unweighted mean over all determinations on the material, is used. If this mean turns out to be zero or less, then a 0.0001 or similar value is assigned to signal a blank. These rules are computational conventions that avoid halting the computer because an RSD_j is not positive.

6.2 OUTLIERS

Analysis of ILS data can be complicated by the presence of outliers. These are values lower or higher than one might expect from normally variable data. Many ASTM committees and other standards organizations operate with special rules for detecting and dropping outliers. In fact it is rare to see ILS data with eight or more laboratories without noticing one outlier laboratory. A row-column model described by Mandel (1995) and associated diagnostic tests from Proctor (2000) are available to guide in reviewing ILS data; the SAS code needed for these calculations can be found in Appendix C.

6.3 IMPUTING FOR MISSING NEGATIVES

Almost any invocation of reporting limits wastes information whenever an estimated concentration could have been provided—even when it is negative. In order that the average of measured concentrations on the blank will be zero requires that roughly half of them will be negatives. When these negatives are correctly included all formulas will work properly. However, some instruments will automatically read them as zero or technicians will report missing. If other materials in the ILS bracket the detection limit such a blank can be omitted. However, if the reproducibility standard deviation of the blank is needed, then imputation can be used. Appendix B gives the background of two imputation methods and compares them over some simulations. It turns out that nonparametric imputation serves well for duplicates, but parametric imputation does best for imputing laboratory means.

6.4 COMPUTER CODE

Both the log-log and the hybrid models can be fit from the listing of reproducibility standard deviations and concentrations that is routinely provided as the precision statement from the ILS. Computer instructions in SAS code are given in Appendix C for fitting both models. The log-log model can be fit using a hand-held calculator that does linear regression, but the nonlinear hybrid model requires a computer.

6.5 JACKKNIFE STANDARD ERRORS

Once programming is done that allows converting an ILS precision statement into a detection limit, then it is a simple further step to calculate standard errors via the Tukey (1958) jackknife. One removes data from one laboratory and then estimates detection limit from the remaining data. This is done for each laboratory. In each case a pseudo-value is calculated by multiplying the overall estimate by L (the number of laboratories) and subtracting $(L-1)$ times the part-data estimate. The rationale is that each estimate acts as a mean so that multiplying it by L or by $(L-1)$ yields a sum and the difference is thus a single observation. Denote by s the sample standard deviation of pseudo-values so that $s/\sqrt{L}$ is the (jackknife) standard error of the estimated detection limit. This standard error is to be preferred to the crude approximation of (5.6), but comparing them can be instructive.

6.6 SUMMARY OF PRINCIPAL STEPS IN CALCULATING AND EVALUATING $c_{\mathrm{DL,LL}}$

The following brief outline will be followed for each of the examples.

1. Obtain a listing of concentrations (c_j or $\bar{y}_j$) and (reproducibility) standard deviations, s_{jR}, for the materials.

2. Compute $\mathrm{RSD}_j = s_{jR}/c_j$ and examine the sequence of decreasing RSDs from the material of lowest concentration to the highest and locate $c_{\min}$ between first two concentrations showing increase.

3. Take logs (omitting blank if there is one) and apply (5.2) to get a and b.

4. Calculate c_0 from (5.4) and then $c_{\mathrm{DL,LL}}$ by (5.5).

5. If the data are from an ILS, then apply the jackknife calculation to get $s/\sqrt{L}$, the standard error of $c_{\mathrm{DL,LL}}$. Divide this standard error by $c_{\mathrm{DL,LL}}$ and compare it to CV from (5.6) to evaluate adequacy of the estimate $c_{\mathrm{DL,LL}}$.

7. ILS EXAMPLE CALCULATIONS

7.1 CHLOROBENZENE IN REAGENT WATER

Fifteen laboratories participated in an ILS to study a capillary column gas chromatography measurement method for determining concentrations of organic compounds in water. The data were reported in ASTM standard D 5790 (2006) and the chlorobenzene in reagent water determinations allow estimation of a detection limit. Only one duplicate was run by each laboratory for each of eight materials. The "study coordinator" made a correction for background concentration. This represents a reduction in laboratory variability, but it was honestly reported and users of the precision statement must keep it in mind. A number of "outliers" were deleted in accord with the policy of that study that recognized that the method was in development and would likely do better than the present data showed. One laboratory, #14, was detected by other diagnostics as on average low but that is rather common to ILS datasets. One "outlier" observation was a "0.00" found for a material with a concentration of 1.00 μg/L, and it was reasoned that a digit had been miscopied. Thus a 1.0001 μg/L was imputed with the trailing digit to signal the imputation.

Table 1 has the data on four of the eight materials in the neighborhood of the detection limit having reference concentrations at 0.88, 1.10, 4.41, and 5.29 μg/L. The reproducibility standard deviations, in this case of just one duplicate, are simply the sample standard deviations for each column of data. The four materials' RSD_j's were found as: 0.527, 0.204, 0.109, and 0.156. The rise in RSD between the third and fourth materials signals that $\gamma_{\min}$ falls between 4.41 and 5.29 μg/L so only the first three materials will be fit under the log-log function.

Formula (5.2) gives $a = -1.09885$ and $b = -0.79247$ and formula (5.5) the value for $c_{\mathrm{DL,LL}} = 0.99970$. The lengthy decimals arise from using the SAS code of Appendix C. Dropping lab #1 and applying the estimation formulas to the part-data of 14 laboratories leads to $c_{\mathrm{DL,LL}} = 1.04214$ and thus the first pseudo-value becomes $15(0.99970) - 14(1.04214) = 0.40554$. The standard error from the complete collection of pseudo-values is 0.27273. The succinct rounded report on estimated detection limit thus becomes "1.00 μg/L $\pm$ 0.27μg/L."

When $\mathrm{RSD}_{\mathrm{Hy}}$ is fit one finds $h_2 = 0.12913$ as the estimate for φ and $g_2 = 0.009806$, the estimate of γ, so that by (5.7) $c_{\mathrm{DL,Hy}} = 1.129$. For this example the two estimates are quite close.

Table 1. Test results from ILS on chlorobenzene in reagent water from $L = 15$ laboratories, $M = 4$ materials (reference concentrations as column headings) and $D = 1$ duplicate, reported in ASTM, D 5790 (2006).

Laboratory	conc= 0.88	conc= 1.10	conc= 4.41	conc= 5.29
1	1.08	1.2400	4.45	5.71
2	2.35	0.9600	4.53	5.24
3	1.30	1.3000	4.90	6.80
4	1.20	1.4000	3.90	4.80
5	2.20	0.9300	4.90	4.00
6	1.21	1.1000	4.50	5.37
7	1.20	1.2000	4.40	4.90
8	1.10	1.0000	4.30	5.80
9	0.80	1.0001	5.30	5.50
10	1.30	1.7000	4.70	6.60
11	1.10	1.2000	4.10	5.30
12	1.00	1.3000	4.90	5.40
13	1.20	1.1000	4.80	5.60
14	0.55	0.7900	3.33	3.65
15	1.00	1.3000	4.70	5.80

7.2 CADMIUM IN WATER

An ILS for measurements of cadmium in water was done by five Michigan laboratories, applying inductively coupled plasma atomic emissions spectroscopy (ICP/AES) on three materials with five replicate measurements. The data of Table 2 were reported by Bhaumik and Gibbons (2005) whose purpose was not specifically to estimate a detection limit. Their aim was a confidence interval, but the data can serve to illustrate estimation of detection limit where the alternative hybrid model seems called for.

The precision statement for the data in Table 2 lists the reproducibility standard deviation for the blank as $s_{1R} = 3.91881$ and standard deviations of $s_{2R} = 4.17207$ and $s_{3R} = 7.67998$ for reference concentrations of 20 and 100. Taking logs of the RSD_j and of the c_j one can readily find the slope and intercept of (5.2) as $a = 0.2947$ and $b = -0.6215$. Application of the estimate for c_0 from formula (5.4) shows that $c_0 = 17$ μg/L. The estimated RSD at this concentration is $\text{RSD}_0 = 0.23$, which is below 1/3 and so the detection limit must be less than 17.0 μg/L. At concentration γ_0, which is estimated by c_0, the RSD curve changes from hyperbola below c_0 to power curve above c_0. The estimate $c_{\text{DL.LL}}$ equals three times the constant reproducibility standard deviation of s_{1r} or 11.76 μg/L. The hybrid model-based estimate turns out to be $c_{\text{DL.Hy}} = 12.00$ which is essentially the same. Comparing estimates of quantitation limit, $c_{\text{QL,LL}} = 65.62$ and $c_{\text{QL,Hy}} = 52.63$.

By dropping each laboratory's data in turn and re-estimating the parameters one finds the estimates 11.78, 13.11, 6.13, 13.19, and 13.19 as the five laboratories are dropped one at a time. The estimate overall was 11.76 and so for the first pseudo-value one finds $5*11.76-4*11.78 = 11.68$. The remaining pseudo values were 6.36, 34.28, 6.04, and 6.04 and the standard error of the means is found to be 5.46. Thus the estimate of detection limit

Table 2. Input Flat File of Data for Cadmium in Water Prepared in Standard Format for Entry to ILS Precision Statement Computer Code (Proctor 2000). (First line is title and second has codes telling format; in this case the 3 tells that duplicates are listed vertically with labs in rows, 5 labs, 3 materials, and 5 duplicates, followed by data.)

Cadmium (μg/L) in water by ICP/AES from Bhaumik and Gibbons (2005)

3 5 3 5

−3.000	10.000	92.000
4.000	20.000	100.000
−4.000	17.200	97.800
3.000	24.000	100.000
3.100	19.100	109.000
−0.060	17.815	90.455
0.010	17.305	87.610
0.115	16.570	85.550
−0.055	17.360	89.925
0.340	18.120	90.070
−7.400	27.100	107.400
−2.100	19.400	108.100
−11.400	9.000	83.800
−11.100	10.500	81.900
−1.400	19.300	94.200
1.000	21.000	96.000
−2.126	16.049	90.650
0.523	16.082	89.388
−2.000	17.000	91.000
−0.551	15.489	85.867
0.000	18.000	91.000
0.000	19.000	101.000
0.000	19.000	102.000
−1.000	18.700	92.700
0.038	19.790	99.884

is reported as 11.8 ±5.5 ppm. It is apparent that detection limit is not well estimated since its sampling CV = 47%. The approximation formula for CV (5.6) would have suggested, based on $M' = 3$ and $L = 5$, that CV = 18%.

Notice that lab #3 seems out of line. When laboratory #3 is dropped the precision statement gives $s_{1R} = 2.042$, $s_{2R} = 2.838$, and $s_{3R} = 6.639$ from which the estimates are much less—$c_{\mathrm{DL,LL}} = 6.13$ and $c_{\mathrm{DL,Hy}} = 6.28$. The corresponding Tukey standard errors were ±4.4 and ±3.8—also smaller.

By way of verifying how imputation might work, the negatives in Table 2 were replaced by zeroes and then imputed as described in Appendix C. Laboratory #3 was omitted. The resulting reproducibility standard deviation for the blank, s_{1R}, turns out to be 1.956 and agrees well with the original 2.042.

Table 3. Visual ratings (before conversion to concentrations in ng/L by taking 2.5 power and dividing by 10) on aflatoxin in peanut butter by thin layer chromatography. (Each set of three columns are duplicates for one of the four materials. Each row is a laboratory.)

0	0	0	3	3	3	4	5	6	6	7	7
0	0	0	0	3	5	3	5	5	5	5	7
3	5	7	4	4	5	5	7	7	5	7	7
0	0	0	3	4	5	5	6	7	7	7	7
0	0	0	2	3	4	4	5	5	6	6	7
0	0	0	3	4	4	0	5	5	7	7	7
0	0	0	4	6	6	5	6	6	7	7	7
0	0	0	3	4	4	5	5	5	6	7	7
0	0	0	3	3	4	5	5	6	6	7	7
1	2	3	4	4	5	5	5	5	5	7	7
0	0	0	2	2	3	4	5	5	7	7	7
0	0	0	3	3	4	5	5	5	6	7	7
0	0	0	4	5	5	5	5	5	6	7	7
0	0	0	3	3	4	5	6	6	7	7	7
0	0	0	4	4	4	6	6	7	7	7	7
0	1	2	4	4	5	5	6	6	6	7	7
0	0	1	3	5	5	3	5	7	7	7	7
0	1	1	3	4	4	4	5	5	7	7	7
0	0	0	4	4	4	5	5	6	7	7	7
0	1	0	2	3	3	2	4	5	6	6	7

7.3 AFLATOXIN IN PEANUT BUTTER EXTRACT

A third example of an ILS was reported by Beljaars et al. (1973). The analyte was aflatoxin B_1 in a matrix of peanut butter extract. The technique of access was thin layer chromatography and the "instrument" signals were visual ratings. The data appear in Table 3. The four materials were prepared to have 0, 3, 6, and 12 ppb of aflatoxin. The calibration function that the article provided was unrealistic in putting 1's for 0's but it turns out that taking a 2.5 power of the visual ratings and then dividing by 10 provides a reasonable calibration function. Final choice of the 2.5 power arose from trial and error but was initiated by fitting logs of reference concentrations on logs of average visual ratings. This regression suggested the 2.33 power, which was "rounded" to 2.5.

The integer ratings from 0 to 7 in Table 3 were transformed to the concentrations 0, 0.10, 0.57, and so on to $12.96 (= 7^{2.5}/10)$ and the routine ILS calculations were applied. There was some suggestion that laboratory # 3 was out of line and it would ordinarily be omitted. The blank material received many zero measurements that arose from the scoring scheme and imputation seems called for. These considerations permit the comparison of the two models (log-log versus hybrid) in three situations—(Case 1) full data, (Case 2) laboratory 3 omitted but no imputation, and (Case 3) laboratory 3 omitted and imputation supplied.

Including the results from laboratory #3 will inflate the estimate of the standard deviation of material #1, the blank. In that case its standard deviation will be similar to that of material #2 and such constancy is fit well by the Kanzelmeyer hybrid function. Thus for

Table 4. Precision statements for aflatoxin B1 in peanut butter, non-blank materials from 20 laboratories in first statement. All materials for 19 laboratories with imputation for blanks in second statement.

Material number	Consensus concentration $\bar{y}_j$	Reference concentration c_j	Reproducibility standard deviation s_{jR}	Relative standard deviation, RSD_J
2	3.0	3.0	1.8579	0.6193
3	6.3	6.0	2.8658	0.4776
4	11.8	12.0	2.2747	0.1896
1	–	0.0001	1.031	10310
2	3.0	3.0	1.87463	0.62488
3	6.3	6.0	2.36165	0.39361
4	11.8	12.0	2.07639	0.17303

Case 1 the estimates agree, $c_{\text{DL,LL}} = 7.05$ and $c_{\text{DL,Hy}} = 6.97$.

Once laboratory #3 is dropped, then the estimate of the standard deviation of the blank becomes small due to the censoring (causing many zeroes). In this case the standard deviation of material #1 is decidedly different from that for material #2 and the hybrid model fits poorly. In fact the estimate for g_2 was 0.1609, which caused $c_{\text{DL,Hy}}$ from (4.4) to become imaginary so the fitting was shifted from RSD to the s_{jR} and $c_{\text{DL,Hy}} = 4.79$. The log-log estimate remained at $c_{\text{DL,LL}} = 6.71$. The hybrid model predicts the three highest material's RSDs to be, 0.45, 0.28, and 0.22. The log-log model predicts 0.68, 0.37, and 0.20. The empirical RSDs (shown in Table 4) are 0.62, 0.44, and 0.18 and thus are much better fit by the log-log model.

After applying the nonparametric imputation scheme of Appendix B, the reproducibility standard deviation of the blank rises to $s_{1R} = 1.031$, but $g_2 = 0.11195$ and thus from fitting RSDs the hybrid estimate is imaginary. Upon fitting to the s_{jR} quantities the $c_{\text{DL,Hy}} = 5.58$, which is not far from the $c_{\text{DL,LL}} = 6.71$. Omitting the blank, a Case 4 situation, would allow $c_{\text{DL,Hy}} = 6.67$. Such a choice might have been made if imputation were mistrusted.

When the jackknife procedure is applied to $c_{\text{DL,LL}}$ the standard error of the pseudo values is 0.66. This detection limit estimate of 6.71 thus has about a 10% CV, which is in fair agreement with the (5.6) approximate CV $= 1/\sqrt{2M'L}$ for $M' = 2$ and $L = 19$. The jackknife standard error for the $c_{\text{DL,Hy}} = 5.58$ was found to be 1.17 for a 21% CV, which discourages its use.

8. SINGLE LABORATORY CALIBRATION EXAMPLES

8.1 EXAMPLE OF EPA'S METHOD DETECTION LIMIT APPROACH

As part of EPA's assessment of alternative methods for determining detection limit, a specimen dataset from a single laboratory was analyzed by various "approaches" and the results compared to EPA's MDL (method detection limit) approach. No method worked

Table 5. Reported data from EPA (Oct. 2004) Appendix C, Table 1 for 1, 1, 1, 2-tetrachloroethane in mu-grams per liter. [Columns are spiked concentration (conc), average recorded concentration (ave), and standard deviation of recorded (std), followed by two versions of RSD—first dividing by spiked concentration (rsd1) and then by average concentration (rsd2).]

Obs	conc	ave	std	rsd1	rsd2
1	0.010	0.0016	0.0018	0.18000	1.12500
2	0.015	0.0010	0.0017	0.11333	1.70000
3	0.020	0.0007	0.0010	0.05000	1.42857
4	0.035	0.0057	0.0036	0.10286	0.63158
5	0.050	0.0081	0.0024	0.04800	0.29630
6	0.075	0.0263	0.0202	0.26933	0.76806
7	0.100	0.0295	0.0039	0.03900	0.13220
8	0.150	0.0536	0.0046	0.03067	0.08582
9	0.200	0.0991	0.0156	0.07800	0.15742
10	0.350	0.2350	0.0078	0.02229	0.03319
11	0.500	0.3744	0.0257	0.05140	0.06864
12	0.750	0.6193	0.0262	0.03493	0.04231
13	1.000	0.8368	0.0814	0.08140	0.09728
14	2.000	1.9560	0.0980	0.04900	0.05010
15	5.000	5.0994	0.2382	0.04764	0.04671
16	10.000	10.4453	0.5469	0.05469	0.05236

as well as EPA's but much of the reason for this is found in the data. The data appear in Table 5 and consist of the "spike" or reference concentration (conc), and an average measured concentration (ave), both of these in μg/L (parts per billion), along with the standard deviations (std) of the measured concentrations from samples of seven duplicate determinations.

The MDL approach, as initiated by Glaser et. al. (1981), uses a "discovery" stage whereby judgment and experiences are combined to make a reasonable guess as to where the detection limit will occur and then focuses on only the data (based on the spiked concentration) at or nearby to that guessed concentration. This was the material with the 0.075 ppb spike at the putative detection limit. The method looked then to the 0.05 ppb spiked material and found the standard deviations too different from that at 0.075, but then used the 0.1 ppb spiked material's standard deviation. Even a brief glance at the series of standard deviations in Table 5 would suggest that the 0.0202 value is an outlier, but their approach was focused on that one. The method pooled the two standard deviations, 0.0202 from the 0.075 ppb and 0.0039 from the 0.1 ppb spiked case to get 0.015, which was then multiplied by the "k-value" (an equivalent to the three of the simple definition) of 2.71 to give the MDL at 0.041 ppb.

In accord with the fitting procedure of this article one checks to see that the RSD $= 1/3$ is bracketed and if one just looks at rsd1 $=$ std/conc, the answer is that no fit is possible since maximum rsd1 is only 0.18. The reason is the very poor performance of the measurement method; notice that ave (measured amount) is far from equal to conc

Table 6. Eye-determined percent inhibitions in four AChE biosensors (W303A, F345A, M301A, and WT) at five concentrations (10, 1.01, 0.10, 0.01 and 0.001 ng/L) of insecticide.

Obs	pirimi-mth conc	M301A/W303A	F345A	M301A	WT
1	10.000	92.0	82	93.0	19.0
2	10.000	92.0	82	93.0	20.0
3	10.000	93.0	83	92.5	20.0
4	1.010	81.0	20	85.0	−2.0
5	1.010	81.0	21	85.0	0.0
6	1.010	82.0	22	86.0	2.0
7	0.100	41.0	0	31.0	0.0
8	0.100	41.0	3	31.0	3.0
9	0.100	42.0	6	32.0	2.0
10	0.010	23.0	0	10.0	0.0
11	0.010	23.0	3	10.0	0.0
12	0.010	23.5	5	11.0	0.5
13	0.001	16.0	−2	3.0	2.5
14	0.001	18.0	2	3.0	3.0
15	0.001	21.0	5	4.0	3.0

(reference concentration) for the lowest materials. However, both std and ave are in the same metric (nonlinearly related to conc) so the RSD to trust is rsd2 = std/ave. Even more, the lowest eight materials were included in the log-log fit because the $\gamma_{\min}$ bound was placed beyond the dip to rsd2 = 0.08582. The detection limit must be estimated in the conc (reference) metric so the log-log fit is taken by fitting logs of rsd2 to logs of conc. One finds $c_{\mathrm{DL,LL}} = 0.060$ ppb, which is somewhat similar to EPA's MDL of 0.041 ppb. The $c_{\mathrm{DL,Hy}} = 0.057$ ppb.

8.2 A MULTIVARIATE CALIBRATION

The following example is of interest to see how the log-log fit performs as a generic definition of detection limit that an individual laboratory might use, even when calibration must be done within the laboratory. We put the term "detection" into a search of chemistry articles and out came, "Insecticide Detection through Protein Engineering of *Nippostrongulus brasilienseis* Acetylcholinesterase B," in the September 15, 2005, issue of *Analytical Chemistry*. The article, by Schulze et. al. (2005), reported that, " ... the detection limit (signal-to-noise ratio ≥ 3) could be decreased from 10 μg/L to a value as low as 1 ng/L."

The article presented "data" as plots of signals, AChE inhibition (%) for four enzymes, by concentration of insecticide (pirimiphos methyl). We used "eye fitting" to convert the heights and error bars to observations. The article mentioned "$n \geq 3$" and $n = 3$ was used in all cases. The resulting "data" for the four enzymes are in Table 6. We requested the actual data from the authors but have not received a response as yet.

These data require a multivariate calibration, but even with just one (univariate) signal

there are choices to make among computational methods. Two regressions are commonly used for getting estimated concentrations from instrument signals—(1) the classical regression of signal on concentration is inverted or (2) a regression of concentration on signal is used directly to predict the estimated concentrations. We favor the latter for its simplicity and ease when used in the multivariate mode. Brown (1982) mentioned that two experts, Eisenhart and Williams, favor (1) and two others, Krutchkoff and Hoadley, favor (2). In the multivariate setting one computes the prediction equation by regressing reference concentration on the four signals and then uses predicted concentrations as the "measured" or estimated concentrations from which to calculate standard deviations and thereby RSDs. SAS code for the calculations is included in Appendix C.

The approach is optimistic in that it is intended to mimic what one would find if unknown materials were to produce the signals which were then converted via the calibration curve to estimated concentrations and we were then privileged to know the true concentrations. Recognizing this possible "optimism" bias caused us to provide two versions of the calibration curve. One is based on just the four linear main effects and the other includes quadratic terms as well. The simple version had an $R^2 = 0.9937$ and the other had $R^2 = 0.999973$, which argues strongly that quadratic terms are necessary. After fitting logs the estimated detection limits were found as $c_{\mathrm{DL,LL}} = 0.678$ μg/L from the simple linear one and $c_{\mathrm{DL,LL}} = 0.043$ μg/L for the quadratic. The lower detection limit, at 43 ng/L, suggests that the authors were a bit optimistic, but by only one order of magnitude.

9. CONCLUSIONS

The main objective of this work has been to show how a return to the simplicity of Kaiser's definition of detection limit leads to simple computations and ones that should be of use in standards. The definition applies to an experimental setup that provides replicated results for a number of materials having concentrations near zero. The interlaboratory study (ILS) is archetypical, but calibrations experiments can also provide the $\mathrm{RSD}_j - c_j$ pairs to be fit as shown in the last two examples. Shortcomings of the experimental setup—such as materials not covering the low concentrations and laboratories not being "representative"—cannot be cured by any definition. However, a clear definition with an explicit statistical model allows better evaluation of the estimate through comparisons of the alternatives, being two at present the log-log fit versus hybrid fit, and comparisons between the potential approximate standard error and the realized jackknife standard error.

The log-log model has not been used, as far as we are aware, to calculate a detection limit but the hybrid model appears in two ASTM standards, D 6512 (2003) and D 6091 (2003). The Interlaboratory Quantitation Estimate (IQE) of D 6512 is simply defined and, in fact, estimates γ_{QL}. The 99%/95% Interlaboratory Detection Estimate (IDE) of D 6091 goes back to the assay experiment to require specification of error percents and is not simple. One could argue for dropping mention of the percentages and adopting the simple definition for the IDE. Then the problem might arise of choosing between the log-log and the hybrid.

It could be easily resolved if they were found to be equal (within a jackknife stan-

dard deviation say) and this would ordinarily be the case. A substantive issue might arise concerning the relative chemical realism of the curve determined by the α and β (with bounding γ_0 and $\gamma_{\min}$) of (4.2) versus the curve determined by the φ and γ (with possibly some bounding) of (4.3). These curves, and others reflecting the way measurement precision changes with concentration, are the result of laboratory economics, laboratory customs, laboratory equipment, and so on, as well as the result of the chemical reactions and there is no single best form for all applications. Having two versions, both relatively easily fit, would seem to be more than adequate for general use in standards.

The imputation schemes suffer from complexities and perhaps imputation should be avoided. However, use of nonparametric imputation on duplicates and parametric imputation of laboratory means deserves a trial. A secondary contribution of the nonparametric imputation method will be if it convinces instrument makers and chemists to avoid censoring. It is offered as the method of choice when a reproducibility standard deviation, s_{1R}, is needed for the blank material but negatives have been suppressed.

Detection limit definition and estimation should be viewed as primarily a problem of fitting an invertible variance function. Once assay experiments and surveys used in regulatory and commercial applications are characterized by their variances rather than by error probabilities then perhaps their statistics can be simplified just as this article has done for those from interlaboratory studies and calibration experiments. The issues of costs and benefits should be the primary concerns in regulatory and commercial applications, not the statistics.

APPENDIXES

A. GUIDE TO NOTATION

RSD = relative measurement error standard deviation (std. dev.), a ratio of a sample standard deviation divided by a concentration, both in concentration units.

y_{ijk} = measurement on jth material by ith lab of kth duplicate.

$\bar{y}_j$ = jth material average.

$\bar{y}_{ij}$ = average for jth material by ith laboratory.

L = number of laboratories in ILS.

M = number of materials in ILS but M' = number of materials fit under log-log model.

D = number of duplicate measurements in ILS.

μ_j = true concentration of jth material.

B_{ij} = random (additive) effect of ith laboratory for jth material in ILS model.

ε_{ijk} = random (additive) effect for kth duplicate replication by ith laboratory on jth material.

σ_{jL} = laboratory std dev. for jth material and s_{jL} = ANOVA estimate of σ_{jL}.

σ_{jr} = repeatability std. dev. for jth material and s_{jr} = ANOVA estimate of σ_{jr}.

σ_{jR} = reproducibility std. dev. for jth material and s_{jR} = ANOVA estimate of σ_{jR}.

dev_j = RSD_j-Predicted-RSD-from-Hybrid-Model, with f_{1j} and f_{2j} as derivatives.

σ_R = reproducibility std. dev. at concentration c.

$c = \mu$ = true concentration.

η = a ratio of a population measurement error standard deviation divided by the "true" concentration.

At page 2 only, α = probability of rejecting when null (zero) is true and β = probability of accepting when null is false.

η_j = population relative std. dev. (RSD) of jth material.

$f_\eta(c)$ = an invertible function giving fitted values of η at concentration c.

γ_{DL} = population detection limit such that $f_\eta(\gamma_{\text{DL}}) = 1/3$.

γ_{QL} = population quantitation limit such that $f_\eta(\gamma_{\text{QL}}) = 1/10$.

RSD_{LL} = log-log RSD function and RSD_{Hy} = hybrid RSD function.

α = log-log parameter and a = estimate of α.

β = log-log parameter and b = estimate of β.

γ_0 = Lower concentration under log-log curve marking transition between hyperbola and power curve and c_0 = estimate of γ_0.

γ_{min} = Upper range of concentrations under the log-log curve and c_{min} = estimate of γ_{min}.

φ = constant parameter of Kanzelmeyer or hybrid curve and h_2 = estimate of φ.

γ = multiplier parameter (of c^2 term) in Kanzelmeyer or hybrid curve and g_2 = estimate of γ.

$\gamma_{\text{DL,LL}}$ = population detection limit and $c_{\text{DL,LL}}$ = estimated detection limit by log-log fit.

$\gamma_{\text{DL,Hy}}$ = population detection limit and $c_{\text{DL,Hy}}$ = estimated detection limit by hybrid fit.

$s = \sqrt{\sum (pv_i - p\bar{v})^2 / (L-1)}$ = sampling std. dev. of pseudo-values pv_i.

$\text{CV}(c_{\text{DL,LL}})$ = sampling coefficient of variation of $c_{\text{DL,LL}}$.

$y_{(n/2)}$ = median of n censored or zero-or-less (ZOL) observations.

z_i = ith normal score.

B. IMPUTING CENSORED MEASUREMENTS

The simple definition requires that all measurements be on the concentrations scale. Reported measurements provided by the calibration curve can legitimately be negative and should be reported as such. However, when negatives appear they may sometimes get censored. They can be spotted in data as an excess of zeroes and were thus called zero-or-less (ZOL). This practice wastes about half of the data when estimating the measurement error standard deviation of the blank.

Such censoring is called type I by Kendall and Stuart (1961, vol. 2, pp. 522ff.) in that the cutoff is known (0 in this case) and the number censored is random. The nonparametric imputation scheme arises from supposing that the center of the distribution is at zero and that the distribution is symmetric. If the calibration curve is linear this implies that a "good" transformation has been found for instrument responses and the distribution of measurements can be expected to be roughly symmetric. Thus one uses the spacing of the known positives to be that for the unknown negatives.

The procedure, in five steps, is:

1. Order all observations, including zeroes, to give the order statistics for which $y_{(n)}$ is the largest and $y_{(n-1)}$ is the next largest, etc.

2. Find the median, $y_{(n/2)}$ say, Notice that $y_{(n/2)}$ may be zero.

3. Subtract twice the median from the largest, change sign, and impute it for the smallest, which gives $y_{(1)} = -(y_{(n)} - 2y_{(n/2)})$.

4. Subtract twice the median from the next largest, change sign, and impute it for the next smallest, which gives $y_{(2)} = -(y_{(n-1)} - 2y_{(n/2)})$.

5. Continue as far as possible.

The second scheme, based on normal scores, is not as "distribution free" since it supposes a normal distribution and will be called parametric imputation. The procedure first computes the n normal scores by Blom's (1958) formula,

$$z_i = \Phi^{-1}\left(\frac{i - 3/8}{n + 1/4}\right). \tag{B.1}$$

The Φ function returns cumulative probabilities under the normal distribution; the i are the integers from 1 to n, the number of ZOLs. The order statistics are regressed on the corresponding normal scores, omitting the zeroes, and the prediction equation then uses the normal scores corresponding to the omitted zeroes to predict their order statistics and the predictions become the imputed values.

Both schemes were applied to simulated standard normally distributed data. Sample sizes were set at 2, 3, 4, 5, 10, 20, 30, and 40 so as to cover usual values of D and L in ILS. In each case the original standard deviation without censoring was calculated and then the two standard deviations from the two schemes of imputation were found. Only 100 iterations were run for each of the sizes, but the pattern was quite clear. For the smaller

sizes of 2, 3, 4, and 5 the average estimates were 0.74, 0.83, 0.81, and 0.82 for the nonparametric scheme and 0.16, 0.37, 0.52, and 0.62 for the parametric one so nonparametric imputation seems better. For the larger sizes of 10, 20, 30, and 40 the averages were 0.93, 0.95, 0.94, and 0.95 for nonparametric imputation and 0.98, 0.98, 0.97, and 0.97 for the parametric method. Correlations between the original standard deviation estimate and the two imputation-based estimates were 0.72 and 0.29 for small sizes and 0.71 and 0.54 for larger sizes. That is, the nonparametric method was twice as efficient for small sizes and still somewhat more efficient even for larger sizes.

One may look upon the under-estimation as worrisome but generally it is not. It was found for normally distributed simulated data, while realistic measurement data are generally somewhat leptokurtic—they have longer tails and often are found to contain "outliers." This feature will tend to increase the estimates of standard deviation and thus overcome the underestimation.

The aflatoxin data of Section 7.3 had many ZOL and did suggest that nonparametric imputation will be downwardly biased if the average concentration of the blank, when negatives are reported, is negative. Simulations support this in that the average for nonparametric imputation dropped from 0.95 down to 0.65 for the size 40 case but for parametric imputation changed from 0.97 to 0.98, when simulations used a normal mean of -0.5. Thus parametric imputation is somewhat more robust to shift in mean. These results lead to recommending nonparametric imputation for duplicates and parametric imputation of laboratory means.

As an example of the calculations consider the ILS for cadmium in water in Section 7.2, in which there are $D = 5$ duplicates. The duplicates from the first laboratory were (Table 2) -3.0, 4.0, 3.1, 3.0, -4.0. Upon censoring, the negatives are set equal to zero and the observations ordered, whereupon the largest has its sign changed and replaces the final zero, the next largest has its sign changed and replaces the next zero and so forth. The resulting data with imputations become 4.0, 3.1, 3.0, -3.1, and -4.0.

After applying nonparametric imputation to duplicates from laboratories, except #3, the lab means are found to be 0.600, 0.002, and two zeroes. The four normal order scores are -1.049, -0.299, 0.299, and 1.049 so that imputation puts -0.475 and -1.073 for the zeroes.

C. SAS CODE FOR FITTING LOG-LOG MODEL AND HYBRID MODEL

SAS code is available online at *http://www.amstat.org/publications/jabes/upload/index.cfm?fuseaction=ViewIssues&pub=JABES*. This code can be copied into the SAS Program Editor and run to give estimates under both the log-log model and the hybrid model. Change places of the datasets to see all estimates.

[Received September 2006. Revised September 2007.]

REFERENCES

ASTM, Designation E 691 (2006), "Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method," *Annual Book of ASTM Standards*, 14.02, West Conshohocken, PA: ASTM.

ASTM, Designation D 5790 (2006), "Standard Test Method for Measurement of Organic Compounds in Water by Capillary Column Gas Chromatography/Mass Spectrometry," *Annual Book of ASTM Standards*, 11.02, West Conshohocken, PA: ASTM.

ASTM, Designation D 6091 (2006), "Standard Practice for 99%/95% Interlaboratory Detection Estimate (IDE) for Analytical Methods with Negligible Calibration Error," *Annual Book of ASTM Standards*, 11.01, West Conshohocken, PA: ASTM.

ASTM, Designation D 6512 (2006), "Standard Practice for Interlaboratory Quantitation Estimate," *Annual Book of ASTM Standards*, 11.01, West Conshohocken, PA: ASTM.

Beljaars, P. R., Verhulsdonk, C. A. H., Paulsch, W. E., and Leim, D. H., (1973), "Collaborative Study of Two Dimensional Thin Layer Chromatographic Analysis of Aflatoxin B1 in Peanut Butter Extracts, Using the Antidiagonal Application Technique," *Journal of the AOAC International*, 56, 1444–1451.

Bhaumik, D. K., and Gibbons, R. D. (2005), "Confidence Regions for Random-Effects Calibration Curves with Heteroscedastic Errors," *Technometrics*, 47, 223–230.

Blom, G. (1958), *Statistical Estimates and Transformed Beta Variables*, New York: Wiley.

Brown, P. J., (1982) "Multivariate Calibration," *Journal of the Royal Statistical Society*, Series B, 44, 287–307.

Cox, C. (2005), "Limits of Quantitation for Laboratory Assays," *Journal of the Royal Statistical Society, Applied Statistics*, Series C, 54, 63–76.

Currie, L. A. (1968), "Limits for Qualitative Detection and Quantitative Determination," *Analytical Chemistry*, 40, 586–593.

——— (1988), "Detection: Overview of Historical Societal, and Technical Issues," in *Detection in Analytical Chemistry* (chap. 1), ed. Currie, L. A. ACS symposium series, 361, Washington, DC: American Chemical Society.

Davidian, M., Carroll, R. J., and Smith, W. (1988), "Variance Functions and the Minimum Detectable Concentration in Assays," *Biometrika*, 75, 549–556.

EPA, U. S. Environmental Protection Agency (Oct. 2004), "Revised Assessment of Detection and Quantitation Approaches," EPA-821-R-03-005, 254 pages. Available online at *http://www.epa.gov/waterscience/methods/det/rad/rad.pdf*.

Gallant, A. R. (1987), *Nonlinear Statistical Models*, New York: Wiley.

Gibbons, R. D., and Coleman, D. E. (2001), *Statistical Methods for Detection and Quantification of Environmental Contamination*, New York: Wiley.

Glaser, J. A., Forest, D. L., McKee, G. D., Quave, S. A., and Budde, W. L. (1981), "Trace Analyses for Wastewaters," *Environmental Science and Technology*, 15, 1426–1435.

Hamaker, H. C., (1987), "Repeatability and Reproducibility: Some Problems in Applied Statistics," in *Design, Data, and Analysis by Some Friends of Cuthbert Daniel* (chap. 5), ed. C. L. Mallows, New York: Wiley.

Horwitz, W. (1982), "Evaluation of Analytical Methods Used for Regulation of Foods, and Drugs," *Analytical Chemistry,* 54, 1, 67A–76A.

Kaiser, H. (1970a), "Quantitation in Elemental Analysis," *Analytical Chemistry*, 42, 2, 24A–38A.

——— (1970b), "Part II Quantitation in Elemental Analysis," *Analytical Chemistry*, 42, 4, 26A–58A.

Kendall, M. G., and Stuart, A. (1961), *The Advanced Theory of Statistics, Volume 2: Inference and Relationship*, New York: Hafner.

Lehmann, E. L. (1986), *Testing Statistical Hypotheses* (2nd. ed.), New York: Wiley.

——— (1991), *Theory of Point Estimation*, Pacific Grove, CA: Wadsworth.

Mandel, J. (1995), "Structure and Outliers in Interlaboratory Studies," *Journal of Testing and Evaluation*, 23, 364–369.

Olivieri, A. C., Faber, N. M., Ferre, J., Boque, R., Kalivas, J. H., and Mark, H. (2006), "Uncertainty Estimation and Figures of Merit for Multivariate Calibration," *Pure and Applied Chemistry*, 78, 633–661.

Proctor, C. H. (2000), "Diagnostic Scores for Interlaboratory Study Data," *Journal of Testing and Evaluation*, 28, 307–334.

Rocke, D. M., and Lorenzato, S. (1995), "A Two-Component Model for Measurement Error in Analytical Chemistry," *Technometrics*, 37, 176–184.

Schulze, H., Muench, S. B., Villatte, F., Schmid, R. D., and Bachmann, T. T. (2005), "Insecticide Detection through Protein Engineering of *Nippostrongylus brasiliensis* Acetycholinesterase B," *Analytical Chemistry*, 77, 5823–5830.

Tukey, J. W. (1958), "Bias and Confidence in Not-Quite Large Samples," *Annals of Mathematical Statistics*, 29, 614.