STATISTICAL ANALYSIS

Uncertainty—A Chemist's View

WILLIAM HORWITZ
U.S. Food and Drug Administration, Center for Food Safety and Applied Nutrition, Washington, DC 20204

Complete characterization of the performance of analytical methods requires an evaluation of the halo of uncertainty bracketing the reported result. Achieving a satisfactory estimate of this uncertainty is more important than how this estimate is produced. Enumeration of all conceivable error components—the so-called error budget approach—is one way to estimate the uncertainty, but it is not the only way. In fact, when applied to analytical chemistry this approach is likely to (*1*) overlook important variables and double count others, (*2*) avoid considering unknown and unknowable interactions and interferences, and (*3*) adjust for missing variables with an uncontrollable "Type B" component. The problem is one of experimental design. Alternative and more efficient ways of estimating uncertainty in analytical chemistry include the Youden ruggedness procedure, accompanied by a bonus of optimization, and the all-encompassing interlaboratory method-performance trial.

Scientists involved in regulatory control of consumer commodities and the environment have a practical interest in extending control of analytical measurements beyond establishment of the true value and into control of the variability. Our task is the establishment and enforcement of legal standards dealing with everyday problems. As a starting point we have to accept, as our basis for judgment, criteria imposed by legal requirements, which express uncertainty in terms such as "preponderance of evidence" and "beyond a reasonable doubt."

By definition, the true value is nearly always within the range of the highest and lowest values assigned to an expanded uncertainty, somewhat equivalent to a 95% confidence interval. The unusual aspect of this uncertainty from a chemist's point of view is that it consists not merely of the random error component of the error model but also of the systematic error component.

Received January 30, 1998. Accepted by GL February 11, 1998.

Presented at the National Institute of Science and Technology, Gaithersburg, MD, Dec. 2, 1997. An abbreviated version was given at the meeting on uncertainty held at the Royal Society of Chemistry, London, United Kingdom, December 8, 1997, under the assigned title of "A North American View on Uncertainty."

The concept of uncertainty is purported to be explained in a "bible" endorsed by 8 international organizations and published by the International Organization for Standardization (ISO) (1). The U.S. National Institute of Science and Technology (NIST) also produced a similar document (2) that was shorter but almost as incomprehensible to the ordinary chemist as the ISO explanation. The NIST document states that the uncertainty statement was developed primarily "for reporting results of determinations of fundamental constants, fundamental metrological research, and international comparisons of realizations of SI units," and it applied specifically to Standard Reference Materials (SRMs). I suspect it was expanded to ordinary analytical chemical measurements by default, without much investigation of its implications.

As a result of U.S. participation in the Codex Alimentarius Committee on Methods of Analysis and Sampling, we had to look into the implications of the utility of "uncertainty" as applied to chemical measurements. The Codex Alimentarius is an international activity under the auspices of the United Nations' Food and Agriculture Organization and World Health Organization for the purpose of establishing standards and practices for control of international trade in foods. At the last 2 meetings, the discussion alerted us to a requirement of some national accrediting organizations that requested a standard uncertainty budget for each of the hundreds of analytical methods used in the typical food laboratory for examining components and contaminants of food. Therefore, we had to look into the problem to be in a position to prepare comments and develop a policy. We ultimately analyzed uncertainty from the point of view of error structure, estimation, and practical statistical problems.

The Error Structure of Analytical Chemistry

The idea behind the uncertainty concept is unquestionably sound. Chemists have been taught to calculate the standard deviation of a set of data, and they understand that this standard deviation describes the distribution of values obtained by their analyses. They have been told by practically every textbook of analytical chemistry to include in their reports an indication of the uncertainty of their results, but this admonition has fallen mostly on deaf ears. One of the most important reasons chemists have not eagerly jumped on the bandwagon of uncertainty is that the reported statistic applied to trace analysis is frightfully large, especially when the "t" correction factor is also applied. If we consider that the typical repeatability standard deviation of aflatoxin analyses within a laboratory at the 10 μg/kg

level is 15%, the expanded uncertainty from this component alone is 30%. But results of the U.S. Food and Drug Administration (FDA) are not accepted per se as correct. Defendants in legal actions have the right to present their results to a court on a par with the government. Thus we have to allow for reproducibility standard deviation among laboratories.

We have found this among-laboratory variability to be about double the within-laboratory variability. Therefore we are faced with an expanded uncertainty from random error alone of 60% in these analyses, a value that does not inspire much confidence in the reliability of our work. Such large random error components swamp any allowance we might want to make for systematic errors arising from losses in extracting aflatoxins from food and from diffusion during measurement by thin-layer chromatography (TLC). We still use TLC for this quantitation because the analytical error is only 10% of the total error of the analysis. The major error, typically 90%, arises from sampling inhomogeneous bulk commodities such as peanuts and corn, where a single kernel can contain the equivalent of 100 mg aflatoxin/kg and a million kernels will contain none.

Modern Analytes

Modern analytical chemistry has been proceeding relentlessly in the direction of analyzing smaller amounts of material containing lower concentrations of analytes. The ultimate has been reached in reports of single-molecule detection systems. (What is the uncertainty of single-molecule detection?) As the quantity measured becomes smaller, the relative uncertainty in its measurement becomes larger.

Modern analytes are also far more complex than the inorganic elements and compounds of a half-century ago. The complexity is not merely in the structure of organic molecules but also in their presence in a melange of homologous series. When you have isolated and characterized a product in a synthetic mixture or in a biological tissue, you can be sure that it is accompanied by a number of relatives that also have to be isolated and characterized and their pertinence to your problem determined.

Quality Control

About 2 decades ago, with introduction of mandatory good laboratory practices into the pharmaceutical industry, laboratories were required to demonstrate to government officials or to third-party assessors that the results they produce are reliable. This requirement now has developed into a large bureaucratic operation in some countries and is rapidly moving in that direction in the United States under the label of accreditation. Most of these programs rely primarily on review of documentation, presumably reflecting performance. The huge increase in international trade of food products has produced a demand for equivalency of analytical results, particularly in pesticide residue and natural contaminant analyses and microbiological assays. Equivalency means that analytical results produced by a laboratory in the importing country should correspond to those produced by a laboratory in the exporting country. (We will not consider microbiological assays here.) Equivalency requires good control of variability and minimal differences in bias (systematic error) by the laboratories involved.

Error Control

About a century ago, when government control of trade in commodities was introduced for fertilizers, feeds, foods, and drugs, differences in analytical results were thought to arise from use of different analytical methods. Therefore, for many decades the major emphasis in regulatory control was standardization of analytical methods. Every agricultural segment developed its own trade association and its professional society to recommend methods of analysis to be used as a basis for paying the grower for the raw commodity and for controlling the quality of the finished product by the manufacturer. Government laboratories developed their own organization that supplied methods of analysis that would be used for enforcement rather than for establishment of economic value.

Change in Analytes of Interest

As government regulatory agencies shifted their emphasis from economic adulteration and misbranding to safety of the food supply, the analytes of interest became those present in micro and trace quantities rather than in macro amounts. These analytes no longer were amenable to classical macro gravimetric and volumetric techniques that depended on the skill of the analyst. Analytical chemistry became instrumental chemistry, demanding electronic and computer skills. Stoichiometric chemistry gave way to separation chemistry. Chromatographic separations, instrumental measurements, and mass spectral identification have revolutionized analytical chemistry.

Shift in Error Structure

The shift in analytical technique was accompanied by a shift in the error structure of analytical chemistry. Methods in vogue a half-century ago were used to introduce chemists to the fundamentals of chemical reactions and physical chemistry. Results were traceable to the basic SI units. My textbook of analytical chemistry contained lengthy chapters on calibration and use of weights and balances, volumetric flasks, pipets, and burets. Modern textbooks almost ignore these topics, plunging into electronics, electrochemical methods, chromatographic separations, and spectrometry—mass, radiation, nuclear magnetic, and vibrational. Modern textbooks also contain a chapter on statistics and error analysis, including an introduction to confidence intervals and propagation of uncertainty with considerable emphasis on calibration, but not of weights, volumes, or voltages, but rather of physical properties and characteristics such as light intensity as revealed by recorder responses and computer printouts, presmoothed and outlier-purged for your convenience.

(**a**) *Traceability.*—The first major shift was a loss of traceability to SI units. Metrologists finally are beginning to realize this fact and have loosened their descriptions of traceability processes by substituting the phrase "suitable standards" or its equivalent for "SI units," thus sanctioning the existence of surrogates such as certified reference materials and inhouse standards. In the early stages of developing a method of analysis for isolating and identifying unknown contaminants, toxicants, pollutants, metabolites, and biologically active constituents, a living organism is usually the indicator of the analyte's pres-

ence or absence. When the responsible analyte is identified and characterized, a few milligrams of a fairly pure material, separated by preparative chromatography or other techniques, must serve as the standard. Such standards suffice in exploratory research, but their use is undoubtedly accompanied by a large but unknown systematic error. They qualify as standards only by necessity. Of the group of naturally occurring, biologically active contaminants, such as 2,3,7,8-tetrachlorodibenzo-*p*-dioxin, polychlorinated biphenyls, and polynuclear aromatics, it takes about 2 decades before a fully qualified SRM becomes available. Fortunately, we have been able to develop life-saving drugs, identify biologically active compounds and their mechanism of action, and control life-threatening contaminants without the luxury of the traceability and uncertainty concepts.

(**b**) *Ratio of systematic to random error.*—The second shift in the error structure of measurements was a change in the relative proportions of systematic to random error. When analytical reactions were based on stoichiometry, most errors were systematic, arising from solubility and contamination of precipitates, incomplete reactions and extractions, and competition from interferences. The laws covering these occurrences were well known, and the deviations they caused were functions of time, temperature, and concentration. When these potential variables were controlled, small consistent systematic errors resulted.

However, the magnitude of these deviations usually were determined experimentally, accompanied by random measurement errors, which were generally of second order and could be neglected.

(**c**) *Stoichiometric to calibration chemistry.*—Modern analytical chemistry, however, is no longer exclusively "stoichiometric chemistry." It is "calibration chemistry," based on the availability of a standard. Without a standard, there is no link to a specific analyte; the link is only to a physical property. Many of the steps of our analytical methods are designed to isolate the single analyte of interest to avoid mistaking its response from that of other, more innocuous analytes. Much analytical chemistry is devoted to attaining specificity, or at least selectivity, in our analyses. We are indeed fortunate that chromatography evolved just in time to handle the explosion in pesticides, food additives, contaminants, and pollutants made available by modern industry.

But what we achieve in one area of measurement is compensated for by deterioration in another. We are able to isolate and measure micro-, nano-, and picograms of these complex compounds in such complicated matrixes as biological tissues, but we have to pay for it in terms of reduced reliability—larger systematic errors and poorer precision—and considerably greater economic cost.

Quantitation Limit

Lloyd Currie of NIST suggested that the limit of reliable measurement in analytical chemistry is reached when the relative precision exceeds about 10%. This figure, however, is too conservative for practical purposes. Such a limit would make impractical the enforcement of action levels against what toxicologists regard as threats to human health by chlorinated pesticides, polynuclear aromatic hydrocarbons, mycotoxins, marine toxins, and ubiquitous air pollutants such as NO_x and SO_2. The FDA has reported examining packaged milk for 2,3,7,8-tetrachlorodibenzo-*p*-dioxin at the parts-per-quadrillion (pg/kg; 10^{-15}) level (3). But they did not tell you at what cost in resources and uncertainty. Quotations from commercial laboratories for this determination are of the magnitude of $1000 per value, and the expanded uncertainty undoubtedly approaches 100%. Routine analyses of pesticide residue in foods have a relative standard deviation within laboratories (RSD_r) of 10–15% at the fraction of a part-per-million (mg/kg; 10^{-6}) level. Similarly, routine aflatoxin and drug residue determinations yield RSD_r values of 15–20% at the 10 ppb (μg/kg; 10^{-12}) level. These values must be doubled when variability among different laboratories is also considered. When the Delaney (zero-tolerance) clause was added to the Food, Drug, and Cosmetic Act in 1954, 1 ppm was considered as the determination limit of using elegant analytical chemistry.

These values encompass only random error. The FDA has stated that we should allow for recovery variability in our examinations of tissues for pesticide and drug residues of 80–110% at concentrations above 0.1 mg/kg and 60–120% below that level (4). These specifications recognize that systematic error can be measured only by a process that unfortunately entails the presence of random error. At the level of interest, it appears that the 2 components are of approximately equal magnitude.

Consequences of Violating the Limit of Quantitation

The focus here is on the magnitude of deviations and variabilities that we have indicated and their practical effect. The distribution of measurements taken in this region is so broad that an occasional overlap of zero is expected. When this occurs with an analyte, a false negative is reported; when it occurs with a blank, a false positive is observed. In our reviews of interlaboratory studies of mycotoxins (5) and polychlorinated contaminants (6), we had to invent a new crude method performance parameter, "fraction positive," as reported by the group of laboratories on each low-level test sample.

Theory of False Negatives

Assuming we have a normal distribution of analytical estimates—and Thompson and Howarth (7) argue strongly that this is the case—the appearance of false negatives is a mathematical consequence of working with high relative standard deviations (RSDs). If we consider a normal distribution with a mean of 1 unit and an RSD of 100%, the mean minus 1 standard deviation is 0. From normal distribution tables, about 84% of values would be in the positive region and 16% would be in the negative region. We avoid the apparent paradox of negative concentrations by pointing out we are measuring differences in a physical property such as light intensity or voltage. Such differences can take on any value, positive, negative, or zero. Some of these signals transformed to concentrations through a calibration curve become the false negatives. If you have never seen data with RSDs approaching 100%, you have not examined the legitimate multiple laboratory data from interlaboratory studies of analytes such as lead, aflatoxins, or dioxin congeners at very low levels, $\lesssim 10^{-9}$. Every laboratory supplies

what it believes is a valid estimate, supported by quality control parameters. Yet taken as a group, they are seen to contain invalid data. The brilliant Tomas Hirschfeld was aware of this inconsistency when he wrote (8) that in extending the lower limits of analytical chemistry we may be setting goals that "involved beating the Heisenberg principle, Shannon's limit, or the second law of thermodynamics, all of which can take rather surprising forms under extreme conditions."

Biological Variability

For completeness, I will mention the error structure of biological properties. Here, all error is random by default; you never know or can even approximate the "true" value. How tall should a tree be or how many fish are in the sea? You can measure biological properties physically or chemically or by count, but there is no way of knowing or guessing what the property should be. The error structure of biology is truly "floating," not anchored or traceable to anything. Biological properties do have considerable variability, as demonstrated by the basic disagreements among toxicologists regarding interpretation of toxicity studies.

Summary

To summarize what has been stated so far, the nature of analytical chemistry has changed from the metrology of stoichiometric reactions to the utilization of calibration curves produced by computer-controlled instruments. The magnitude of measurements has shifted from macro quantities (%) to traces ($<10^{-6}$). Simultaneously, the relative proportions of systematic error to random error has become inverted, with random error now often predominating over systematic error. Further, the total error has been pushed into regions where uncertainties are measured in terms of the number of false negatives and false positives, changing the question from "how much" to "is the analyte present at all."

Estimation of Uncertainty

Overall Equation

Estimation of uncertainty by the ISO approach is based on the classical equation used in the theory of errors:

$$Y = f(X_1, X_2, \ldots X_N)$$

The uncertainty documents (1, 2) assume that the quantity Y is a function determined by a number of variables $X_1, X_2, \ldots X_N$. Small differences in Y are considered to arise from small differences in the constituent variables. The total change in Y is composed of a series of partial derivatives of Y with respect to X, holding all others constant and summing the results. Mathematically, a similar partial derivative expression applies to variance and thus to uncertainty. For this expression to hold, however, all errors must be independent so that no "cross terms" develop. This is the basis for the "error budget" approach to calculating uncertainty by the ISO procedure. All possible sources contributing to the total error are listed, and their extremes are estimated.

Estimating Values for Individual Error Components

According to ISO (1), components of uncertainty are derived by construction of an "error budget" that lists all sources of systematic and random variability. The maximum magnitude of each is estimated, and the total is obtained. Some components are estimated by truncation of a Taylor's series by assuming that δx is small. In trace analysis, δx of magnitude >10% cannot be assumed to be small. To the extent that the assumption of linearity is not valid, factors estimated in this manner are likely to differ considerably from their experimental values.

The Guide (1) is aware of the concept of "uncertainty of uncertainty" or the standard deviation of the standard deviation of the mean (standard error [se] in the United States). Most chemists are unaware of this second-order variability. Table 1 shows how uncertain a standard deviation can be when calculated from only a few observations. Given these values, I do not understand how we analytical chemists could have been delivering respectable analyses that reflect reality. When we can check our analytical values by an entirely independent assay, such as recovery of metals from ores or titrating the therapeutic effect of vitamins on depleted rats, the verifications are remarkably close.

EURACHEM Presentation

The European organization devoted to improving analytical work, called EURACHEM, tried to assist analytical chemists with respect to uncertainty by publishing a companion volume (9) to the basic ISO exposition, using examples from analytical chemistry. All examples in the ISO and NIST documents are from physics or engineering. Unfortunately, this valiant attempt did not succeed. The examples, although based on analytical techniques, emphasized physical aspects of the variabilities—such as volumes of flasks, temperature corrections, and weighing variabilities—and overlooked important chemical sources of variabilities such as the effect of the acid base CO_2 on the endpoint of the acid–base titration and different strengths of the parent acid that was diluted for leaching cadmium and lead from ceramic food utensils. Errors arising from chemical effects were handled as "Type B" effects, based on the best guesses of experienced analysts, when they could have been obtained almost as easily by independent replication experiments. The apparent unfamiliarity of the authors with some

Table 1. The standard deviation of the standard error as a function of number of observations (1)

Number of observations	Standard deviation of se, %
2	76
3	52
4	42
5	36
10	24
20	16
30	13
50	10

laboratory operations led to double counting of some errors, by assessing them directly and then by unconsciously including them in more-inclusive operations. Some of these exercises resemble the medieval polemics of determining how many angels can dance on the head of a pin or the more modern counterpart characterized by the phrase "bean counters."

Unexplained was why the entire allowable range of specifications for volumetric ware, such as pipets, has to be included in the overall uncertainty when the actual volume of the pipet was determined in the course of establishing the uncertainty of the calibration operation. Similarly, we fail to understand why the entire range of allowable temperatures must be included in the uncertainty calculation when this is a fixed, local variable in actual practice, easily determined by reading a common thermometer. Fortunately, such uncertainties generated by physical variables are relatively small compared with the variabilities produced by chemical operations.

Finally, in many cases, uncertainties emerging from analyses did not match variabilities determined experimentally. The uncertainty of the acid–base titration could be reduced experimentally by a factor of 2 by reversing the roles of the acid and alkali (10). Furthermore, how can physical variables be represented by rectangular distributions when the fundamental error equation suggests the components are normally distributed variables? The answer lies in inventing the Type B uncertainty "based on scientific judgment" that permits estimating (guessing) the extreme limits the property can take and assuming equal probability for all values within the range. This is equivalent to a rectangular distribution. The best estimate of the uncertainty of a rectangular distribution is that interval, Δ, divided by $\sqrt{3}$.

Although there is no clear demarcation, typically physical measurements are made singly. If multiple measurements are made, the first is usually completed before the next one is begun. Chemical operations usually are done in groups, and the same operation is conducted on the group before the next operation is begun. Automation is characteristic of routine chemical control operations starting with holders with positions for as many as 96 test solutions. Immunoassays are conducted from plates containing 96 wells. Even classical gravimetric and volumetric operations ("wet chemistry") often are performed on a dozen test portions consecutively. A single calibration built into a computer automatically responds to the signal, transforms the current or voltage into an amount, and prints out the analyte concentration, taking into consideration the mass of the test portion. Local reference materials are placed into the series that can be used as a check on systematic error, random error, and drift or for quality control operations. The nature of the test operations has several implications: (*1*) Chemical measurements are more likely to be related; that is, within a series, they lack statistical independence. (*2*) Chemical measurements are more likely to be undertaken under reproducibility conditions (changed principle, method, observer, instrument, standard, location of use, and time); physical measurements are more likely to be undertaken for local use, that is, under repeatability conditions (same procedure, observer, instrument, location, and time). (*3*) The magnitude of the total error of many chemical measurements are several decades greater than those of physical measurements. Chemical measurements have relative errors of the order of units and tens of percent; many physical measurements are given in parts per million, that is, accurate to the fifth and sixth significant figure.

In other words, chemical measurements and physical measurements operate in entirely different environments—completely different ballparks—and we can expect different effects and influences.

Assumptions

Attempting to calculate errors by summation, metrologists have overlooked numerous assumptions inherent in the use of the propagation of error theory as applied to the high variabilities of modern trace analysis. The assumptions are listed by H.H. Ku of the National Bureau of Standards (11) as follows: (*1*) In the Taylor expansion of a nonlinear continuous function, $W = F(x, y)$, where W is some physical property and x and y are measured values with random errors ε_i, terms containing higher orders of ε_x and ε_y are neglected. (2) The estimate of a specific value of W, $\hat{w}$, is normally distributed for large n, but is it still a good approximation for small n? (3) Estimates always assume that measurements of x and y are independent, but is this still a reasonable assumption when the same calibration curve is always used? Ku follows up with looking into the higher moments of the normal distribution. Analytical chemists rarely have enough values to investigate the second-order moment (variance) of their distributions, let alone the higher moments of skewness and elongation.

Most revealing in Ku's chapter is his further discussion of neglecting higher order terms when the standard deviations are small in comparison with their respective means. ". . . the second and higher order partial derivatives evaluated at the means do not give rise to abnormally large numbers. This is usually true in the field of physical science, since errors of measurement are usually of the order of 1 part in 1000, or parts per million . . ."

Most of the examples given in the discussions of uncertainty are from physics and engineering. The EURACHEM document is about the only one that has looked into the field of chemistry, except for a very recent exposition (12) by the same authors.

Alternative Means of Calculating Uncertainty

The ISO Guide (1) provides a way of calculating uncertainty that has been accepted in metrology and physics and adopted, probably by default, by national standards organizations as applicable to all other measurements. It also has been accepted by accrediting organizations who send out assessors to evaluate laboratories according to ISO Guides. I understand that these assessors expect to find error budgets in place for each of the hundreds of analytical methods routinely used in food laboratories today. The 2 organizations that have evaluated implementation of the concept of uncertainty in analytical chemistry have endorsed the utility of the idea but not the rigid implementation through the error budget approach. The Analytical Methods Committee of the Royal Society of Chemistry of the UK (13) advocates an all-inclusive "top-down" approach obtained through an interlaboratory study that includes even systematic

errors (biases) of each laboratory in the sample of laboratories, as transformed to a random error. The Nordic Analytical Committee, representing the food laboratories of the Scandinavian countries, also rejected the error budget model as too complex for food laboratories. They too provided a "top-down" approach but based on experimental data generated in the individual laboratory (14). They characterized their uncertainty as "internal reproducibility standard deviation" and instructed that systematic errors be corrected for. Analysis of certified reference materials, participation in proficiency testing schemes, and use of recoveries and reference methods (for empirical analytes) were advocated for eliminating systematic error.

Chemical Measurement Process versus Measurement

The concepts of accuracy and precision apply to the chemical measurement process, not to individual measurements (15). Natrella recommended that, in the case of very high accuracy and precision as in wavelength measurements with 8–10 digits, the last significant figure should carry the burden of indicating reliability (15). When the imprecision is not negligible, however, she recommended use of an uncertainty statement consisting of the outer bound of the systematic error increased by at least twice the standard error. This is the format currently recommended for "expanded uncertainty."

Formulation of Uncertainty: Error Budget

The ISO approach (1, 12) lists all conceivable sources of uncertainty, assigns an uncertainty to each item in the form of a standard deviation, and sums them in quadrature (takes the square root of the sum of squares of the standard deviations). Recommendations for obtaining uncertainty statements are given in a more understandable, if abbreviated, form for analysts by Ellison et al. (12). The analyst is to write a clear statement of the relationship among the measurand (usually analyte concentration) and the parameters on which it depends, such as weight of test portion, volume of solvents, concentration of standards, chromatographic conditions, instrument characteristics, etc. These usually involve times, temperatures, weights, and volumes, as well as chemical relationships. A standard deviation is assigned or measured for each item, random and systematic. These standard deviations are squared, the squares are summed, and the square root of the sum is taken as a "standard uncertainty." This term is doubled to give "expanded uncertainty," a range analogous to the 95% confidence interval.

The advantages and disadvantages of the error budget approach to expressing uncertainty, based on the presentation by Ellison et al. (12), are shown in Table 2.

Experimental Design

Experimental design considers the most appropriate selection of samples to determine the distribution of a population. In many cases, a random selection of samples is the most efficient design. A random sample is not very efficient when variability and interactions are extremely high, as in agricultural field trials, where the idea of experimental design originated. Here, assigning samples in a definite pattern to cover the available area is required to provide adequate coverage of extreme values. The experimenter is not interested in absolute values but in relative values of different factors. Experimental material is easily available for replication, and no standards are required. In physics and metrology, we have the other extreme: Fewer factors and interactions are involved, variability is low, standards may be available, and corrections for systematic errors can be applied. Usually the property is well defined, and a measurement procedure need not be specified. Analytical chemistry is intermediate: Variability is moderate, interactions (interference) are considerable, and measurement methods often are lengthy and have an influence on the measurement. Therefore, the measurement procedure must be specified in detail. In both chemistry and physics, absolute values are usually sought, although in many cases of manufacturing operations and characterization of material properties, relative values are satisfactory.

According to Youden (16), there are many ways of obtaining an estimate of the error in the difference, or in the ratio of the 2 magnitudes associated with the comparison of 2 items. The direct repetition of measurements is the simplest approach that involves the least computation. Simple repetition is vulnerable to "memory" on the part of the operator. Also there is often a failure to provide the opportunity for errors to manifest themselves. "The actual value on a certificate may be a random error from the viewpoint of a national laboratory but it is a systematic error from the viewpoint of the calibrating laboratory."

In the various sciences, the source and magnitude of errors and uncertainties are different, although at the boundaries, these characteristics may blend into each other. But there is no a priori reason that these estimates can or must be developed in the same way. The error budget approach is appropriate for metrology, because all or many of the variables affecting the measurement can be anticipated and their magnitude estimated. But try to conceive setting up an error budget for expressing the uncertainty in the yield of corn from a fertilized plot when you have no control over the critical variables of temperature and rainfall. Uncertainty would have to be formulated on a historical basis, by using a high probability that the extremes will be avoided. In physical chemistry, experiments would be performed much like in physics, with most variables anticipated and controlled, but biochemistry would mimic agriculture. Analytical chemistry would vary over the full spectrum, depending on the complexity of the matrix and the potential for unanticipated interactions.

Alternative Means of Developing Uncertainty

The Analytical Methods Committee of the Royal Society of Chemistry of the United Kingdom has proposed an alternative method of estimating uncertainty, designated as the "top-down" approach (13). Here the all-inclusive variability in the analytical environment is obtained by an interlaboratory study. A sample of laboratories analyzing homogeneous, uniform materials by a common method provides an actual value for the property, usually analyte concentration. The variance can be apportioned into 2 components: A pooled within-laboratory random component and a mean among-laboratory bias that

Get Your Own Copy— Subscribe To The *Journal Of AOAC INTERNATIONAL!*

The Scientific Association Dedicated to Analytical Excellence®

AOAC INTERNATIONAL®

	North America (USA, Canada, Mexico)			Outside North America		
	*Member 1998	Nonmember 1998	Institutional 1998	*Member 1998	Nonmember 1998	Institutional 1998
1 year	☐ $181	☐ $249	☐ $270	☐ $232	☐ $300	☐ $320
2 years	☐ $331	☐ $452	☐ $490	☐ $432	☐ $554	☐ $591
Single Copies	☐ $ 58	☐ $ 58	☐ $ 58	☐ $ 66	☐ $ 66	☐ $ 66
Courier delivery outside U.S.				☐ $233	☐ $233	☐ $233

AOAC-J, c/o Signet Lockbox, P.O. Box 75198, Baltimore, MD 21275-5198

Please enter my subscription to the ***Journal of AOAC INTERNATIONAL*** as checked above. I have enclosed payment of $_____ (Subscriptions are on a calendar basis, 6 issues: Jan., Mar., May, July, Sept., Nov.)

☐ Check Enclosed (U.S. funds drawn on U.S. banks only) ☐ Charge my ☐ VISA ☐ MasterCard ☐ Diners Club ☐ American Express ☐ JCB ☐ Eurocard

Card Number_____ Exp. _____ Signature_____ Date _____

Send to: Name _____

Employer _____

Street Address _____

City _____ State(Province) _____ Postal Code _____

Country _____

*Member Number _____ (Must be provided to get member price.)

Become A Member Of AOAC INTERNATIONAL And Reap These Benefits:

AOAC INTERNATIONAL®

The Scientific Association Dedicated to Analytical Excellence®

- Opportunities to participate in the unique AOAC methods validation processes
- Professional advancement
- Peer recognition
- Discounts on meeting registration fees, courses, and all AOAC publications
- Membership directory
- Free subscription to *Inside Laboratory Management*—AOAC's monthly magazine
- Assistance in obtaining methodology information
- Job ads in *Inside Laboratory Management*
- **All for only U.S. $75 per year!**

_____ Application for AOAC INTERNATIONAL Membership _____
(Please Print)

Name _____

Employer _____

Address _____

City _____ State (Province) _____ Postal Code _____

Country _____ Phone Number_____ Fax Number_____

I am with ☐ private industry, ☐ a government agency, ☐ academia, ☐ other _____
(please specify)

Education (Scientific degrees): _____

Title and brief description of duties: _____

I hereby apply for membership in AOAC INTERNATIONAL. If accepted, I agree to abide by its Bylaws.

Signature_____ Date _____

To become a member, complete and sign this application and return it with a check for $75.00 or complete credit card information to:

AOAC INTERNATIONAL-J, c/o Signet Lockbox, P.O. Box 75198, Baltimore, MD 21275-5198

(U.S. funds drawn on U.S. banks only)

Become A Member Of AOAC INTERNATIONAL And Reap These Benefits:

- Opportunities to participate in the unique AOAC methods validation processes
- Professional advancement
- Peer recognition
- Discounts on meeting registration fees, courses, and all AOAC publications
- Membership directory
- Free subscription to *Inside Laboratory Management*—AOAC's monthly magazine
- Assistance in obtaining methodology information
- Job ads in *Inside Laboratory Management*
- **All for only U.S. $75 per year!**

AOAC INTERNATIONAL®

The Scientific Association Dedicated to Analytical Excellence®

Application for AOAC INTERNATIONAL Membership

(Please Print)

Name _____

Employer _____

Address _____

City _____ State (Province) _____ Postal Code _____

Country _____ Phone Number _____ Fax Number _____

I am with ☐ private industry, ☐ a government agency, ☐ academia, ☐ other _____
(please specify)

Education (Scientific degrees): _____

Title and brief description of duties: _____

I hereby apply for membership in AOAC INTERNATIONAL. If accepted, I agree to abide by its Bylaws.

Signature _____ Date _____

To become a member, complete and sign this application and return it with a check for $75.00 or complete credit card information to:

AOAC INTERNATIONAL-J, c/o Signet Lockbox, P.O. Box 75198, Baltimore, MD 21275-5198

(U.S. funds drawn on U.S. banks only)

Get Your Own Copy— Subscribe To The *Journal Of AOAC INTERNATIONAL*!

The Scientific Association Dedicated to Analytical Excellence®

AOAC INTERNATIONAL®

	North America (USA, Canada, Mexico)			Outside North America		
	*Member 1998	Nonmember 1998	Institutional 1998	*Member 1998	Nonmember 1998	Institutional 1998
1 year	☐ $181	☐ $249	☐ $270	☐ $232	☐ $300	☐ $320
2 years	☐ $331	☐ $452	☐ $490	☐ $432	☐ $554	☐ $591
Single Copies	☐ $ 58	☐ $ 58	☐ $ 58	☐ $ 66	☐ $ 66	☐ $ 66
Courier delivery outside U.S.				☐ $233	☐ $233	☐ $233

AOAC-J, c/o Signet Lockbox, P.O. Box 75198, Baltimore, MD 21275-5198

Please enter my subscription to the ***Journal of AOAC INTERNATIONAL*** as checked above. I have enclosed payment of $ _____ (Subscriptions are on a calendar basis, 6 issues: Jan., Mar., May, July, Sept., Nov.)

☐ Check Enclosed (U.S. funds drawn on U.S. banks only) ☐ Charge my ☐ VISA ☐ MasterCard ☐ Diners Club ☐ American Express ☐ JCB ☐ Eurocard

Card Number _____ Exp. _____ Signature _____ Date _____

Send to: Name _____

Employer _____

Street Address _____

City _____ State(Province) _____ Postal Code _____

Country _____

*Member Number _____ (Must be provided to get member price.)

Table 2. The advantages and disadvantages of the "error budget" uncertainty approach based upon the presentation by Ellison et al. (12)

Advantages	Disadvantages
Permits interpretation of reliability of a measurement	Different methods of estimating uncertainty lead to different values
Requires correction of known systematic errors	If an estimate of systematic error is not available, one is postulated on the basis of experience
Allows for uncertainty of correction factors	
Is characteristic of chemical measurement process	Uncertainty may be too large for the purpose
Does not require reference to a true value	Approach does not provide an estimate of systematic error to permit correction
Quantitates all contributing components as standard deviations	Local constants are expanded unnecessarily to distributions
Treats estimates from experiment, prior knowledge, and professional judgment equally	Quality of estimate is not considered; approach is not entirely objective
Is quicker and less costly than interlaboratory method performance study	It is impossible to predict all environmental factors that might be introduced by different laboratories
Accounts for all contributing factors	It is impossible to anticipate all potential sources of uncertainty
Provides for comparability in uncertainty estimates	Addition or omission of factors by different estimators can lead to discrepancies in uncertainty estimates
Does not exclude use of third-party review, quality assessments, interlaboratory comparisons, in-house control materials, and certified reference materials	Local estimates are not generally applicable to other laboratories
Permits method improvements from knowledge of uncertainties	"Fit-for-purpose" criterion limits need for improvement
Provides all necessary data through routine quality control procedures	Quality control provides only final results, not components
Permits incorporation of uncertainty from sampling	Uncertainty from sampling is a local constant, characteristic of individual lots
Considers adventitious sources of uncertainty	Adventitious sources are strictly a local constant, not a constituent of the measurement process
Takes advantage of certified reference materials to provide uncertainty information in a consistent manner	
	Calculated uncertainty at ultratrace levels often very large in relation to its limit.

summarizes the systematic errors found in the laboratories sampled. If a reference material is available so much the better, but such an opportunity is not necessary and is unavailable in the case of ill-defined analytes. The chief objection is the cost of organizing and conducting the interlaboratory method-performance study, but there is no alternative when allowance must be made for differences among laboratories. The example used by Ellison et al. (12), that later studies of the AOAC method for the determination of cholesterol in fats and oils showed that the recovery is less than the reproducibility figure suggests, is inappropriate. The AOAC method was developed to determine adulteration of vegetable oils by animal fats and not to determine cholesterol content of foods as interpreted by the citation. The criticism is also inappropriate in that recovery measures systematic error while reproducibility measures the random component of uncertainty, which the authors are careful to differentiate.

There is a substantial difference between physical and chemical measurements regarding reliance on individual or multiple laboratories. Metrologists will spend a considerable amount of time on a single measurement to ensure the reliability of their physical standards and the validity of their corrections. They are not very interested in the equivalent values obtained by other laboratories. The characteristics and definitions of physical properties suggest how to measure these features so there is no need for standard methods. The magnitude of physical properties may be known to 8–10 significant figures. The magnitude of most concentration measurements obtained in analytical chemistry is typically 2, and at best 4, significant figures.

The nature of errors is also quite different in the 2 sciences. In physical sciences most experimental errors are systematic

and correctable, such as temperature and pressure effects. Random errors are relatively minor, arising from reading a scale and sometimes from lack of sufficient resolution of the measuring instrument. Metrological measurements are considered very important, and considerable time is spent on each. The trend in analytical chemistry, at least in the food and drug sciences, is toward very rapid, broad-spectrum, nonspecific methods such as near-infrared spectroscopy. There is no need for high accuracy in many of these determinations, because often, each analyte is composed of numerous congeners, each with different biological activity in the very broad sense of the word, whereas the desired effect is the overall biological response, independent of analyte specificity. Furthermore, the quantity of food consumed, which is rarely measured, governs nutritional status more than the composition of the food. Even in safety evaluation, overall effect is desired, such as expressing the different toxicities of dioxin congeners as "toxicity equivalent" to 2,3,7,8-tetrachlorodibenzodioxin. In these cases, identification of analytes is important in order to assign the correct toxicity factor. All of this measurement activity is occurring at levels thought to be impossible to achieve as recently as 2 or 3 decades ago. Many necessary transformations are buried in computer algorithms inaccessible to analysts. Under such circumstances, extracting an uncertainty factor is almost impossible. Even if a semireliable factor could be estimated, it would undoubtedly be so large as to overlap zero. When this unreliability factor appears, mathematical conclusions become irreconcilable with the physical evidence of chromatographic peaks and mass spectral ion responses at predicted mass numbers.

Despite doubts cast on the reality of some of our measurements, a best estimate of analyte concentrations, even below the limit of quantitation, is required to provide an indication of population exposure to ubiquitous toxicants, such as dioxins and lead, or to essential nutrients verging on deficiency, such as zinc or folic acid. The laws of reversion to the mean and of large numbers appear to work in the direction of providing reasonable estimates despite high uncertainties.

Within-Laboratory Validation

In the modern instrumental laboratory, control of uncertainties is based on operations of quality control. Much of modern analytical chemistry of organic matrixes is performed by liquid chromatographic (LC) separations with high resolution obtained by use of high pressure or small-diameter capillary columns. Validation of these systems, discussed in detail by Furman et al. (17), consists of demonstrating that the system is suitable for the intended use, which corresponds to showing that the uncertainty is small enough to give meaningful results and conclusions. Furman et al. used 2 approaches, the first, called modular validation, examines each component of the system not in terms of uncertainty components but its equivalent in terms of calibration. The chromatographic system consists of an absorbent column whose properties have been selected to perform the desired separations, maintained at constant temperature. The analytes in solution, extracted from the original matrix, if necessary, are injected onto the column. A solvent is passed through the column to move analytes along under such conditions of temperature, pressure, and column and solvent polarities that achieve the desired separation. Separated analytes flow through a cell where a detector typically determines absorbance or fluorescence of the solution at a selected wavelength in the visible or ultraviolet region. Changes in absorbance (or other property) initiated by the presence of the analyte is transformed from an analog signal to a digital output for computer calculation of the analyte concentration in the original product. Numerous modifications of the basic system are required for various analytes and matrixes, but the basic procedure is the same.

Some of the more important problems encountered in calculating uncertainties of each component of this elaborate modern analytical system are as follows:

(**a**) *Injection volume.*—A definite volume is manually or automatically injected onto the column. This volume, in microliters, is determined by weighing the delivered volume of a reference liquid. Uncertainties arise from the difference in surface properties of the calibration liquid and the analyte solution and from deterioration of injector parts.

(**b**) *Pump flow rate.*—The volume, pressure, and flow rate of the mobile phase, as well as the polarity of the column, govern the resolution of components as they flow through the column. An attempt is made to control the pump characteristics, an exercise in hydraulic engineering made difficult by deterioration of moving parts and deposition of salts on critical parts.

(**c**) *Detector.*—The separated analyte emerges from the column and flows through the cell of the detector for measurement of absorbance at a definite wavelength or some other property such as fluorescence or an electrochemical response. The wavelength or property must be calibrated, the linearity of the detector response must be verified, and stray light or other extraneous effects must be minimized. Often, the responses of the calibrating system and the test system must be assumed to be identical.

(**d**) *Analog-to-digital converter.*—It is extremely difficult to design electrical simulation systems to check the conversion of the input signal representing the absorbance or other property to the output signal used to calculate the final value.

(**e**) *Data processor and program code.*—The computer program must be checked to ensure that it gives the correct calculated results from peak heights or peak areas to concentrations. Many systems use integrators and workstations with proprietary codes and programs that will not be made public by the manufacturer. The best that can be hoped for is that the manufacturer has tested the program and found it to be accurate.

Furman et al. conclude that because most chemists will not be able to validate proprietary computer programs or perform an individual evaluation of each of the components of a computerized system, such an evaluation cannot guarantee reliable analytical results. Therefore, they propose a holistic approach that evaluates overall performance of the entire system under working conditions. This is equivalent to a within-laboratory method-performance study.

In the holistic approach, the entire system is first tested for linearity, with 4 standard solutions but not 0, over the expected range. This will be used as the primary quality control tool. The

linearity check is supplemented by a daily replicability check of 6 injections of the same standard solution. A repeatability of 1% for the system and a single operator is the initial target value. Drift is controlled by analyzing a standard solution with every set of 5 test solutions or every 2 h. If the performance is satisfactory, test samples are run and calculated for compliance with specifications either manually or automatically by the controlling computer. These LC systems have so many adjustable parameters that different, fully qualified operators are unlikely to make the same choices. These include attenuation, noise rejection, sensitivity of the standard curve, construction of the baseline, data-smoothing parameters, noise rejection levels, points to start and stop peak integration, and choice of signal measurement (height or area). Improvements in columns, supports, detectors, electronics, and data-handling algorithms occur with such frequency that at some future time it is impossible to reproduce the measurement conditions of several months previously.

Other Complex Systems

Complexity is not confined to LC systems. Some other examples of modern analyses that are performed automatically include gas chromatography for pesticide residues, inductively coupled plasma atomic emission spectroscopy for trace elements, and various separation techniques coupled with mass spectrometry for industrial pollutants. It is inconceivable that these techniques could be monitored through error budgets. The only practical way of monitoring their uncertainties is through an overall quality control approach using calibrating solutions, house standards, and certified reference materials. Furman et al. (17) have arrived independently at the conclusion that the "top-down" procedure (holistic) approach is the only way to handle modern analyses. The error budget approach (modular) cannot be used to handle modern analytical problems.

Ruggedness Approach

Youden (18), with his very practical approach to statistics, provided a within-laboratory technique for estimating uncertainty of an analytical method within a single laboratory. This model was formulated initially to replace the multiple experiments required to obtain the effects of a number of different variables changed one at a time. By proper placement of high and low values for each of the potential variables, performance of a single design could indicate which of the factors must be carefully controlled. The high and low values are chosen to reflect the extreme effects that would be expected during the actual conduct of the method. If there are no outstanding differences, the differences can be used to calculate an expected standard deviation. If a certified reference material is available, an estimate of systematic error is also available from the mean value.

Although this model has not been used to calculate uncertainty, it has proved effective for isolating significant variables during the development phase of new methods of analysis and for optimizing existing methods in new applications prior to the performance of an interlaboratory method performance study.

Neglected Alternative Recommendations

The EURACHEM guide "Quantitating Uncertainty in Analytical Measurement" (9) lists a number of alternatives to calculating uncertainty by the error budget process. These are grouped in a section entitled "Experimental Quantification," 5.4.6–5.4.24. They include:

5.4.6—Perform repeatability experiments to provide the standard deviation of random effects from about 15 replicates. These replicates should be independent (not consecutive).

5.4.7—Vary known parameters to provide an uncertainty by statistical means. This is the Youden ruggedness procedure, although not so designated.

5.4.9—Use reference materials to provide the combined effect of many sources of uncertainty.

5.4.12—Use tolerances supplied by certificates or manufacturer's literature.

5.4.13—Use reproducibility data from interlaboratory method-performance studies, although the authors do not believe that they are sufficiently inclusive.

5.4.14—Use data from proficiency-testing schemes but only if the assigned value is traceable and the uncertainty on the assigned value is small compared with the spread of the results. These restrictions unnecessarily remove many methods of product control and food and nutrition analyses that must utilize "empirical" methods. These methods have no systematic error by definition.

5.4.15—Use quality assurance data as a check on previously determined uncertainty values.

5.4.16—Finally, estimation based on judgment is permitted when all else fails—when repeated measurements cannot be performed as in extrapolation to other matrixes, a mismatch of reference materials, empirical methods, and spiking with a single substance applying to the entire class of related compounds. In such cases "a degree of belief" is combined with other components of uncertainty.

It is unfortunate that the error budget technique has been so emphasized that legitimate application of other procedures has been overlooked. The interlaboratory study has been unnecessarily deprecated on the grounds that not all sources, particularly of systematic error, are covered. The statement overlooks the requirement that such studies be designed to apply to or to bracket those specifications that make the method "fit for the purpose." Therefore, the study does not expand uncertainty estimation unnecessarily to cover situations not encountered in practice.

In view of the practical difficulties of transferring uncertainty estimation as practiced in physics and metrology to modern analytical chemistry, we should look for more practical alternatives. One has already been mentioned—internal within-laboratory quality control. This is the simplest procedure that works well for production line analyses for control of analytes expected to be present in a reasonably constant amount, as is the case for pharmaceutical products, well-characterized agricultural products and adjuncts (fertilizers, feeds, pesticide formulations, standardized foods), and clinical chemistry. It would not, however, work well when products vary considerably in composition and identity.

The ISO error budget approach to uncertainty calculation is analogous to the classical approach to experimental design, in which all factors are kept constant except the one under consideration and each factor is examined in turn in this manner. This approach was still favored in physics but was unsatisfactory in an environment where changing one factor also changed other factors, as is the case in many less structured branches of science. Modern analytical techniques no longer can be analyzed by simplistic approaches. The full play of interactions should be permitted, often in the direction of canceling opposing effects. Viewing results from a higher metrological level often permits transformation of hidden systematic errors into random errors that converge to zero with increasing replication.

The error budget approach also has some other minor defects. It is inapplicable to the single nonroutine test that comes into all laboratories and must be handled. The matter of disposition of outliers has not been handled. Indirect evidence from our reviews suggest that as much as 10% of the data reported by analytical chemists may be suspect.

Conclusions

The concept of uncertainty has been developed to inform our customers how much allowance must be made for the possibility that repetition of the test will give a different value. It is the range within which we expect future values obtained from the same test will lie with a high probability. Most customers, even other scientists, are more or less indifferent to how the estimate of uncertainty is derived. All are concerned with its reliability—that other operators using different equipment in different laboratories at different times will supply results within that interval of uncertainty. The pioneers from the national metrology laboratories followed the path of classical science of changing one variable at a time while holding all others constant and noting the effect. Then all the effects are added to arrive at the total effect. This procedure served well as long as the variables could be isolated and did not interact too much with each other. These conditions were met in physics, engineering, and metrology and in measurement of simple physical measurements elsewhere. As soon as the measurement technique became complicated, however, it was no longer possible to isolate individual effects, but rather groups of effects such as dissolution of the test portion, isolation of the analyte, and measurement of the isolated analyte. With automated, computer-controlled equipment, it was no longer possible to isolate any step, and the overall process had to be evaluated through calibration, quality control, and validation.

Laboratory assessors unfortunately have focused on the simple error budget model as the only means of demonstrating uncertainty, despite its inappropriateness in the complex environment of modern analytical chemistry. A number of more appropriate alternative designs already exist that provide more effective evaluation of errors of chemical measurements.

References

(1) International Organization for Standardization (1995) "Guide to the Expression of Uncertainty in Measurement," ISO, Case Postale 56, CH1211 Geneva, Switzerland

(2) Taylor, B.N., & Kuyatt, C.E. (1994) "Guidelines for Evaluating and Expressing the Uncertainties of NIST Measurement Results," NIST Technical Note 1297

(3) U.S. Food and Drug Administration (April 12, 1994) *Fed. Reg.* **59**, 17384–17389

(4) U.S. Food and Drug Administration (Dec. 31, 1987) *Fed. Reg.* **52**, 49589; "General Principles for Evaluating the Safety of Compounds Used in Food-Producing Animals," Center for Veterinary Medicine (Revised July 1994)

(5) Horwitz, W., Albert, R., & Nesheim, S. (1993) *J. AOAC Int.* **76**, 461–491

(6) Horwitz, W., & Albert, R. (1996) *J. AOAC Int.* **79**, 589–621

(7) Thompson, M., & Howarth, R.J. (1980) *Analyst* **105**, 1188–1195

(8) Hirschfeld, T. (1976) *Anal. Chem.* **48**, 16A–31A

(9) "Quantitating Uncertainty in Analytical Measurement" (1995) Available from Laboratory of the Government Chemist, Queens Rd, Teddington, Middlesex TW11 0LY, United Kingdom

(10) Pueyo, M., Obiols, J., & Vilalta, E. (1996) *Anal. Commun.* **33**, 205–208

(11) Ku, H.H. (Ed.) (1969) *Precision Measurement and Calibration*, NBS Special Publication 300 US Government Printing Office, Washington, DC, Vol. 1, p. 336

(12) Ellison, S., Wegscheider, W., & Williams, A. (1997) *Anal. Chem.* 607A–613A

(13) Analytical Methods Committee (1995) *Analyst* **120**, 2303–2308

(14) Nordic Committee on Food Analysis (NMKL) (1997) "Estimation and Expression of Measurement Uncertainty in Chemical Analysis," NMKL Procedure No. 5 (1997) General Secretariat of NMKL, c/o VTT Biotechnology and Food Research, PB 1500, FIN-02044 VTT, Finland

(15) Natrella, M.G. (1993) in *NBS Handbook 91*, U.S. Government Printing Office, Washington, DC, pp 23–31

(16) Youden, W.J. in Ref. (11), 63–133

(17) Furman, W.B., Layloff, T.P., & Tetzlaff, R.F. (1994) *J. AOAC Int.* **77**, 1314–1318

(18) Youden, W.J., & Steiner, E.H. (1975) *Statistical Manual of the Association of Official Analytical Chemists*, AOAC, Arlington, VA, pp 33–36