

P. MEASUREMENT UNCERTAINTY AND NIOSH METHOD ACCURACY RANGE
by David L. Bartley, Stanley A. Shulman, and Paul C. Schlecht, NIOSH/DART

CONTENTS	Page
1. Introduction	208
2. ISO GUM	209
3. The Symmetric Accuracy Range A as Used by NIOSH	212
a. Definition and Approximation	212
b. Uses of the Symmetric Accuracy Range	213
4. Uncertainty and Analytical Lab Procedures	215
a. Validated Method Adoption	215
b. Pooled Quality Control Results	218
c. Continual Method Re-Evaluation	218
d. Limit of Detection and Detection Limit	218
5. Discussion	217
6. Technical Notes	219
Note 1: Example of accuracy range confidence limit	220
Note 2: Single-evaluation correction of bias	221
Note 3: Characterizing effects of uncorrected bias	224
7. References	225
8. Terminology	227

1. INTRODUCTION

Recently, the *ISO Guide to the Expression of Uncertainty in Measurement* (GUM) has come close to being universally adopted as the standardized way to characterize and document measurement uncertainty [1-5]. Since the mid-1970s, accuracy criteria have been an integral part of the evaluations of the sampling and analytical methods used by the National Institute for Occupational Safety and Health (NIOSH), the Occupational Safety and Health Administration (OSHA), the Mine Safety and Health Administration (MSHA), and others. NIOSH has previously published extensive discussions addressing the issue of accuracy as a factor in the development, evaluation, and characterization of analytical methodology. Both traditional method accuracy and new measurement uncertainty concepts are intended to communicate measurement limitations to laboratory clients. Naturally, laboratories are interested in how NIOSH accuracy requirements [6-7] relate to measurement uncertainty.

This chapter provides guidance for achieving consistency in determining measurement uncertainty by those laboratories using NIOSH methods. Minor modifications to NIOSH accuracy measures, and an expansion of ISO GUM to cover situations unique to workplace atmosphere measurement can improve consistency and utility.

ISO GUM proposes pooling estimated variance components from diverse error sources. The square root of the pooled variance estimate is termed the *combined uncertainty* u_c . Multiplication of u_c by a *coverage factor* k (generally in the range of 2 to 3) results in an *expanded uncertainty* U . The purpose of the expanded uncertainty is for each measurement to provide an interval bracketing the *measurand* (the true value of what is to be measured)

to account for errors in both the measurement and the determination of the uncertainty components themselves.

ISO GUM is somewhat sketchy about the coverage factor k . Furthermore, the coverage factor can be interpreted in several ways. Most straightforward is the limited case where the uncertainty components can be re-evaluated each time the method is used (resulting in k proportional to a Student-t quantile). In this case, the covering intervals bracket the measurand for (for example) 95% of the measurements.

Alternatively, the coverage factors based on the Student-t quantile specify intervals containing measurand values at levels of evaluation confidence in the *mean* (i.e., averaging over many method evaluations). In other words, for roughly 50% of method evaluations, intervals used at each measurement contain the measurand value greater than (for example) 95% of the time. The concept is consistent with the statistical theory of tolerance or prediction intervals.

This approach is important to industrial hygiene since workplace air concentrations vary spatially and over time to such a degree that a method cannot be evaluated by simply taking replicate measurements [8]. However, industrial hygiene measurement methods have traditionally required confidence levels greater than 50% in the method evaluation. Generally, 95% confidence in a method validation is required. The different types of confidence levels are reflected simply in the numerical value and interpretation of the coverage factor.

Of equal importance in the industrial hygiene field are details needed to handle systematic error (bias) relative to reference concentration measurements found during method evaluation. For example, the sampling rate of a given diffusive sampler for gases or vapors is generally measured once by the diffusive sampler manufacturer prior to use by multiple clients. As the samplers are not re-calibrated for each use, residual bias exists in the measurements due to uncertainty in sampling rates used [9]. (NIOSH methods typically do not cite performance for passive samplers because agreement among diffusive monitor manufacturers on test protocols has not yet been achieved, and a system of third party evaluation of diffusive monitor manufacturers sampling rates is not available.) Similarly, the calculation of desorption efficiencies may be performed only once or infrequently and can, therefore, introduce residual bias in measurements that use sorbent-captured samples, e.g., charcoal tubes.

In aerosol sampling, detailed knowledge of the particle size-dependent bias of a sampler relative to a sampling convention, such as adopted by ISO/CEN/ACGIH/ASTM [10] for defining respirable dust, is often necessary to judge the usefulness of a given sampler. Each

type of aerosol sampler is characterized by specific particle collection characteristics, and some analytical methods (e.g. silica) may also exhibit particle size effects. Typically the issue of aerosol sampler bias is avoided or minimized in the industrial hygiene field by narrowing use to a specific aerosol sampler. For example, common industrial hygiene practice establishes a single sampler type, such as the 1.7 L/min 10-mm nylon Dorr-Oliver cyclone, for respirable dust sampling in a particular application.

Sensitivity to other environmental factors, referred to in ISO GUM as *influence variables*, must be acknowledged. Suppose a sampler is sensitive to temperature changes that are impractical to measure in the field; i.e., sampler estimates are not temperature corrected. Then, suppose during method evaluation in the laboratory, measurement of this sensitivity is combined with knowledge of the expected temperature variation for a given field application. Putting together both would determine the uncertainty associated with the effect. Examples of the important effects of influence variables - such as wind velocity, temperature, pressure, and fluctuating workplace concentrations - on diffusive monitor uptake rates are common.

2. ISO GUM

ISO GUM presents several concepts. One of these calls for the identification of sources (labeled $j = 1, 2, \dots$) of uncertainty u_j (standard deviation estimate components) in a measurement method and for their classification into Type A or Type B uncertainties. Type A uncertainty is one that has been characterized by a statistically sound approach. In this case, u_j^2 is given by s_j^2 , an unbiased estimate (with ν_j degrees of freedom) of variance σ_j^2 . On the contrary, Type B uncertainty generally requires professional judgment. See Table 1 for examples of possible uncertainty components.

A common example of Type B uncertainty is the conservative assignment (Chapter E) of a 5% relative standard deviation component (without error, i.e., with infinite degrees of freedom) as the random sampling pump uncertainty. As described in ISO GUM, such an assignment would be a result of sampling pump random errors that had a uniform distribution and fell within $\pm \sqrt{3} \times 5\%$ of zero with a probability "for all practical purposes equal to one". Therefore, if it is judged that sampling pump variations are *within* these bounds, then the assignment of 5% as the relative standard deviation component is conservative. Other similar ways of handling Type B uncertainties are found in ISO GUM.

TABLE 1. EXAMPLES OF POTENTIAL UNCERTAINTY COMPONENT SOURCES.

Sampling

personal sampling pump flow rate: setting the pump and subsequent drift
sampling rate of diffusive sampler
sampler dimension (aerosol and diffusive sampling)

Sample handling

sample preparation (e.g., handling silica quasi-suspensions)
sample loss during transport or storage

Analytical

aerosol weighing
recovery (e.g, GC-based methods)
Poisson counting (e.g., in XRD methods)
sensor variation
operator effects giving inter-lab differences (if data from several labs are to be used)

Evaluation

calibration material uncertainty
evaluation chamber concentration uncertainty
other bias-correction uncertainty

Environmental influence parameters

temperature (inadequacy of correction, if correction is made as with diffusive samplers)
atmospheric pressure
humidity
aerosol size distribution (if not measured by a given aerosol sampling method)
ambient wind velocity
sampled concentration magnitude itself (e.g., sorbent loading)

Within the field of industrial hygiene, the quantities u_i are often standard deviation component estimates obtained from a *single* measurement-method evaluation, rather than from replicates. When the estimates are independent, a *combined uncertainty* u_c may be computed (through the propagation of uncertainty approximation) as:

$$u_c = \sqrt{u_1^2 + u_2^2 + \dots} \quad (1)$$

Through a *coverage factor* k , generally approximated conservatively (e.g., see *Technical Note 2* at end of chapter), as equal to 3 for a single method evaluation, an *expanded uncertainty* U may be computed as:

$$U = k \times u_c \quad (2)$$

The purpose of the expanded uncertainty U is to provide intervals, which generally contain measurand values (often referred to as the *true* values). In particular, given a concentration estimate \hat{c} (hats, as here, indicate estimates), the measurand value C is bracketed at better than 95% confidence by intervals of the type:

$$\hat{c} - U < C < \hat{c} + U, \quad (3)$$

at 95% confidence in the method evaluation. The coverage factor k is intended to account for both (1) the fluctuation of the measurement about the measurand value and (2) the uncertainty in the assessment of this fluctuation.

Note: Requiring only *mean* confidence in the evaluation leads to k given in terms of a *Student-t* quantile. Here, however, in fixing the method evaluation confidence (e.g., at 95%), the *chi-square distribution* takes the place of the Student-t distribution. Double confidence levels (in the measurement and evaluation) directly relate to a well-developed [11-19] statistical theory of *tolerance* or *prediction intervals*.

Another point of ISO GUM is semantic. *Uncertainty*, as in common usage, covers only what is *unknown* about a measurement. The *known* but uncorrected systematic deviation or *bias* relative to reference concentrations does not enter into measurement uncertainty.

A related concept, *accuracy*, is defined qualitatively within ISO GUM as the “closeness of agreement between the result of a measurement and a true value of the measurand”. Accuracy can have both random and systematic components. It is not surprising then that if a bias correction is made and if accuracy is quantified reasonably, the expanded uncertainty and an accuracy confidence limit can be equivalent.

As mentioned above, another aspect of ISO GUM deals with *influence factors*. If measurement results are expected to be sensitive to an environmental factor (e.g., ambient temperature), then the effect of such a factor on the measurement method must be measured in the laboratory. Given estimates of the environmental variations expected during method application, influence components of the combined uncertainty can be estimated for inclusion in the uncertainty budget of Eq. 1. Table 1 lists several *influence factors*, which may or may not be significant.

3. THE SYMMETRIC ACCURACY RANGE A AS USED BY NIOSH

a. Definition and its Approximation

The *symmetric accuracy range* A is defined as the fractional range, symmetric about the true concentration C , within which 95% of sampler measurements \hat{c} are to be found. Another way of saying this is:

$$C \times (1 - A) < \hat{c} < C \times (1 + A) \quad \text{for 95\% of measurements } \hat{c} \quad (4)$$

It is clear from this simple definition that the accuracy range function A must increase with both random effects and bias magnitude and therefore, is one means of quantifying accuracy as defined above according to ISO GUM.

More specifically, suppose that estimates \hat{c} are normally distributed about population mean c with standard deviation σ . Then we may characterize random measurement effects in terms of the (true) relative standard deviation $TRSD$ and *bias* of the mean concentration estimate c relative to the true concentration C as:

$$\begin{aligned} \text{bias} &= (c - C) / C \\ TRSD &= \sigma / C \end{aligned} \quad (5)$$

The descriptive definition of Eq. 4 implies that the symmetric accuracy range A increases with both $TRSD$ and bias magnitude $|\text{bias}|$. This feature can be seen directly in the following close approximation to the accuracy range function A , which follows [Ref. 11, 12 for derivation] from the definition in Eq. 4:

$$\begin{aligned} A &= 1.960 \times \sqrt{|\text{bias}|^2 + TRSD^2}, \text{ if } |\text{bias}| < TRSD / 1.645; \\ A &= |\text{bias}| + 1.645 \times TRSD, \text{ otherwise.} \end{aligned} \quad (6)$$

This expression is simple enough for calculation by most hand-held calculators, and it is also a useful starting point for estimating the 95% confidence limit $A_{95\%}$ on the accuracy range as measured during a method evaluation, accounting for evaluation errors.

b. Uses of the Symmetric Accuracy Range

method validation. One application of the symmetric accuracy range is for evaluating measurement methods. As mentioned in Chapter E, a method evaluation consists of a number of measurements taken from replicate samplers at each of several controlled and known concentrations covering the range of expected method application. This type of experiment gives information about the samplers' random errors and also the bias relative to reference concentrations. A confidence limit on the accuracy range can then be computed. One objective in a method suitable for NIOSH application is that the 95% confidence limit $A_{95\%}$ not exceed 25%. A includes both the uncertainty (as the term is used by ISO GUM) and the systematic deviation or bias, so that correction of the bias by the sampler vendor or developer is encouraged by the very statement of this objective. See Eqs. 9-11 below for computing $A_{95\%}$ when bias is negligible.

measurement uncertainty. Suppose then that bias correction has been made. For example, suppose that following evaluation, the sampler is used for future measurement with bias corrected on the basis of its measurement during the evaluation itself. Then computation of the confidence limit $A_{95\%}$ is possible accounting for the residual bias which is uncorrectable due to evaluation limitations, but nevertheless will be present in all future measurements. The quantity $\hat{c} \times A_{95\%}$ in this case forms the counterpart to the expanded uncertainty U of ISO GUM for specifying evaluation confidence at 95%.

The relationship between $\hat{c} \times A_{95\%}$ (with corrected bias) and expanded uncertainty U can be seen most clearly in the case that $A_{95\%}$ is significantly smaller than 100%. In this case, Eq. 4 can be rewritten as the approximation:

$$\hat{c} - \hat{c} \times A_{95\%} < C < \hat{c} + \hat{c} \times A_{95\%}, \quad (7)$$

which means that at 95% confidence in the method evaluation, the inequality bracketing the measurand value C holds at probability $> 95\%$. A study using 10,000-point simulations indicates that $A_{95\%}$ can be as large as 25%, with method evaluation confidence close to 95% using the approximation of Eq. 7. As can be seen directly, Eq. 7 is the analogue to Eq. 3 when $\hat{c} \times A_{95\%}$ is adopted as the *expanded uncertainty* U :

$$U = \hat{c} \times A_{95\%} \quad (8)$$

In the case that bias is known to equal zero (*Technical Note 1* at the end of this chapter), $A_{95\%}$ (at 15 degrees of freedom in the evaluation experiment) is simply:

$$A_{95\%} = 2.8 \cdot TRSD \quad (9)$$

Eq. 8 then gives:

$$\begin{aligned} U &= 2.8 \cdot TRSD \cdot \hat{c} \\ &= 2.8 \cdot u_c \end{aligned} \quad (10)$$

Therefore, the coverage factor k is

$$k = 2.8, \quad (11)$$

consistent with the use of $k = 3$ as a conservative but not excessive value.

The user of a method then may report the expanded uncertainty U in a concentration estimate \hat{c} using Eq. 8, knowing the accuracy range confidence limit $A_{95\%}$ as reported in the method. Of course, this approach relies on the sense of double confidence—in the evaluation and also in the subsequent application.

Note: The European Assessment of Workplace Exposures Technical Committee, CEN TC 137, has adopted a similar viewpoint regarding method performance [20]. In this case, an *overall uncertainty*, defined as $|bias| + 2 \times RSD$, is used to quantify accuracy. When compared to Eq. 6, the overall uncertainty can be regarded as an approximation to the symmetric accuracy range.

4. UNCERTAINTY AND ANALYTICAL LAB PROCEDURES

Interest in measurement uncertainty and ISO GUM is currently finding its way into the criteria for the accreditation of analytical labs [4]. The result will no doubt be high confidence in understanding one component of the combined or expanded uncertainty—namely the analytical component. Several general approaches to controlling and characterizing analytical uncertainty in routine lab practices seem reasonable.

a. Validated Method Adoption

One possibility is for a lab to adopt a published, evaluated method. Such an adoption would require an initial establishment of the method within the lab's capabilities. Equivalence to the published method would be established during this initial phase. Thereafter, the method's uncertainty as documented in the original publication would be claimed for the lab results. Ongoing analysis of a limited number of quality control samples would provide evidence that the method as implemented in the lab remains stable.

An example of this approach is the current practice in some labs that handle sorbent tubes to analyze about 4 lab blanks per set of field samples analyzed. The variability in the blank results are then continually compared to past lab performance so as to detect problems which may occur in analysis. Though the small number of degrees of freedom ($= 3$) does not give a tight figure on the uncertainty, it nevertheless gives assurance that the method is stable.

As a specific example of method evaluation data and documentation of an uncertainty budget, data from $n = 16$ exposures of diffusive samplers in a controlled environment are shown in Table 2. The evaluation is somewhat simplified for this example; a more comprehensive evaluation would also measure effects of wind velocity, humidity, temperature, and concentration time-dependence (potentially significant to diffusive monitoring). Analysis of these data can be handled by an ordinary calculator capable of computing means and standard deviations.

Note that the uncertainty (tR_{SD} / \sqrt{n} , where $n = 16$ is the number of measurements) in the bias is the value that accounts for residual bias due to imperfect correction. Very similarly, the uncertainty in the reference concentration is pooled to arrive at a combined uncertainty. Interestingly, neither of these two contributions corresponds to quantities that vary during sampler application subsequent to its initial evaluation. The background for documenting residual (uncorrectable) bias can be seen in *Note 2* at the end of this chapter.

Table 2. Example of Method Evaluation and Uncertainty Budget

Evaluation experimental results

The following are results from a simplified evaluation of a specific diffusive sampler for o-xylene. There were four experimental runs with four samplers each. The reference concentration set within the exposure chamber is denoted as \hat{C} having an assigned (Type B) relative uncertainty = 1%.

Run	\hat{C} (ppm)	replicates (ppm)	$\Rightarrow TRSD_{inter} [s / \hat{C}]$	biâs
Run 1	\hat{C} (ppm) = 123	139.2 138.2 138.6 145	2.6%	14.0%
Run 2	\hat{C} (ppm) = 101.1	108.3 110 110.7 112.8	1.8%	9.2%
Run 3	\hat{C} (ppm) = 12.7	14.2 15.3 12.9 14.6	7.9%	12.2%
Run 4	\hat{C} (ppm) = 91.3	109 109.2 107.1 105	2.1%	17.8%

Averaging the above bias estimates and pooling the inter-sampler estimates $TRSD_{inter}^2$ results in:
 Mean bias estimate: $biâs = 13.3\%$ from average of $4 \times 4 = 16$ data points.

The TRSD estimate is $TRSD_{inter} = 4.4\%$ having $4 \times 3 = 12$ degrees of freedom.

Uncertainty budget

The following includes bias correction by dividing future concentration estimates by $(1 + \text{bi}\hat{a}s)$ as in *Technical Note 2*.

<i>Source</i>	<i>Component</i>	<i>Category</i>
Inter-sampler	3.9% [= 4.4% / (1+.133)]	Type A
Bias Correction Uncertainty	0.97% [= 3.9% / $\sqrt{16}$]	Type A
Ref Concentration Uncertainty	0.5% [1% / $\sqrt{4}$, but not / (1 + <i>bi</i> \hat{a} s)]	Type B

Combined (Relative) Uncertainty $u_c = \sqrt{3.9\%^2 + 0.97\%^2 + 0.5\%^2} = 4.0\%$

Expanded (Relative) Uncertainty $U = k \times u_c = 12.1\%$ ($k = 3$).

Notes: Here $k = 3.0$. A more accurate determination based on the chi-square quantile at 12 degrees of freedom and $\text{prob} = 0.05$, gives $k = 2.97$, which is consistent with conventional use of 3 as a conservative value.

Again, an expanded (relative) uncertainty U means that with greater than 95% of future bias-corrected estimates \hat{c} , true concentrations C are bracketed by:

$$\hat{c} \times (1 - U) < C < \hat{c} \times (1 + U)$$

at 95% confidence in the above evaluation experiment. Generally, a quality control program is required to ensure that the method remains stable following evaluation.

Note that many methods (e.g., those based on sorbent tubes) employ personal sampling pumps, in which case normally a 5% (Type B) component representing sampling pump uncertainty would be included in the uncertainty budget.

Note also that the inter-sampler component includes both analytical and sampling sub-components. Further refinement of the inter-sampler component may perhaps be useful for improving a method, but is not needed for establishing confidence intervals around (true) measurand values.

Note further that storage effects require estimating and inclusion in the budget if considered significant.

b. Pooled Quality Control Results

Another approach utilizes a large number (e.g., 50) of the most recent quality control sample results. By pooling uncertainty values, a running method evaluation can be effected. The result is (1) a direct measure of the analytical uncertainty of the method as implemented in the lab, and (2) a means of detecting any problems that may creep into a method during routine use. Note that a running uncertainty average is similar to a partial method evaluation and not to a method re-evaluation at each measurement because consecutive running averages are strongly correlated.

This approach is adopted within a current MSHA procedure for the analysis of silica. A sampling filter is dissolved and re-deposited onto an analysis filter where the silica is quantified by infrared absorption. From each batch of samples to be analyzed, an analysis filter is retained for re-dissolution, re-deposition, and re-analysis within a subsequent batch. The result is a large number of pairs of nearly identical samples, which can give a running estimate of the method's analytical uncertainty.

c. Continual Method Re-Evaluation

A third possibility, the closest to the original measurement approach of GUM uses a large number (e.g., 30) of *independent* control samples for each application measurement. This is the most expensive approach, but also may give the best estimate of the analytical uncertainty, especially in cases where uncertainties may be measurement-dependent. Because many more evaluative measurements per application measurement are needed, this approach is not easily implemented for most industrial hygiene applications.

As an example of this approach [2], suppose that a lab estimates only a 30-day average concentration of a given gas or vapor. Further, every day a measurement is taken of a known calibration gas concentration. Then, if the method is expected to behave similarly for measurements of gas and field samples calibrations, the 30 control samples give analytical uncertainty estimates that differ month-to-month and from field measurement to field measurement.

d. Limit of Detection and Detection Limit

When the concentration is low, approaching the method uncertainty u_c , concepts of the limit of detection (LOD) and a related detection limit L_D may be useful. LOD is used for controlling false positives when asserting the presence of a substance. On the other hand, the detection limit specifies what measurand value (e.g., concentration) is required so that the false negative rate is negligible when the substance is actually present. The limits can refer to the analytical measurement only or, as in this section, to the entire sampling and analytical measurement method.

In the following examples, several often realistic assumptions are made. The standard deviation in concentration estimates is assumed constant (i.e., independent of the sampled concentration), unlike the commonly occurring constant *relative* standard deviation at larger concentrations. Also, bias (or uncertainty in its correction) is assumed to be negligible. More

complicated cases generally have specific difficulties that are best be approached by a statistician.

With these assumptions, in terms of the combined uncertainty u_c the limit of detection is traditionally [23, 24] taken to be:

$$LOD \equiv 3 \times u_c \quad , \quad (12)$$

and the detection limit may be defined as

$$L_D \equiv 2 \times LOD \quad . \quad (13)$$

Note that u_c includes the uncertainty associated with correction, if any, with blank subtractions. See Ref. [5] for a detailed example.

After LOD and L_D have been determined for a method they may be used as follows. A substance may be asserted as present if an estimate \hat{x} exceeds LOD . Moreover, if unknown (true) concentration X exceeds L_D , an estimate \hat{x} is likely to exceed LOD . Given the above definitions and assumptions, the false positive rate r on asserting presence is closely equal to the non-detection rate.

Note: If the combined uncertainty u_c is determined from a method evaluation providing an effective number (as in Technical Note 2) degrees of freedom, then at 95% confidence in the method evaluation, the false positive rate r is limited by:

$$r < 1 - \Phi \left[3 \times \sqrt{\chi_{0.05, \nu}^2 / \nu} \right] \quad (14)$$

where Φ is the cumulative normal function. For example, if $\nu = 15$, then $r < 3.5\%$.

5. DISCUSSION

The approach presented here to document method accuracy range and uncertainty relates to the statistical theory [11, 16-22] of *tolerance or prediction intervals*. This theory was originally developed in simplified form to predict the range of future measurements of a normally distributed random variable on the basis of n initial measurements. The *initial measurements* are analogous to the method evaluation, whereas the *future measurements* represent method application subsequent to evaluation.

Because of measurement cost, workplace assessments cannot at present be conducted in such a way that continual re-evaluation is done at each measurement. The prediction interval approach given here shows a less costly way to document measurement uncertainty in those cases where a method has been initially evaluated and then used many times without re-evaluation. The approach closely follows ISO GUM. Of course, for such an approach to actually make sense, an adequate quality control program must be instituted so that the measurement method remains stable during the time of its application following evaluation.

Several generalizations and variations of the material presented in this chapter are possible. The relative standard deviation and relative bias sometimes depend on the concentration sampled in a complicated way, requiring special attention. See, for example Ref. [24]. Also, asymmetric confidence intervals are sometimes required. Single-sided intervals are useful in some instances, e.g., alarm systems, as well as in quantifying limits of detection or quantitation, described briefly above. Ways to handle environmental influence parameters may also be complicated. See, for example Ref. [9]. In any case, despite the complexities possible, the examples given in this chapter may help to characterize method uncertainty in a reasonable manner.

6. TECHNICAL NOTES

Note 1: *Mainly for the statistically minded.*

Example of accuracy range confidence limit:

Suppose it is known that the bias is zero. For example, an exposure standard may be set that specifies a given sampling and analytical method. In this case, the hazardous concentration may be said to be *operationally defined*. Operationally defined methods include NIOSH Methods 7400 and NIOSH 5040.

If the bias is zero, Eq. 6 simplifies to:

$$A = 1.960 \times TRSD .$$

Furthermore, if the relative standard deviation is estimated as \hat{TRSD} with ν degrees of freedom (computed using the Smith-Satterthwaite approximation [1,13-14] if $TRSD$ has more than one component), then the 95% confidence limit on $TRSD$ is:

$$TRSD_{95\%} = \hat{TRSD} \times \sqrt{\nu / \chi_{\nu,0.05}^2} ,$$

where $\chi_{\nu,0.05}^2$ is a 5% quantile value for the chi-square distribution, which can be read from a table in most elementary statistics texts. This determines the 95% confidence limit on the accuracy range itself as:

$$A_{95\%} = \hat{A} \times \sqrt{\nu / \chi_{\nu,0.05}^2} .$$

If $\nu = 15$, $\sqrt{\nu / \chi_{\nu,0.05}^2} = 1.44$, giving coverage factor $k = 2.8$ (Eq. 11).

Note that $TRSD_{95\%}$ can be interpreted as a conservative estimate of $TRSD$ and therefore can be treated as a Type B uncertainty with infinite number of degrees of freedom as described following Table 1.

Note 2: For statisticians only.

Single-evaluation correction of bias:

Details are given here illustrating the tolerance interval approach, bias correction, imprecise reference concentrations, and the use of the symmetric accuracy range function. The derivation is not entirely general, but is given here for guidance in handling the myriad possibilities in measurement uncertainty. Though the derivation is slightly complicated, the result obtained is simple.

Suppose that estimates \hat{c} having an as-yet-unknown constant *bias* relative to true concentrations C (not necessarily constant) may be modeled as:

$$\hat{c} = C \cdot [1 + bias + \hat{\varepsilon}],$$

where the random variable $\hat{\varepsilon}$ is approximately normally distributed about zero with variance $TRSD^2$. For evaluating the method, assume that reference concentration measurements \hat{C} can be made simultaneously and modeled by:

$$\hat{C} = C \cdot [1 + \hat{\varepsilon}_{ref}],$$

where $\hat{\varepsilon}_{ref}$ has variance $TRSD_{ref}^2$, assumed known accurately. Measure n values of the ratio $\hat{\theta}$:

$$\hat{\theta} \equiv \hat{c} / \hat{C} = 1 + bias + \hat{\varepsilon}_c$$

and compute estimates *biâs* and $TRSD_c^2$ at $n - 1$ degrees of freedom, where the approximately normally distributed random variable $\hat{\varepsilon}_c$ has variance $TRSD_c^2$ given by:

$$TRSD_c^2 = TRSD^2 + (1 + biâs)^2 \times TRSD_{ref}^2$$

(to the order of $TRSD$, neglecting Cauchy effects of reciprocals of random variables).

Future bias-corrected measurements \hat{x}' of unknown concentration X can be defined in terms of raw measured values \hat{x} as:

$$\hat{x}' \equiv \hat{x} / (1 + biâs).$$

The residual corrected $bias'$ is then given by:

$$bias' = (bias - bi\hat{a}s) / (1 + bi\hat{a}s).$$

If n is large enough, $|bias'|$ will be small enough that the corrected symmetric accuracy range A' can be accurately approximated (Eq. 6) as:

$$A'^2 = a^2 1.960^2 / (1 + bi\hat{a}s)^2,$$

where the unknown a^2 is:

$$a^2 = (bias - bi\hat{a}s)^2 + TRSD^2,$$

whose confidence limit is now required.

First note that the expected value of the first term is:

$$E[(bias - bi\hat{a}s)^2] = TRSD_c^2 / n.$$

Therefore, an estimate \hat{a}^2 for a^2 can be constructed as:

$$\begin{aligned} \hat{a}^2 &\equiv TR\hat{R}SD_c^2 / n + TR\hat{R}SD^2 \\ TR\hat{R}SD^2 &\equiv TR\hat{R}SD_c^2 - (1 + bi\hat{a}s)^2 \times TRSD_{ref}^2. \end{aligned}$$

Expressed in terms of $TR\hat{R}SD^2$, \hat{a}^2 is

$$\begin{aligned} \hat{a}^2 &= TR\hat{R}SD^2 \\ &+ \frac{1}{n} TR\hat{R}SD^2 \\ &+ \frac{1}{n} \cdot (1 + bi\hat{a}s)^2 \times TRSD_{ref}^2. \end{aligned}$$

Each term can now be identified, forming the basis for an uncertainty budget: the first is the (uncorrected) method uncertainty (squared); the second and third reflect the bias-correction uncertainty owing to finiteness of the validation experiment and the uncertainty in the reference concentration (as here measured n times).

A confidence limit a_β^2 at confidence level β (e.g., 95%) on a^2 is now constructed using \hat{a}^2 :

$$a_\beta^2 = K^2 \cdot \hat{a}^2,$$

where the constant K is to be determined so that

$$\text{prob}\{a_{\beta}^2 < a^2\} = 1 - \beta.$$

First of all, the distribution of \hat{a}^2 is approximated as chi-square:

$$\nu_{\text{eff}} \frac{\hat{a}^2}{E[\hat{a}^2]} \approx \chi^2,$$

where ν_{eff} is determined as with the Smith-Satterthwaite [23-24] approximation, forcing variances to agree; often $\nu_{\text{eff}} \approx n-1$.

Now,

$$a_{\beta}^2 < a^2 \Leftrightarrow \hat{a}^2 < K^{-2} a^2,$$

or, in other words:

$$\chi^2 < \nu_{\text{eff}} (K^{-2} a^2) / E[a^2].$$

[Note that $E[\hat{a}^2] = E[a^2]$, but $\neq a^2$, unlike Note 3.]

Remembering that a^2 depends on the estimate $bi\hat{a}s$, K is given as a solution of the following integral equation:

$$\begin{aligned} 1 - \beta &= \int_{-\infty}^{+\infty} d bi\hat{a}s F[bi\hat{a}s] \int_0^{\nu_{\text{eff}} \cdot K^{-2} \cdot a^2 / E[a^2]} d\chi^2 P_{\nu_{\text{eff}}}[\chi^2] \\ &= \int_0^{\nu_{\text{eff}} \cdot K^{-2}} d\chi^2 P_{\nu_{\text{eff}}}[\chi^2] \times (1 + O[1/n^2]), \end{aligned}$$

where the correction $O[1/n^2]$ is easily proved by expanding the integrand about $TRSD_c^2$ in $(bi\hat{a}s - bias)^2$. Therefore, the following simple asymptotic expression for K^2 results:

$$K^2 \approx \nu_{\text{eff}} / \chi_{1-\beta, \nu_{\text{eff}}}^2.$$

Thus, the coverage factor k is approximated as:

$$k \approx 1.960 \cdot \sqrt{\nu_{\text{eff}} / \chi_{1-\beta, \nu_{\text{eff}}}^2},$$

which is less than and close to 3.0, if the effective number of degrees of freedom $\nu_{\text{eff}} \geq 12$, and $\beta = 0.95$.

In summary, the bias uncertainty is pooled together with the uncertainty components in $TRSD^2$.

It should be remembered, however, that only $TRSD$ refers to quantities, which vary at each of the future measurements following the initial evaluation.

Note 3: For statisticians.

Characterizing effects of uncorrected bias:

If the systematic error (bias) is non-zero, confidence limits on the accuracy range A may be approximated as follows. The Smith-Satterthwaite approximation is generalized in approximating estimates \hat{A} in terms of a chi-square random variable χ_v^2 for the two cases in Eq 6 by:

$$\hat{A} / A = \begin{cases} \sqrt{\chi_v^2 / \nu}, & \text{if } |bias| < RSD / 1.645, \\ \chi_v^2 / \nu, & \text{otherwise} \end{cases}$$

The effective number of degrees of freedom ν is determined by forcing the variance of χ_v^2 to reproduce the estimated variance of \hat{A}^2 or \hat{A} in their respective cases:

$$\nu = \begin{cases} 2A^4 / \text{var}[\hat{A}^2], & \text{if } |bias| < RSD / 1.645 \\ 2A^2 / \text{var}[\hat{A}], & \text{otherwise} \end{cases}$$

Calculation of $\text{var}[\hat{A}^2]$ or $\text{var}[\hat{A}]$ is generally straightforward and depends on specifics of the evaluation experiment and on significant influence parameters. The confidence limit $A_{95\%}$ is then determined as in Eq 9:

$$A_{95\%} = \begin{cases} \hat{A} \times \sqrt{\nu / \chi_{0.05, \nu}^2}, & \text{if } |bias| < RSD / 1.645 \\ \hat{A} \times \nu / \chi_{0.05, \nu}^2, & \text{otherwise} \end{cases}$$

This expression has been found [12] quite accurate, exhibiting negligible effects from the discontinuity: The chi-square approximation is expected to be worst when $|bias|$ is large relative

to *TRSD*. As an example, suppose the uncertainty has the following components: 5% from pump error and also a 5% analytical relative standard deviation. Suppose *bias* = 20%. Suppose bias and the analytical uncertainty are measured with $\nu = 15$ degrees of freedom. Then 10,000-point simulations indicate that the calculated $A_{95\%}$ is slightly conservative, giving 96% confidence.

7. REFERENCES:

- [1] ISO [1993]. Guide to the expression of uncertainty in measurement. Geneva, Switzerland: International Organization for Standardization.
- [2] ISO 11222 [2002]. Air quality - Determination of the uncertainty of the time average of air quality measurements, International Organization for Standardization. Geneva, Switzerland: International Organization for Standardization.
- [3] EURACHEM/CITAC [2000]. Quantifying Uncertainty in Analytical Measurement, 2nd edition, Eurachem.
- [4] ISO 17025 [1999]. General requirements for the competence of testing and calibration laboratories. Geneva, Switzerland: International Organization for Standardization.
- [5] ISO 15767 [2003]. Workplace atmospheres - controlling and characterizing errors in weighing collected aerosols. International Organization for Standardization, Geneva, Switzerland: International Organization for Standardization.
- [6] Busch KA [1977]. SCP Statistical Protocol, in: Taylor DG, Kupel RE, and Bryant JM, eds. Documentation of the NIOSH Validation Tests. U.S. Department of Health, Education, and Welfare, Public Health Services, Centers for Disease Control, National Institute for Occupational Safety and Health, DHEW (NIOSH) Publication. No. 77-185.
- [7] Kennedy ER, Fischbach TJ, Song R, Eller PM and Shulman SA [1995]. Guidelines for Air Sampling and Analytical Method Development and Evaluation. Cincinnati, Ohio: U.S. Department of Health and Human Services, Public Health Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication No. 95-117, Cincinnati.
- [8] Vaughn NP, Chalmers CP, Botham RA. [1990]. Field comparison of personal samplers for inhalable dust. *Annals of Occupational Hygiene*, 34 (6):553-573.
- [9] ASTM D 6246 [2002]. Standard practice for evaluating the performance of diffusive samplers: West Conshohocken, Pennsylvania: American Society for Testing and Materials International.
- [10] ASTM D 6062M [1996]. Standard Guide for Personal Samplers of Health-Related Aerosol Fractions: West Conshohocken, Pennsylvania, American Society for Testing and Materials International, and references to ISO/CEN/ACGIH therein.
- [11] Bartley DL [2001]. Definition and Assessment of Sampling and Analytical Accuracy. *Annals of Occupational Hygiene* 45 (5):357-364.
- [12] Bartley DL and Irwin KC [2002]. Accuracy in measurements using diffusive samplers, Proceedings of the International Conference on Diffusive Samplers, Montpellier, France.
- [13] Hald A [1952]. A statistical theory with engineering applications. New York: John Wiley & Sons.
- [14] Wald A [1942]. Setting tolerance limits when the sample is large, *Annals of Mathematical Statistics* 13:289-299.
- [15] Wald A [1943]. An extension of Wilk's method for setting tolerance limits. *Annals of Mathematical Statistics* 14: 5-55.

- [16] Wald A , Wolfowitz J [1946]. Tolerance limits for a normal distribution. *Annals of Mathematical Statistics* 17:208-215.
- [17] Wilks AS [1941]. Determination of sample sizes for setting tolerance intervals. *Annals of Mathematical Statistics* 12:91-96.
- [18] Wilks AS [1942]. Statistical prediction with special reference to the problem of tolerance intervals, *Annals of Mathematical Statistics* 13:400-409.
- [19] Aitchison J, and Dunsmore IR [1975]. *Statistical prediction analysis*, Cambridge, England: Cambridge University Press.
- [20] EN 482 [1994] *Workplace atmospheres - General requirements for the performance of procedures for the measurement of chemical agents*. Brussels: Comité Européen de Normalization.
- [21] Smith HF [1936]: *Journal of the Council of Scientific and Industrial Research* 9:211-212.
- [22] Satterthwaite FE [1946]. An Approximate Distribution of Estimates of Variance Components. *Biometrics* 2:110-114.
- [23] ACS [1983]. Subcommittee on environmental improvement. *Anal Chem* 55:2210-2218.
- [24] Currie LA [1997]. Detection: International update, and some emerging dilemmas involving calibration, the blank, and multiple detection decisions. *Chemometrics and Intelligent Laboratory Systems* 37:151-181.

7. TERMINOLOGY

A — *symmetric accuracy range*, relative (%) range of 95% of a method's measurements about the (true) measurand

$A_{95\%}$ — 95% confidence limit on the symmetric accuracy range

$bias$ — mean concentration estimate *bias* relative to the (true) measurand

$bi\hat{a}s$ — bias estimate

\hat{c} — concentration estimate

C — true concentration

\hat{C} — reference concentration (estimate)

k — *coverage factor*, a constant containing confidence information for obtaining the expanded uncertainty U as a factor of the combined uncertainty u_c

L_D — *detection limit* (for controlling false negatives)

LOD — *limit of detection* (for controlling false positives)

n — number of measurements in a method evaluation

s_j^2 — unbiased estimate of variance σ_j^2

σ_j^2 — jth population variance component

$TRSD$ — (true) relative standard deviation

$TRSD_{ref}$ — (true) relative standard deviation in reference concentration (estimates)

u_j — jth *uncertainty* component, an estimated standard deviation

u_c — *combined uncertainty*, pooled uncertainty components

U — *expanded uncertainty*, a value giving intervals bracketing the (true) measurand at given confidence in the measurement and method evaluation

ν_j — degrees of freedom in an estimate s_j^2

$\chi_{\nu,0.05}^2$ — chi-square quantile. This quantity by definition exceeds the chi-square variable at probability = 5%. Note that many tables use the notation $\chi_{\nu,0.95}^2$ for this quantity.

In general, hats represent estimates. Primes indicate bias-corrected quantities.