

Review

# Calibration, uncertainty, and recovery in the chromatographic sciences

L.E. Vanatta<sup>a,\*</sup>, D.E. Coleman<sup>b</sup>

<sup>a</sup> Air Liquide—Balazs™ Analytical Services, Box 650311, MS 301, Dallas, TX 75265, USA

<sup>b</sup> Alcoa Technical Center, MST-C, 100 Technical Drive, Alcoa Center, PA 15069, USA

Available online 17 February 2007

## Abstract

This paper reviews calibration-, uncertainty-, and recovery-related documents from 10 consensus-based organizations. The main points from each treatise are summarized. Also included is a critique of the various approaches, as well as recommendations for a statistically sound protocol that is more compatible with chromatographic data.

© 2007 Elsevier B.V. All rights reserved.

**Keywords:** Calibration; Data analysis; Measurement uncertainty; Prediction interval; Recovery; Regression

## Contents

1. Introduction .....	48
2. Publications from “uncertainty-budget” organizations .....	49
2.1. EURACHEM documents (references [1–3]) .....	49
2.1.1. Reference [1] .....	50
2.1.2. Reference [2] .....	50
2.1.3. Reference [3] .....	51
2.2. Additional documents (references [5–10]) .....	51
3. Publication from “regression-diagnostics” organization (reference [11]) .....	52
4. Publication from “neutral” organization (reference [12]) .....	52
5. Critique of publications reviewed .....	52
6. Recommended approach .....	53
Acknowledgments .....	53
Appendix A. Background on organizations .....	53
A.1. A2LA .....	53
A.2. AIHA .....	54
A.3. APLAC .....	54
A.4. CCIL .....	54
A.5. CITAC .....	54
A.6. EA .....	54
A.7. EURACHEM .....	54
A.8. EUROLAB .....	55
A.9. ICH .....	55
A.10. IUPAC .....	55
Appendix B. Definitions .....	55
B.1. Accuracy .....	55
B.2. Bias .....	55
B.3. Calibration Curve .....	56

\* Corresponding author. Tel.: +1 972 9957541; fax: +1 972 9953204.  
E-mail address: [lynn.vanatta@airliquide.com](mailto:lynn.vanatta@airliquide.com) (L.E. Vanatta).

B.4.	Combined Standard Uncertainty	56
B.5.	Coverage factor	56
B.6.	Error (of Measurement)	56
B.7.	Expanded Uncertainty	56
B.8.	Linearity	56
B.9.	Measurand	56
B.10.	Measurement	56
B.11.	Measurement Procedure	56
B.12.	Method Validation	56
B.13.	Precision	56
B.14.	Random Error	57
B.15.	Repeatability (of a measuring instrument)	57
B.16.	Repeatability (of results of measurements)	57
B.17.	Repeatability Limit “r”	57
B.18.	Repeatability Standard Deviation	57
B.19.	Reproducibility	57
B.20.	Reproducibility Limit “R”	57
B.21.	Reproducibility Standard Deviation	57
B.22.	Standard Deviation	57
B.23.	Standard Uncertainty	58
B.24.	Systematic Error	58
B.25.	True Value	58
B.26.	Trueness	58
B.27.	Uncertainty (of Measurement) i.e. Measurement Uncertainty	58
B.28.	Validation	58
Appendix C.	Recommended approach for chromatographic-data analysis. Article is from <i>American Laboratory</i> and is reprinted with permission from the publisher, International Scientific Communications: Statistics in Analytical Chemistry Part 25 –Calibration Summary	58
C.1.	Step 1: Design the calibration study	58
C.2.	Step 2: Perform the calibration study in the laboratory	59
C.3.	Step 3: Diagnose the calibration data statistically	59
C.4.	Calibration/Regression	59
References		60

## 1. Introduction

Today, chromatographs are a staple in most analytical-chemistry laboratories. As such, these instruments generate an enormous amount of data, which are used to make a variety of business and regulatory decisions. Thus, these analytical results often lead to far-reaching consequences. It follows that there needs to be a mechanism for assessing the quality of these numbers. At this juncture, statistical analysis becomes a useful tool.

However, raw chromatographic data present a fundamental problem; i.e., the values that the detector generates are *not* in immediately useful units. Instead, results are reported as, e.g., absorbance units or peaks areas (or heights). Hence, a process is needed to transform these values into, typically, concentrations. This process is the statistical technique of regression. (In the regression process, a model is fitted to the data via a fitting technique; the model is a mathematical equation that relates the response,  $y$ , to the true value,  $x$ ; a fitting technique determines how much influence each measurement has on the curve.) With these “regression-based” instruments, a series of known analyte concentrations (i.e., standards) are analyzed and the responses recorded. Regression is used to establish a relationship between the two sets of variables data; in this context, the statistical tech-

nique is known as “calibration” and its result is a “calibration curve.”

To be certain that any calibration expression is adequate, important statistics-related questions must be asked. It should be noted that although all of the following questions are important, perhaps the most important (and often neglected) query is the last one regarding uncertainty.

- (1) Was the design of the calibration study appropriate?
- (2) Were the resulting data analyzed in a statistically sound manner?
  - (2a) Is the chosen model adequate?
  - (2b) Is the fitting technique appropriate?
  - (2c) How much bias does the curve reflect?
  - (2d) What is the uncertainty in any sample result that is estimated from the curve?

Typically, calibration standards are prepared in a pure solvent that is compatible with the particular chromatograph. Actual samples rarely are in such “nice” form and often have to go through a preparation process before they can be analyzed chromatographically. Thus, there is a very real possibility that the concentration estimate from the calibration curve will *not* reflect 100% recovery. Therefore, once an appropriate calibration curve

has been established, the behavior of the sample matrix(s) must be investigated. Statistical analysis again is called upon to help assess recovery; questions similar to the above must be answered.

A great deal has been written about calibration and recovery matters. Furthermore, many consensus-based organizations have been formed to set standards for generating, evaluating, and reporting analytical data. Many governmental and regulatory agencies are accepting and relying on these standards.

Since this review: (1) cannot cover all of the literature and (2) is not intended to be used for political or commercial purposes, the focus is on the publications of these consensus-based groups. No documents from governmental or for-profit organizations are included. A further restriction is that all treatises must be freely available via the Internet. The reason for this narrowing is two-fold. First and most important, most of the widely accepted and utilized documents fall into this category; ones that do not are typically very similar in scope and content. Second, implementing this restriction means that the interested reader can easily obtain the full text of any citation (Web sites are given in Reference section). A full list of the organizations whose publications are included in this review is given in [Appendix A](#). Also given is a brief synopsis of the membership and purpose of each group. As can be seen from these summaries, the associations with the most internationally based membership are: (1) EURACHEM, (2) ICH, and (3) IUPAC. The documents from the first two groups are some of the most widely referenced and used. (Readers who are interested in ISO documents that are related to this review are referred to the Quality section of the ISO catalog, especially section 03.120.30, Application of statistical methods. This catalog is available on the ISO Web site, <http://www.iso.org>.)

In the reviewed literature on analytical data, by far the most discussed topic is the subject of uncertainty. On this topic, the writings generally take one of two approaches: (1) development of an “uncertainty budget” or (2) evaluation via regression diagnostics. The first strategy is the overwhelming favorite and is central to the documents from all of the associations in [Appendix A](#) except IUPAC and ICH. The first exception discusses the second protocol; the second exception remains neutral, in that the document does not address uncertainty.

The remainder of this paper is divided according to how the “parent” organization approaches uncertainty. Following the discussion on the contents of the publications, the pros and cons of the writings are given. Last, a recommended strategy is presented. Statistical topics that are addressed are: (1) the calibration process, (2) measurement uncertainty, and (3) recovery. The following subjects are outside the scope of this review: (1) detection limit, (2) quantitation limit, (3) reporting limit, (4) identity confirmation/selectivity/specificity, (5) ruggedness/robustness, (6) traceability, and (7) outliers. Three appendices are included. [Appendix A](#) was mentioned in the preceding paragraph. [Appendix B](#) contains definitions of terms specific to this statistical discussion; all entries are from Reference [1]. [Appendix C](#) is the full text of a recently published article that summarizes the approach recommended by the authors of this review.

## 2. Publications from “uncertainty-budget” organizations

Because of their broadly based membership and widely accepted publications, the works of EURACHEM are discussed in detail. Subsequently, differences found in the works of other “budget” groups are given.

### 2.1. EURACHEM documents (references [1–3])

EURACHEM has published three main documents [1–3] related to analytical methods.

Ref. [1], “The Fitness for Purpose of Analytical Methods” (1998), was the result of a Working Group comprised of representatives from Belgium, Germany, UK, Hungary, Sweden, USA, The Netherlands, Switzerland, Denmark, Czech Republic, Finland, Ireland, and Austria. This guide concentrates on method validation:

- “1. The process of establishing the performance characteristics and limitations of a method and the identification of the influences which may change these characteristics and to what extent. Which analytes can it determine in which matrices in the presence of which interferences? Within these conditions what levels of precision and accuracy can be achieved?”
- “2. The process of verifying that a method is fit for purpose, i.e. for use for solving a particular analytical problem.”

Ref. [2] is EURACHEM/CITAC Guide, “Quantifying Uncertainty in Analytical Measurement,” Second edition (2000). This document was written by a Working Group comprised of representatives from EURACHEM, CITAC, AOAC, IAEA, and EA. Countries represented were UK, Switzerland, Belgium, Germany, The Netherlands, Sweden, Austria, China, USA, Australia, and Japan. According to the Foreword, this revised edition of the EURACHEM/CITAC Guide “. . .stresses that the procedures introduced by a laboratory to estimate its measurement uncertainty should be integrated with existing quality assurance measures, since these measures frequently provide much of the information required to evaluate the measurement uncertainty.”

Furthermore, the Foreword states that the document is application oriented, showing how the 1993 ISO document (“Guide to the Expression of Uncertainty in Measurement [4]”) “. . .may be applied in chemical measurement.” (The ISO Guide was written by members from ISO, BIPM, IEC, IFCC, IUPAC, IUPAP, and OIML. Per the EURACHEM/CITAC Guide, the ISO treatise “. . .formally established general rules for evaluating and expressing uncertainty in measurement across a broad spectrum of measurements.”) In Section 1.1 (of the Scope and Field of Application section of ref. [2]), the ISO Guide is said to be “. . .applicable at all levels of accuracy and in all fields. . .”

Ref. [3] (“Harmonised Guidelines for the Use of Recovery Information in Analytical Measurement, 1995”) was jointly prepared by IUPAC, ISO, AOAC, and EURACHEM. Members of the Working Party were from Switzerland, Austria, Denmark, Norway, Germany, Spain, USA, The Netherlands, Belgium,

Australia, UK, Hungary, New Zealand, Sweden, Italy, and Portugal. As is stated in the report's Foreword, the guide is general in scope and gives overall recommendations concerning recovery.

### 2.1.1. Reference [1]

The fitness-of-purpose treatise is discussed first, since a working method must be developed and validated before meaningful uncertainty measurements can be made. Within the context of calibration, the earliest topic is the working range of concentrations. Section 6.27 states that "Within the working range there may exist a linear response range. Within the linear range signal response will have a linear relationship to analyte concentration or property value. The extent of this range may be established during the evaluation of the working range. Note that regression calculations on their own are insufficient to establish linearity. To do this a visual inspection of the line and residuals may be sufficient; objective tests, such as 'goodness-of-fit' tests, are better still [15–17]. In general linearity checks require points at least 10 different concentrations/property values." An accompanying table ("Working and Linear Range—Quick Reference") emphasizes the use of visual inspection of the regression plot itself and of the residual pattern to determine linearity. The table also states, "If variance of replicates is proportional to concentration then use a weighted regression calculation rather than a non-weighted regression. In certain circumstances it may be better to try to fit a non-linear curve to the data. Functions higher than quadratic are generally not advised"

The document splits accuracy into two components: trueness and precision. Trueness typically is assessed via the use of certified reference material or via comparison with an established method. In the first case, the mean and standard deviation from replicate measurements are calculated and used for trueness evaluation. Section 6.34 recommends that, "To check against an alternative method, compare results from the two methods for the same sample or samples." Finally, Section 6.36 states, "For most purposes, however, acceptability of bias should be decided on the basis of overall bias measured against appropriate materials or reference methods, taking into account the precision of the method, any uncertainties in reference material values, and the accuracy required by the end use. Statistical significance tests are recommended."

Precision is discussed in Section 6.37. "'Precision' is normally determined for specific circumstances which in practice can be very varied. The two most common precision measures are 'repeatability' and 'reproducibility'. They represent the two extreme measures of precision which can be obtained. Repeatability (the smallest expected precision) will give an idea of the sort of variability to be expected when a method is performed by a single analyst on one piece of equipment over a short timescale, i.e. the sort of variability to be expected between results when a sample is analysed in duplicate. If a sample is to be analysed by a number of laboratories for comparative purposes then a more meaningful precision measure to use is reproducibility (this is the largest measure of precision normally encountered, although it does formally exclude variation with respect to time). ... Precision is usually stated in terms of standard deviation or relative standard deviation. Both repeatability and reproducibility are

generally dependent on analyte concentration, and so should be determined at a number of concentrations and if relevant, the relationship between precision and analyte concentration should be established." Repeatability and reproducibility limits are also defined (see Appendix B of this paper).

The topic of measurement uncertainty is mentioned in this guide, but only briefly. Readers are referred to documents such as ref. [2], the contents of which are discussed next in this paper.

Recovery is addressed in Sections 6.46 and 6.47, and in an accompanying table ("Recoveries-Quick Reference"). The use of spiked samples (at various concentrations) or certified reference materials are recommended. For spikes, recovery (%) is defined as:

$$\frac{(C1 - C2)}{C3} \times 100,$$

where C1 = concentration determined in fortified sample  
C2 = concentration determined in unfortified sample  
C3 = concentration of fortification"

### 2.1.2. Reference [2]

Once a method has been validated, measurement uncertainty can be addressed via this EURACHEM/CITAC guide. Measurement uncertainty is defined in Section 2.1 to be:

"A parameter associated with the result of a measurement, that characterises the dispersion of the values that could reasonably be attributed to the measurand." Uncertainty might be expressed in terms of a standard deviation or a confidence interval. Furthermore, this uncertainty typically contains more than one component. Type A components "... may be evaluated from the statistical distribution of the results of series of measurements. ..." Type B components "... are evaluated from assumed probability distributions based on experience or other information."

In Section 2.3, uncertainties are further categorized as follows:

**Standard uncertainty:** An uncertainty component that is expressed as a standard deviation.

**Combined standard uncertainty,  $u_c(y)$ :** The total uncertainty, which is "... an estimated standard deviation equal to the positive square root of the total variance obtained by combining all the uncertainty components, however evaluated, using the law of propagation of uncertainty."

**Expanded uncertainty,  $U$ :** "... provides an interval within which the value of the measurand is believed to lie with a higher level of confidence.  $U$  is obtained by multiplying  $u_c(y)$  ... by a coverage factor,  $k$ . The choice of the factor  $k$  is based on the level of confidence desired. For an approximate level of confidence of 95%,  $k$  is 2." Typically,  $U$  is the value that should be reported.

The Guide recommends that uncertainty (which represents a range of values) be distinguished from error, which is discussed in Section 2.4. Error has both a random and a systematic component, and is defined as "... the difference between an individual result and the true value of the measurand." Correction of a result for error, which is a single number, is encouraged.

In Section 4.1, the actual steps for estimating measurement uncertainty are described:

- (1) Identify the measurand clearly.
- (2) List the sources of uncertainty.
- (3) Estimate each source.
- (4) Calculate the combined uncertainty.

Sections 5 and 6 provide detailed guidance for steps 1 and 2. Steps 3 and 4, which involve statistical techniques, comprise Sections 7 and 8 (respectively) and are discussed here. In many cases, data for estimating uncertainties are available from existing data (such as QA/QC results), and the use of these reports is encouraged. General procedural advice is given in Section 7; directions depend on the source of the data.

Instructions and formulas for estimating and combining uncertainties are the subject of Section 8. First, all standard uncertainties must be converted (if necessary) to standard deviations. Second, the combined standard uncertainty is estimated. The general expression for this estimate is:

$$u_c(y(x_1, x_2, \dots)) = \sqrt{\sum_{i=1,n} c_i^2 u(x_i)^2} = \sqrt{\sum_{i=1,n} u(y, x_i)^2},$$

where  $y(x_1, x_2, \dots)$  = a function of  $x_1, x_2$ , etc.;  $c_i$  = a sensitivity coefficient;  $=$  the partial differential of  $y$  with respect to  $x_i$ ;  $u(y, x_i)$  = the uncertainty in  $y$  arising from the uncertainty in  $x_i$ ;  $=$  “the square of the associated uncertainty expressed as a standard deviation multiplied by the square of the relevant sensitivity coefficient.”

The above formula is more complicated if the variables are not independent. In some cases, the general expressions can be simplified. Both situations are detailed in the text. Also, Section 8.2.4 states that “. . . when an uncertainty contribution is associated with the whole procedure, it is usually expressed as an effect on the final result. In such cases, or when the uncertainty on a parameter is expressed directly in terms of its effect on  $y$ , the sensitivity coefficient . . . is equal to 1.0.”

The third and final step in the estimation process is to determine the expanded uncertainty, which is obtained from the combined standard uncertainty by multiplying by the appropriate coverage factor. Recommendations for choosing this factor are given, but in general, a value of 2 is selected.

Six examples are worked in detail in Appendix A. Example A4 (Determination of organophosphorus pesticides in bread) involves the use of a gas chromatograph and is the example of interest for this review article. The example utilizes data from the method-validation process. A schematic of the overall analysis is given in Fig. A4.4; sources of uncertainty are detailed in cause-and-effect diagrams (Figs. A4.2, A4.5, and A4.6; the last includes branches for the validation study). Data from precision and spiking studies are listed in Tables A4.2 and A4.3, respectively. Data from the validation study showed that the instrumental response was linear within the working concentration range; single-point calibration of the instrument was conducted in order to generate a reference-peak intensity. Details are given for calculating the three sequential uncer-

tainties described above; uncertainty results are summarized in Tables A4.4 and A4.5.

### 2.1.3. Reference [3]

The topic of recovery (of the analytes from the matrix) is the subject of this document. Section 3 presents an overview of various approaches for addressing recovery. The possibilities discussed are use of: (1) matrix reference material, (2) surrogates, (3) isotope dilution, (4) spiking, and (5) internal standards. Section 6 and the Appendix deal with calculating the recovery and its associated uncertainty. The recovery,  $R$ , is given in the Appendix by the expression:

$$R = \frac{c_{\text{obs}}}{c_{\text{ref}}},$$

where  $c_{\text{obs}}$  = the observed concentration obtained from the analytical method, and  $c_{\text{ref}}$  = the true concentration in the reference material.

The uncertainty associated with  $R$  is  $u_R$ . If the reported data have been corrected for recovery, then  $u_R$  is included in the uncertainty budget via the formula:

$$\left(\frac{u_{\text{corr}}}{c_{\text{corr}}}\right) = \sqrt{\left(\frac{u_c}{c}\right)^2 + \left(\frac{u_R}{R}\right)^2},$$

where:  $u_{\text{corr}}$  = the combined uncertainty;  $c_{\text{corr}}$  = the corrected result,  $c/R$ ;  $u_c$  = uncertainty associated with the raw result; and  $c$  = the raw result.

Finally,  $u_{\text{corr}}$  is multiplied by the coverage factor,  $k$ , to determine the expanded uncertainty,  $U$ .

## 2.2. Additional documents (references [5–10])

In the following discussion of these six documents, content that duplicates what has been presented in Section 2.1 above is not repeated here; only additional details are included.

Per the Introduction, the A2LA Guide [5] “. . . provides guidance on estimation of uncertainty based on reproducibility estimates and control charting.” In the discussion of Type A uncertainty, details of various probability distributions (rectangular, triangular, Normal, U, Poisson) are included in Section 3.3.2. Section 3.10 calls for a thorough evaluation of uncertainty, not just the construction of an uncertainty-budget table, which only summarizes the results.

The AIHA Guidelines document [6] centers on the Type A and Type B approaches to uncertainty. The text encourages the use of quality-control data if Type B components are not significant.

APLAC’s document [7] includes a Section (2.6) on uncertainty that comes from sampling. In Section 6 on Chemical Testing, the use of reproducibility data is encouraged. Also, in Section 6.2, it states, “In the field of chemical analysis it is considered acceptable to group sources of uncertainty together; . . .”

CCIL’s Protocol [8] relies solely on the use of QC data and was developed for use by environmental testing laboratories. The document states, “Uncertainty values obtained from this procedure must be regarded as **estimates**. . . **It is our intent**

**with this procedure to arrive at an estimate of a 95% confidence level uncertainty value that can be assumed to apply to 95% (or more) of the samples that a laboratory receives for a given test.**" (Emboldened text is as it appears in the CCIL text.) Key equations and a step-by-step description of the protocol are given.

The EA Guidelines [9] suggest the use of quality-control data, proficiency-testing data, and prior-study data (Sections 6.4, 6.5, and 6.7, respectively).

EUROLAB's report [10] also allows the use of "grouped data" (e.g., QC results) in estimating uncertainty. Example 1 in Section 4 outlines three approaches to estimating the uncertainty associated with an ion-chromatography analysis. The first estimation uses proficiency-test results, the second utilizes control-chart data, and the third relies on validation data in the standard method that was used.

### 3. Publication from "regression-diagnostics" organization (reference [11])

The referenced article [11] was prepared by the members of the IUPAC's Analytical Chemistry Division, Commission on General Aspects of Analytical Chemistry. The Commission had members from Germany, USA, Sweden, The Netherlands, UK, Poland, Belgium, Switzerland, Japan, South Africa, Russia, Brazil, Czechoslovakia, Czech Republic, Hungary, India, Ireland, New Zealand, Portugal, South Korea, and Turkey.

The Synopsis states that, "This IUPAC nomenclature document has been prepared to establish a uniform and meaningful approach to terminology, notation, and formulation for calibration in analytical chemistry." The text is replete with statistical formulas and covers the topics of: (1) calibration functions and models, (2) least-squares calibration, (3) evaluation of calibration errors, (4) linearity, (5) trends in response variance, and (6) standard addition.

In discussing calibration errors, the document introduces the concept of prediction intervals in Section 3.2.9; formulas are included. Section 3.3 addresses the situation where the variance of the response trends with concentration; in such situations, the fitting technique of weighted least squares is needed and is shown. The article does not address how to compute prediction intervals in the weighted-least-squares situation. Instead, the document avoids this complexity (and an expansion of scope) by advising the reader that the formulas are similar. Section 7 outlines the use of standard addition, but ends with the following caveat, "Although standard addition calibration is an unsafe method if linearity in the range  $x < x_0$  is not experimentally verified but only supposed, there is scarcely an alternative when matrix effects are seriously suspected."

### 4. Publication from "neutral" organization (reference [12])

An ICH Expert Working Group wrote this Guideline [12] and in doing so, combined previous documents Q2A ("Text on Validation of Analytical Procedures") and Q2B ("Guideline on Validation of Analytical Procedures: Methodology").

No specific statistical procedures or formulas are given. Indeed, the Introduction states, "This document presents a discussion of the characteristics for consideration during the validation of the analytical procedures included as part of registration applications submitted within the EC, Japan and USA. . . . Furthermore, this text presentation serves as a collection of terms, and their definitions, and is not intended to provide direction on how to accomplish validation." To that end, a discussion of the concepts of: (1) accuracy, (2) precision, (3) linearity, and (4) range are included. Also given is a table recommending when each of these concepts should be evaluated, depending on the purpose of the analytical method. No mention of uncertainty is made.

### 5. Critique of publications reviewed

With the exception of reference [11], the above publications have the admirable goal of attempting to construct a general approach to measurement, especially measurement uncertainty. Unfortunately, there are several reasons why such documents typically are not adequate for use with regression-based instruments such as chromatographs. First, while calibration of such instruments is assumed, no definitive guidelines are given for carrying out the procedure or for incorporating calibration error into the uncertainty formulas. The magnitude of calibration error varies with the method, matrix, and analyte, but should never be assumed to be negligible.

Second, the uncertainty-budget approach requires the division of the overall uncertainty into components, evaluation of these parts, and then the combining of the results. Such a protocol often: (1) is tedious and involves complicated, detailed calculations, (2) can be unduly conservative (i.e., the total error variation can be overestimated), and (3) does not incorporate any relationship between error standard deviation and concentration level. Additionally, estimation error can be introduced at each step in the estimating process, since sample standard deviations are intrinsically noisy.

Third, no definitive recommendations are made for addressing imperfect recovery, a problem that affects many chromatographic procedures, as can be revealed via a plot of recovered concentrations versus true concentrations. For such a graph, the linear slope and intercept typically will have statistically significant values that are different from 1 and 0, respectively.

Fourth, most discussions of calibration stress linearity and range. This emphasis often leads to two unfortunate consequences. First, there is an assumption that any method's response data will be linearly related to the true concentrations, and that something is "wrong" if such an outcome does not materialize. However, not all data behave linearly; in such situations, the preferred outcome is to find an appropriate non-linear model to explain the data. Second, there is a tendency to see how wide a concentration range can be assumed to be linearly related to the response data. Instead, the concentration range should be limited to the values that need to be covered by the method. In general, this limited approach leads to the selection of the most appropriate calibration (or recovery) curve, along with the

most statistically sound uncertainty estimate. With all methods, extrapolation of the relevant curve is not wise.

Reference [11] is a step in the right direction, in that this document details the fundamentals behind regression and uncertainty intervals. Lacking, though, is a guide that the chromatographer can follow in calibrating the instrument and conducting recovery studies.

## 6. Recommended approach

The recommendation of the authors is that a regression-based approach be taken to the generation and analysis of chromatographic data. The IUPAC document [11] provides the basics for such a protocol, and does so in a conceptually logical and rigorous manner, with a helpful flowchart. However, no real-data examples are provided in the document. Also absent are practical details, which would allow the analyst to carry out this strategy in his or her own laboratory. To address these deficiencies, a step-by-step plan is proposed. The overall approach has been summarized recently [13]; the text is repeated in Appendix C, with permission from the publisher, International Scientific Communications. The details (along with real-world examples) have been given in a series of articles [14–37] appearing in *American Laboratory*. These installments have been contributed by the authors without compensation, and are based on a combination of statistical fundamentals and application experiences; all papers are available gratis via the Internet. Table 1 summarizes

the subject matter of each part. Readers who are interested in more details are referred to the first 24 parts of the series.

## Acknowledgments

The authors wish to express deep appreciation to the following people for their help in locating material relevant to this review: John P. Auses (Alcoa Technical Center), Nancy E. Grams (Advanced Earth Technologies), David S. Jackson (United States Food and Drug Administration), Steven J. Jones (Air Liquide – Balazs), Bernard A. Olsen (Eli Lilly and Company), Dale Rushneck (Interface Inc.), and Don Saunders (Eli Lilly and Company).

## Appendix A. Background on organizations

### A.1. A2LA

American Association for Laboratory Accreditation (established in 1978) (from the Fact Sheet at <http://www.a2la.org/general/structure.cfm>)

“Membership in A2LA is open to any individual, institution or corporation interested in supporting its mission. Members pay dues, elect the Board of Directors, attend meetings sponsored by the Association, . . .”

“As of January 1, 2006, A2LA has 1803 accredited laboratories in 48 States in the US and several foreign countries.”

Table 1  
Titles and citation information for Parts 1 through 24 of the *American Laboratory* series, “Statistics in Analytical Chemistry”

Part #	Title	Citation information	Comments on content
1	A New American Laboratory Column	34 (September 2002) 44	Introduction of the authors' backgrounds and of future topics
2	Measurement and the Real-Number Line	34 (November 2002) 38	
3	Calibration: Introduction and Ordinary Least Squares	35 (January 2003) 18	
4	Calibration: Uncertainty Intervals	35 (March 2003) 60	
5	Calibration: Calibration Design	35 (June 2003) 30	Fundamental concepts
6	Calibration Design	35 (July 2003) 19	Example designs
7	A Review	35 (September 2003) 31	
8	Calibration Diagnostics	35 (November 2003) 40	Modeling of standard deviation's behavior
9	Calibrations Diagnostics (Continued)	36 (February 2004) 64	Residual patterns
10	Calibration Diagnostics (Concluded)	36 (March 2004) 46	Lack-of-fit test and prediction intervals
11	Calibration Example	36 (May 2004) 40	Straight-line model with ordinary least squares is adequate (SL/OLS)
12	Calibration Example 2	36 (July 2004) 38	SL/OLS is adequate
13	Calibration Example 3	36 (September 2004) 28	SL/weighted least squares (WLS) is adequate
14	Calibration Example 4	36 (December 2004) 32	Quadratic model with WLS is adequate
15	Calibration Example 5	37 (February 2005) 48	Cubic model with WLS is adequate
16	Calibration Example 5 (Concluded)	37 (May 2005) 37	Use of an inadequate model
17	Calibration Example 6	37 (August 2005) 32	No model is adequate
18	Calibration Example 6 (Concluded)	37 (September 2005) 24	Use of piece-wise calibration
19	Internal Standards	37 (December 2005) 23	
20	Recovery Curves	38 (February 2006) 54	
21	Recovery Curves and Internal Standards (Concluded)	38 (April 2006) 30	
22	Lack-of-Fit Details	38 (June/July 2006) 34	
23	Degrees of Freedom in the Lack-of-Fit Test	38 (October 2006) 38	
24	Glossary	38 (November/December 2006) 25	Compilation of all statistical terms and symbols used in previous articles

Explanatory comments on an article's content are given where appropriate.

“The American Association for Laboratory Accreditation (A2LA) is a nonprofit, non-governmental, public service, membership society. The mission of A2LA is to provide comprehensive services in laboratory accreditation and laboratory-related training. Services are available to any type of organization, be it private or government. Laboratory accreditation is based on internationally accepted criteria for competence (ISO/IEC 17025:2005). A2LA also offers programs for accreditation of inspection bodies, proficiency testing providers, reference material producers and product certification bodies.”

#### A.2. AIHA

American Industrial Hygiene Association (founded in 1939) (from <http://www.aiha.org/Content/AboutAIHA/aboutaiha-splash.htm>)

Organizational Members are primarily from the USA, but also include ones from Canada, UK, Brazil, Taiwan, Japan, Kuwait, China, Philippines, and Saudi Arabia.

“The American Industrial Hygiene Association is one of the largest international associations serving the needs of occupational and environmental health professionals practicing industrial hygiene in industry, government, labor, academic institutions, and independent organizations.”

“...AIHA is a nonprofit organization with more than 75 local sections.”

#### A.3. APLAC

Asia Pacific Laboratory Accreditation Cooperation (initiated in 1992) (from <http://www.aplac.org>).

Full Member countries are Australia, Brunei Darussalam, Canada, China, Hong Kong (China), India, Indonesia, Japan, South Korea, Malaysia, Mexico, Mongolia, New Zealand, Pakistan, Papua New Guinea, Philippines, Singapore, Taiwan, Thailand, USA, and Vietnam.

“Asia Pacific Laboratory Accreditation Cooperation (APLAC) groups accreditation bodies in the Asia Pacific region responsible for accrediting calibration, testing and inspection facilities. APLAC’s principal objectives are to foster the development of competent laboratories and inspection bodies in member economies, to harmonize accreditation practices in the region and with other regions, and to facilitate mutual recognition of accredited test, measurement and inspection results. . . .”

#### A.4. CCIL

Canadian Council of Independent Laboratories (founded in 1993) (from <http://www.ccil.com/about.html#1>)

Members are firms within Canada. “CCIL members are taxpaying business entities, unaffiliated with any academic or governmental institution or with any outside industrial company or trade group which in any manner might affect the firm’s ability to conduct investigations, render reports or give professional counsel objectively and without bias.”

“CCIL and its member firms share a commitment to protect the public through education, adhering to a code of ethics, encouraging good performance and reliability among members, and establishing fair and just fee guidelines. The association participates with standards-writing bodies and conducts surveys for the benefits of its member firms. The CCIL promotes certification by standards bodies or through special programs to meet specific needs.”

#### A.5. CITAC

Co-Operation on International Traceability in Analytical Chemistry (initiated in 1993) (from <http://www.citac.cc>)

Working Group Members are from Argentina, Australia, Austria, Belgium, Brazil, Chile, China, Finland, France, Germany, Greece, Hong Kong (China), India, Ireland, Israel, Japan, South Korea, Mexico, New Zealand, Russia, South Africa, The Netherlands, UK, USA.

“CITAC . . . arose out of an international workshop held in association with the Pittsburgh Conference in Atlanta in March 1993. The aim of this workshop was to discuss how analytical activities could be developed to meet the needs of the 21st century, and it identified a wide variety of issues to be addressed to ensure that analytical measurements made in different countries or at different times are comparable. These range from the development of traceable reference materials and methods to the harmonisation of analytical quality practices.

“The CITAC Initiative aims to foster collaboration between existing organization (*sic*) to improve the international comparability of chemical measurement.”

#### A.6. EA

European co-operation for Accreditation (formed in 1997) (from <http://www.european-accreditation.org>)

Full Members are Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, The Netherlands, Norway, Poland, Portugal, Croatia, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, and UK.

EA is concerned with the accreditation of laboratories, as well as inspection and certification organizations. “The EA missions are

To ensure transparency of the operations (including assessments) and results of its members

To ensure common interpretation of the standards they use  
To manage a peer evaluation system, consistent with the international practice. . .

To support and promote mutual recognition and acceptance of accredited conformity assessment services and results”

#### A.7. EURACHEM

EURACHEM (established in 1989) (from <http://www.eurachem.ul.pt>)



Member countries are Albania, Austria, Belgium, Croatia, Cyprus, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Lithuania, Luxembourg, Malta, The Netherlands, Norway, Poland, Portugal, Romania, Russia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, Ukraine, and UK.

“Eurachem is a network of organisations in Europe, having the objective of establishing a system for the international traceability of chemical measurements and the promotion of good quality practices. It provides a forum for the discussion of common problems and for developing an informed and considered approach to both technical and policy issues.”

“Membership of Eurachem is open to countries within the European Union and the European Free Trade Association, the European Commission and European countries recognized by the EU and EFTA as accession states.”

#### A.8. EUROLAB

EUROLAB (created in 1990) (from “Eurolab information,” “Our Profile” at <http://www.eurolab.org>)

Active Members are Austria, Belgium, Cyprus, Czech Republic, Denmark, Finland, France, Germany, Greece, Iceland, Italy, Latvia, The Netherlands, Norway, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, Switzerland, and UK.

“Objectives:

**Representation** by formulating and voicing the opinion of European laboratories regarding political and technical issues having a direct impact on their activity, both on the European scene and worldwide.

**Coordination** by interfacing with all European organizations having activities of interest to the laboratory community, and striving to avoid duplication of efforts and activities.

**Action** by providing adequate means for exchange of information and experience, such as the publication of our Position Papers, Technical Reports, Newsletter, Seminars, and Working Groups, etc.

**Promoting** cost-effective testing, calibration and measurement services, for which the accuracy and quality assurance requirements should be adjusted to actual needs.”

#### A.9. ICH

International Conference of Harmonisation on Technical Requirements for Registration of Pharmaceuticals for Human Use (established in 1990) (<http://www.ich.org/cache/compo/276-254-1.html>)

Six Parties are directly involved in ICH activities. These Parties are EU (European Commission—European Union), EFPIA (European Federation of Pharmaceutical Industries and Associations), MHLW (Ministry of Health, Labour and Welfare, Japan), JPMA (Japan Pharmaceutical Manufacturers Association), FDA (United States Food and Drug Administration), and PhRMA (Pharmaceutical Research and Manufacturers of America).

“... (ICH) is a unique project that brings together the regulatory authorities of Europe, Japan and the United States and

experts from the pharmaceutical industry in the three regions to discuss scientific and technical aspects of product registration.

“The purpose is to make recommendations on ways to achieve greater harmonisation in the interpretation and application of technical guidelines and requirements for product registration in order to reduce or obviate the need to duplicate the testing carried out during the research and development of new medicines.

“The objective of such harmonisation is a more economical use of human, animal and material resources, and the elimination of unnecessary delay in the global development and availability of new medicines whilst maintaining safeguards on quality, safety and efficacy, and regulatory obligations to protect public health.”

#### A.10. IUPAC

International Union of Pure and Applied Chemistry (formed in 1919) (from <http://www.iupac.org/general/about.html>).

“IUPAC is an association of bodies, National Adhering Organizations, which represent the chemists of different member countries. There are 45 National Adhering Organizations, and 20 other countries are also linked to IUPAC in the status of Associate National Adhering Organizations.” The countries in the National Adhering Organizations are Argentina, Australia, Austria, Bangladesh, Belarus, Belgium, Brazil, Bulgaria, Canada, Chile, China, Croatia, Czech Republic, Denmark, Egypt, Finland, France, Germany, Greece, Hungary, India, Ireland, Israel, Italy, Jamaica, Japan, Jordan, Kuwait, The Netherlands, New Zealand, Norway, Pakistan, Poland, Portugal, Puerto Rico, Russia, Serbia and Montenegro, Slovakia, Slovenia, South Africa, South Korea, Spain, Sweden, Switzerland, Turkey, Ukraine, UK, and USA.

“The International Union of Pure and Applied Chemistry (IUPAC) serves to advance the worldwide aspects of the chemical sciences and to contribute to the application of chemistry in the service of Mankind. As a scientific, international, non-governmental and objective body, IUPAC can address many global issues involving the chemical sciences.”

### Appendix B. Definitions (all are from ref. [1])

#### B.1. Accuracy

The closeness of agreement between a test result and the accepted reference value. Note: *The term accuracy, when applied to a set of test results, involves a combination of random components and a common systematic error or bias component.* [No reference given]

A quantity referring to the differences between the mean of a set of results or an individual result and the value which is accepted as true or correct value for the quantity measured. [IUPAC Compendium of Chemical Technology, 1985]

#### B.2. Bias

The difference between the expectation of the test results and an accepted reference value. Note: *Bias is the total*

systematic error as contrasted to random error. There may be one or more systematic error components contributing to the bias. A larger systematic difference from the accepted reference value is reflected by a larger bias value. [ISO 3534-1]

Characterises the systematic error in a given analytical procedure and is the (positive or negative) deviation of the mean analytical result from the (known or assumed) true value. [IUPAC Compendium of Chemical Technology, 1985]

The difference between the limiting mean ( $\mu$ ) and the true value ( $\tau$ ); i.e.,  $\Delta = \mu - \tau$ . [IUPAC 'Orange' Book]

### B.3. Calibration Curve

Graphical representation of measuring signal as a function of quantity of analyte [AOAC–PVMC]

### B.4. Combined Standard Uncertainty

$u_c(y)$ —Standard uncertainty of the result of a measurement when the result is obtained from the values of a number of other quantities, equal to the positive square root of the sum of terms, the terms being the variances or co-variances of these other quantities weighted according to how the measurement result varies with these quantities. [ISO GUM]

### B.5. Coverage factor

$k$ —numerical factor used as a multiplier of the combined standard uncertainty in order to obtain an expanded uncertainty. Note A coverage factor is typically in the range 2 to 3. [ISO GUM]

### B.6. Error (of Measurement)

The result of a measurement minus the true value of the measurand. Note: *Since a true value cannot be determined, in practice a conventional true value is used.* [VIM 1993]

The value of a result minus the true value. [IUPAC Compendium of Chemical Technology, 1985]

### B.7. Expanded Uncertainty

$U$ —quantity defining an interval about a result of a measurement that may be expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the measurand. Note 1 *The fraction may be regarded as the coverage probability or level of confidence of the interval.* Note 2 *To associate a specific level of confidence with the interval defined by the expanded uncertainty requires explicit or implicit assumptions regarding the probability distribution characterized by the measurement result and its combined standard uncertainty. The level of confidence that may be attributed to this interval can be known only to the extent to which such assumptions can be justified.* Note 3 *An expanded uncertainty  $U$  is calculated from*

*a combined standard uncertainty  $u_c$  and a coverage factor  $k$  using:  $U = k \times u_c$ .* [ISO GUM]

### B.8. Linearity

Defines the ability of the method to obtain test results proportional to the concentration of analyte. Note: *The Linear Range is by inference the range of analyte concentrations over which the method gives test results proportional to the concentration of the analyte.* [AOAC–PVMC]

### B.9. Measurand

Particular quantity subject to measurement. Note: *Specification of a measurand may require statements about quantities such as time, temperature and pressure.* [VIM 1993]

### B.10. Measurement

Set of operations having the object of determining a value of a quantity. [VIM 1993]

### B.11. Measurement Procedure

Set of operations, described specifically, used in the performance of measurements according to a given method. Note: A measurement procedure is normally recorded in a document that is sometimes itself a measurement procedure or measurement method and is usually in sufficient detail to enable the operator to carry out a measurement without additional information. [VIM 1993]

### B.12. Method Validation

1. The process of establishing the performance characteristics and limitations of a method and the identification of the influences which may change these characteristics and to what extent. Which analytes can it determine in which matrices in the presence of which interferences? Within these conditions what levels of precision and accuracy can be achieved? 2. The process of verifying that a method is fit for purpose, i.e. for use for solving a particular analytical problem. Note: *1. is applicable where a method is developed without any particular problem in mind. 2 is applicable where a method is being developed for a specific purpose. In analytical chemistry the other commonly encountered use of the term validation is in the context of instrumentation. Instrument validation is used to describe the process of establishing that an instrument at any given moment is able to perform according to its design specification This process might be achieved for example by means of calibration or performance checks.* [No reference given]

### B.13. Precision

The closeness of agreement between independent test results obtained under stipulated conditions. Note: *Precision depends*

only on the distribution of random errors and does not relate to the true value or specified value. The measure of precision is usually expressed in terms of imprecision and computed as a standard deviation of the test results. “Independent test results” means results obtained in a manner not influenced by any previous result on the same or similar test object. Quantitative measures of precision depend critically on the stipulated conditions. **Repeatability and Reproducibility** are particular sets of extreme conditions. [ISO 3534-1]

A measure for the reproducibility of measurements within a set, that is, of the scatter or dispersion of a set about its central value. [IUPAC Compendium of Chemical Technology, 1985]

#### B.14. Random Error

Result of a measurement minus the mean that would result from an infinite number of measurements of the same measurand carried out under repeatability conditions. Note: *Random error is equal to error minus systematic error. Because only a finite number of measurements can be made, it is possible to determine only an estimate of random error.* [VIM 1993]

The difference between an observed value ( $x_i$ ) and the limiting mean ( $\mu$ ); i.e.  $\delta = x_i - \mu$ . [IUPAC ‘Orange’ Book]

#### B.15. Repeatability (of a measuring instrument)

Ability of a measuring instrument to provide closely similar indications for repeated applications of the same measurand under the same conditions of measurement. [IUPAC ‘Orange’ Book]

#### B.16. Repeatability (of results of measurements)

Closeness of the agreement between the results of successive measurement of the same measurand carried out in the same conditions of measurement. [IUPAC ‘Orange’ Book]

#### B.17. Repeatability Limit “r”

The value less than or equal to which the absolute difference between two test results obtained under repeatability conditions may be expected to be with a probability of 95%. Repeatability (limit) is given by the formula:

$$r = t_{\infty} \times \sqrt{2} \times \sigma_r$$

where  $t_{\infty}$  is the Student’s two tailed value for  $\nu = \infty$  for a given confidence (normal confidence level state is 95% where the value is 1.96), and  $\sigma_r$  is the standard deviation measured under repeatability conditions (see A20.3). [ISO 3534-1]

#### B.18. Repeatability Standard Deviation

The standard deviation of test results obtained under repeatability conditions. Note: *This is a measure of dispersion of the distribution of test results under repeatability conditions. Sim-*

ilarly “repeatability variance” and “repeatability coefficient of variation” could be defined and used as measures of the dispersion of test results under repeatability conditions. [ISO 3534-1]

#### B.19. Reproducibility

Precision under reproducibility conditions, i.e. conditions where test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment. Note: *A valid statement of reproducibility requires specification of the conditions changed. Reproducibility may be expressed quantitatively in terms of the dispersion of the results.* [ISO 3534-1]

#### B.20. Reproducibility Limit “R”

The value less than or equal to which the absolute difference between two test results obtained under reproducibility conditions may be expected to be with a probability of 95%. Reproducibility (limit) is given by the formula:

$$R = t_{\infty} \times \sqrt{2} \times \sigma_R$$

where  $t_{\infty}$  is the Student’s two tailed value for  $\nu = \infty$  for a given confidence (normal confidence level state is 95% where the value is 1.96), and  $\sigma_R$  is the standard deviation measured under reproducibility conditions (see A21). [ISO 3534-1]

#### B.21. Reproducibility Standard Deviation

The standard deviation of test results obtained under reproducibility conditions. Note: *This is a measure of dispersion of the distribution of test results under reproducibility conditions. Similarly “reproducibility variance” and “reproducibility coefficient of variation” could be defined and used as measures of the dispersion of test results under reproducibility conditions.* [ISO 3534-1]

#### B.22. Standard Deviation

This is a measure of how values are dispersed about a mean in a distribution of values: the standard deviation  $\sigma$  for the whole population of  $n$  values is given by:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n}}$$

In practice we usually analyse a sample and not the whole population The standard deviation  $s$  for the sample is given by:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} \quad [\text{No reference given}]$$

### B.23. Standard Uncertainty

$u(x_i)$ —Uncertainty of the result of a measurement expressed as a standard deviation. [ISO GUM]

### B.24. Systematic Error

Mean that would result from an infinite number of measurements of the same measurand carried out under repeatability conditions minus a true value of the measurand. Note: *Systematic error is equal to error minus random error. Like true value, systematic error and its causes cannot be known.* [VIM 1993]

### B.25. True Value

Value consistent with the definition of a given particular quantity. Note: *This is a value that would be obtained by a perfect measurement. True values are by nature indeterminate. The indefinite article “a” rather than the definite article “the” is used in conjunction with “true value” because there may be many values consistent with the definition of a particular quantity.* [VIM 1993]

### B.26. Trueness

The closeness of agreement between the average value obtained from a large set of test results and an accepted reference value. Note: *The measure of trueness is normally expressed in terms of bias. The reference to trueness as “accuracy of the mean” is not generally recommended.* [ISO 3534-1]

### B.27. Uncertainty (of Measurement) i.e. Measurement Uncertainty

Parameter associated with the result of a measurement, that characterises the dispersion of the values that could reasonably be attributed to the measurand. Note: *The parameter may be, for example, a standard deviation (or a given multiple of it), or the width of a confidence interval. Uncertainty of measurement comprises, in general, many components. Some of these components may be evaluated from the statistical distribution of the results of a series of measurements and can be characterized by experimental standard deviations. The other components which can also be characterized by standard deviations, are evaluated from assumed probability distributions based on experience or other information. It is understood that the result of the measurement is the best estimate of the value of the measurand and that all components of uncertainty, including those arising from systematic effects, such as components associated with corrections and reference standards, contribute to the dispersion.* [VIM 1993]

### B.28. Validation

Confirmation by examination and provision of objective evidence that the particular requirements for a specified intended use are fulfilled. [ISO 8402:1994]

## Appendix C. Recommended approach for chromatographic-data analysis. Article is from *American Laboratory* and is reprinted with permission from the publisher, International Scientific Communications: Statistics in Analytical Chemistry Part 25 – Calibration Summary

In a modern chemical-analysis laboratory, virtually all of the testing equipment must be calibrated periodically. However, there is not a universally applicable calibration procedure that can be used in all cases. Much of the problem results from the fact that analytical instruments fall into one of two classes (i.e., instruments that immediately provide the results in usable units and instruments that do not).

An example of the former class is a balance scale. The object for which mass measurement is desired is placed on the balance pan and the numerical value of the mass appears on the readout. The calibration of this type of equipment is quite straightforward: Place a mass (or masses) of known amount(s) on the balance, read the results, and adjust as necessary.

An example of the latter class is a chromatograph. When an analyte reaches the detector, the signal is reported out by the computer in arbitrary units (e.g., peak area or peak height). Calibration in this instance is not as rapid as in the above example. A more involved procedure (i.e., regression) is needed to convert the raw data into useful units (typically concentration).

Since the beginning of this *American Laboratory* column, the overriding theme has been the statistics behind the sound calibration of these “regression-based” instruments. This article summarizes the basics that have been presented. For definitions of statistical terms that are used, the reader is referred to Part 24 in this series (*American Laboratory*, Nov/Dec 2006).

Regression-based calibration should have two objectives: 1) a curve that will transform sample data into concentration units (and do so without bias), and 2) provide (at a user-chosen confidence level) a statistically sound estimate of the uncertainty in any reported concentration. To accomplish these goals, three main steps are involved. First, a calibration study must be designed. Second, the study must be performed carefully in the laboratory. Third, the data must be evaluated statistically using a set of calibration (or, more generally, regression) diagnostics. This third step will ensure that the resulting curve meets the two calibration objectives listed above.

### C.1. Step 1: Design the calibration study

In designing the study, two decisions must be made: 1) the number of different concentrations of standard solutions that will

be prepared, and 2) the number of each solution's replicates that will be analyzed. In every study, there must be sufficient numbers of both concentrations and replicates to allow for: 1) detection of curvature in the data, 2) modeling of response standard deviation (to see if this statistic trends with concentration), and 3) use of the calibration curve (to predict sample concentrations) without extrapolation at either end. Additionally, the design may need to be adjusted if the analyst is pursuing a low detection limit or high precision.

In the designing process, it is helpful to propose a model and a confidence level for the calibration curve. A rule-of-thumb starting place is a  $5 \times 5$  design (i.e., five replicates of each of five concentrations, typically including blanks). However, the final plan must be based on the intended use of the calibration curve for sample predictions.

### C.2. Step 2: Perform the calibration study in the laboratory

While this step does not directly involve the use of statistics, a few comments are in order. No matter how carefully the study has been designed, if it is not performed properly in the laboratory, the resulting data will be compromised. Standards should be prepared in the pure solvent that is appropriate for the instrument at hand (see below for comments on dealing with sample-matrix issues). If blanks are included in the study, they must be prepared appropriately for the analytical method being studied. If standard preparation is subject to such things as contamination, or if standards degrade rapidly, appropriate action must be incorporated into the lab work.

### C.3. Step 3: Diagnose the calibration data statistically

This portion of the process involves seven basic parts:

1. *Plot response versus true concentration.* Evaluate the overall shape of the scatterplot.
2. *Determine the behavior of the standard deviation of the responses.* Plot the standard deviation versus concentration and fit with a straight line, using ordinary least squares as the fitting technique (the general equation for the line is: standard deviation =  $g + hx$ ). If the p-value for the slope is significant (i.e.,  $<0.01$ ), then the standard deviation trends with concentration; in such cases, weighted least squares must be used to fit proposed curves to the calibration data themselves. The formula for the weight is:

$$([g + hx]^{-2}) / (\text{Avg}[\{g + hx\}^{-2}])$$

3. *Fit the proposed model and evaluate  $R_{\text{adj}}^2$ .* Although  $R_{\text{adj}}^2$  is a weak statistical tool, the value should be close to 1.
4. *Examine the residuals for nonrandomness.* The ideal is to have the zero line pass through the mean of each concentration's residuals. In such a case, there will be a random scatter of the points about the zero line. If a distinct pattern (e.g., parabola or sine wave) exists, then the model probably is not adequate. Appearance of a "trumpet effect" indicates

that the standard deviation of the responses may be trending with concentration.

5. *Evaluate the p-value for the slope (and any higher-order terms).* For calibration data, the  $x$ -term will always be significant (i.e., the term's p-value should be  $<0.01$ ). For higher-order models to be appropriate, the coefficients for the additional terms should be significant as well.
6. *Perform a lack-of-fit (LOF) test.* If the p-value is  $<0.05$ , then the model is not adequate. The shape of the residual pattern should be used to help select an alternate model to be tested.
7. *Plot and evaluate the prediction interval.* This important step will indicate the uncertainty in sample estimates that are derived from this curve. The width of the interval will depend on the noisiness of the data and the confidence level that has been chosen.

The previous seven steps center on making two statistically sound choices: a model and a fitting technique. It must be emphasized that these selections are *independent* of each other. The model choice depends on the outcome of the lack-of-fit test (with help from the residual pattern). The fitting-technique choice depends on the behavior of the responses' standard deviations (supported by the presence or absence of a trumpet effect in the residuals plot).

Occasionally, no model that is tested is adequate, or a model that is adequate is not easily inverted for use in estimating sample concentrations. When a less-than-adequate model is selected for the calibration curve, the width of the uncertainty interval must be adjusted to account for the bias that exists. The procedure for this correction is found in Part 16 (*American Laboratory*, May 2005).

If the sample analytes are in a matrix other than pure solvent, and recovery problems are known or suspected, then a recovery study should be conducted postcalibration. Typically, the calibration design can be used for this study, too. Recovered concentrations are calculated via the pure-solvent calibration curve and plotted versus true concentration, thereby generating a second graph. These data are also modeled and diagnosed using the seven steps above. If the recovery is unacceptably low or high, the equation for the model can be used to correct the recovered concentrations to true values. The associated prediction interval gives the overall uncertainty (at the chosen confidence level) for the method, since this interval includes the uncertainty for both the calibration and the spiking processes.

As a final summary, the following is presented in hopes that it can serve as a useful reference for readers. The authors would more than welcome feedback on this summary article, and especially on the reference box.

### C.4. Calibration/Regression

Calibration (or any type of regression) involves the choice of: 1) a model, and 2) a fitting technique. These two choices are made *independently*; the result for one does *not* influence the choice of the other.

Tables A1 and A2 below guide the selection processes.

Table A1  
Matrix for selection of a regression model and fitting technique

	Model	Fitting technique
Practical choices	Straight line or quadratic	Ordinary least squares or weighted least squares
Statistical tests	Lack-of-fit (LOF) test	Slope's p-value (for plot of response standard deviations vs. concentration)

The *residual plot* is helpful for selection of both *model* and *fitting technique*. However, decisions are made also using the LOF test (for the model) and the slope's p-value (for the fitting technique). *Model*: Zero line ideally should pass through the mean of each concentration's residuals. Appearance of a nonrandom pattern (e.g., parabola, sine wave) is indication of lack of fit for the selected model. *Fitting technique*: Appearance of a "trumpet effect" (in either direction) is indication that the standard deviation trends with concentration.

Table A2  
Details on statistical tests used in regression diagnostics

	LOF test	p-value of slope
Null hypothesis	Model does not show lack of fit	Straight line with zero slope explains data adequately
Cutoff for p-value	0.05 (5%)	0.01 (1%)

## References

- [1] The Fitness for Purpose of Analytical Methods, EURACHEM, 1998, <http://www.eurachem.ul.pt/guides/valid.pdf>.
- [2] Quantifying Uncertainty in Analytical Measurement, EURACHEM/CITAC, second ed., 2000, <http://www.eurachem.ul.pt/guides/QUAM2000-1.pdf>.
- [3] Harmonised Guidelines for the Use of Recovery Information in Analytical Measurement, IUPAC/ISO/AOAC International/EURACHEM, 1995, <http://www.eurachem.ul.pt/guides/recovery.pdf>.
- [4] Guide to the Expression of Uncertainty in Measurement, ISO, Geneva, 1993.
- [5] A2LA Guide for the Estimation of Measurement Uncertainty in Testing, American Association for Laboratory Accreditation, 2002, [http://www.a2la.org/guidance/est\\_mu\\_testing.pdf](http://www.a2la.org/guidance/est_mu_testing.pdf).
- [6] Guidelines for Uncertainty Estimation, American Industrial Hygiene Association, 2006, <http://www.aiha.org/1documents/lab/AIHAGuidelinesforMeasurementofUncertainty.pdf>.
- [7] Interpretation and Guidance on the Estimation of Uncertainty of Measurement in Testing (APLAC TC 005), Asia Pacific Laboratory Accreditation Cooperation, 2004, [http://www.aplac.org/documents/web\\_docs/tc\\_005.pdf](http://www.aplac.org/documents/web_docs/tc_005.pdf).
- [8] CCIL Protocol for Estimating Measurement Uncertainty Using QC Data (Type A), Canadian Council of Independent Laboratories, 2003, <http://www.ccil.com/documents/CCIL%20Type%20A%20MU%20Protocol%20V1.0%2023June03.pdf>.
- [9] EA Guidelines on the Expression of Uncertainty in Quantitative Testing (EA-4/16), European co-operation for Accreditation, 2003, <http://www.european-accreditation.org/n1/doc/EA-4-16.pdf>.
- [10] Measurement Uncertainty in Testing (Technical Report No. 1/2002), EUROLAB, 2002, [http://www.eurolab.org/docs/el\\_11-01\\_02\\_7871.pdf](http://www.eurolab.org/docs/el_11-01_02_7871.pdf).
- [11] Guidelines for Calibration in Analytical Chemistry. Part 1. Fundamentals and Single Component Calibration, Pure Appl. Chem. 70 (1998) 993, <http://www.iupac.org/publications/pac/1998/pdf/7004x0993.pdf>.
- [12] Validation of Analytical Procedures: Text and Methodology, Q2(R1), International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, 2005, <http://www.ich.org/LOB/media/MEDIA417.pdf>.
- [13] Statistics in Analytical Chemistry, Part 25, Calibration Summary, Am. Lab., 39 (February 2007) 35, <http://www.iscpubs.com/articles/>.
- [14] Statistics in Analytical Chemistry, Part 1, A New American Laboratory Column, Am. Lab., 34 (September 2002) 44, <http://www.iscpubs.com/articles/>.
- [15] Statistics in Analytical Chemistry, Part 2, Measurement and the Real-Number Line, Am. Lab., 34 (November 2002) 38, <http://www.iscpubs.com/articles/>.
- [16] Statistics in Analytical Chemistry, Calibration: Introduction and Ordinary Least Squares, Am. Lab., 35 (January 2003) 18, <http://www.iscpubs.com/articles/>.
- [17] Statistics in Analytical Chemistry, Part 4, Calibration: Uncertainty Intervals, Am. Lab., 35 (March 2003) 60, <http://www.iscpubs.com/articles/>.
- [18] Statistics in Analytical Chemistry, Part 5, Calibration: Calibration Design, Am. Lab., 35 (June 2003) 30, <http://www.iscpubs.com/articles/>.
- [19] Statistics in Analytical Chemistry, Part 6, Calibration Design, Am. Lab., 35 (July 2003) 19, <http://www.iscpubs.com/articles/>.
- [20] Statistics in Analytical Chemistry, Part 7, A Review, Am. Lab., 35 (September 2003) 31, <http://www.iscpubs.com/articles/>.
- [21] Statistics in Analytical Chemistry, Part 8, Calibration Diagnostics, Am. Lab., 35 (November 2003) 40, <http://www.iscpubs.com/articles/>.
- [22] Statistics in Analytical Chemistry, Part 9, Calibration Diagnostics (Continued), Am. Lab., 36 (February 2004) 64, <http://www.iscpubs.com/articles/>.
- [23] Statistics in Analytical Chemistry, Part 10, Calibration Diagnostics (Concluded), Am. Lab., 36 (March 2004) 46, <http://www.iscpubs.com/articles/>.
- [24] Statistics in Analytical Chemistry, 2007, Part 11, Calibration Example, Am. Lab., 36 (May 2004) 40, <http://www.iscpubs.com/articles/>.
- [25] Statistics in Analytical Chemistry, Part 12, Calibration Example 2, Am. Lab., 36 (July 2004) 38, <http://www.iscpubs.com/articles/>.
- [26] Statistics in Analytical Chemistry, Part 13, Calibration Example 3, Am. Lab., 36 (September 2004) 28, <http://www.iscpubs.com/articles/>.
- [27] Statistics in Analytical Chemistry, Part 14, Calibration Example 4, Am. Lab., 36 (December 2004) 32, <http://www.iscpubs.com/articles/>.
- [28] Statistics in Analytical Chemistry, Part 15, Calibration Example 5, Am. Lab., 37 (February 2005) 48, <http://www.iscpubs.com/articles/>.
- [29] Statistics in Analytical Chemistry, Part 16, Calibration Example 5 (Concluded), Am. Lab., 37 (May 2005) 37, <http://www.iscpubs.com/articles/>.
- [30] Statistics in Analytical Chemistry, Part 17, Calibration Example 6, Am. Lab., 37 (August 2005) 32, <http://www.iscpubs.com/articles/>.
- [31] Statistics in Analytical Chemistry, Part 18, Calibration Example 6 (Concluded), Am. Lab., 37 (September 2005) 24, <http://www.iscpubs.com/articles/>.
- [32] Statistics in Analytical Chemistry, Part 19, Internal Standards, Am. Lab., 37 (December 2005) 23, <http://www.iscpubs.com/articles/>.
- [33] Statistics in Analytical Chemistry, Part 20, Recovery Curves, Am. Lab., 38 (February 2006) 54, <http://www.iscpubs.com/articles/>.
- [34] Statistics in Analytical Chemistry, Part 21, Recovery Curves and Internal Standards (Concluded), Am. Lab., 38 (April 2006) 30, <http://www.iscpubs.com/articles/>.
- [35] Statistics in Analytical Chemistry, Part 22, Lack-of-Fit Details, Am. Lab., 38 (June/July 2006) 34, <http://www.iscpubs.com/articles/>.
- [36] Statistics in Analytical Chemistry, Part 23, Degrees of Freedom in the Lack-of-Fit Test, Am. Lab., 38 (October 2006) 38, <http://www.iscpubs.com/articles/>.
- [37] Statistics in Analytical Chemistry, Part 24, Glossary, Am. Lab., 38 (November/December 2006) 25, <http://www.iscpubs.com/articles/>.