Journal of Occupational and Environmental Hygiene, 1: D37–D41
ISSN: 1545-9624 print / 1545-9632 online

DOI: 10.1080/15459620490432079

Analytical Performance Criteria Reconciling Traditional Accuracy Assessment with the ISO *Guide to the Expression of Uncertainty in Measurement* (ISO/GUM)

Column Editor
Kevin Ashley, PhD

Reported by
David L. Bartley, PhD

BartlEquations, UnLtd.
Recently of the National Institute for Occupational Safety and Health

The past five years have seen a move towards standardizing the documentation of measurement uncertainty through nearly worldwide adoption of the International Organization for Standardization *Guide to the Expression of Uncertainty in Measurement* (ISO/GUM).[1–5] The question then naturally arises as to how to interpret traditional accuracy assessments of measurement methods for industrial hygiene applications.[6–8] Secondly, how are uncertainty problems specific to industrial hygiene to be handled if not covered in detail within the ISO/GUM?

These problems within workplace measurements may be summarized as follows:
(1) The expanded uncertainty U is defined in the ISO/GUM so as to bracket measurand values. Within the ISO/GUM, U depends generally on the student-t distribution. A student-t variable is a ratio of a normally distributed variable to a chi-square variable (such as an estimate s^2 of standard deviation squared). Both numerator and denominator in the ratio must vary to define the student-t variable. With workplace measurements, however, the denominator (s^2) is not remeasured (i.e., does not vary) at each practical method application. This is because workplace air concentrations vary spatially and over time to such a degree[9] that cost precludes method evaluation (i.e., by taking replicate measurements). Variability in terms of s^2 is generally measured only once in an extensive method validation, relying on adequate quality control to maintain stability of a given method. Specifics in terms of prediction in such a practice are required.
(2) Several useful methods have unknown, yet controlled or minimized, residual bias or systematic error. That is, the measured value is off from measurand values by a constant amount in all applications of the methods. Details, again in terms of prediction, are needed.

TRADITIONAL ACCURACY MEASURE

Within workplace air measurements the *symmetric accuracy range A* has been applied for evaluating candidate measurement methods and for documenting the accuracy of methods in their application.[7] 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. Explicitly,

$$C \times (1 - A) < \hat{c} < C \times (1 + A) \qquad \textbf{(1)}$$

for 95% of measurements $\hat{c}$, where the notation $\hat{c}$ is indicative of an estimate. This definition implies that the accuracy range function A must increase with both random effects described by *TRSD* (the relative standard deviation) and the estimate's mean

bias, both expressed relative to the true value C. With estimates $\hat{c}$ approximately normally distributed, the accuracy range A is given simply by[10,11]

$$A = 1.960 \times \sqrt{bias^2 + TRSD^2}$$
$$\text{if } |bias| \text{ is small (i.e., } < TRSD/1.645) \quad \textbf{(2a)}$$
$$A = |bias| + 1.645 \times TRSD \text{ if } |bias| \text{ is large} \quad \textbf{(2b)}$$

One application of the accuracy range A is to express the uncertainty in individual measurements. Suppose the confidence limit $A_{95\%} \ll 100\%$, then Equation (1) can be re-expressed so that for 95% of all method validations,

$$\hat{c} - (\hat{c} \times A_{95\%}) < C < \hat{c} + (\hat{c} \times A_{95\%})$$
$$\text{for} > 95\% \text{ of the estimates } \hat{c} \quad \textbf{(3)}$$

Therefore, $A_{95\%}$ provides intervals bracketing the true concentration C—with a double confidence sense, that is, confidence (95%) in the method validation and confidence (also taken here equal to 95%) in the ultimate application.

SUMMARY OF THE ISO/GUM

The ISO/GUM presents several concepts. One of these calls for the identification of sources (labeled $j = 1, 2, \ldots$) of uncertainty u_j (standard deviation estimate components) in a measurement method and for their classification into Type A or Type B uncertainties. A 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 υ_j degrees of freedom) of variance σ_j^2. On the contrary, Type B uncertainty generally requires professional judgment.

A common example of Type B uncertainty is the conservative assignment 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 the ISO/GUM,[1] such an assignment would be a result of sampling pump random errors having a uniform distribution so that they fall within $\pm\sqrt{3} \times 5\%$ of zero with 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 the ISO/GUM.

In the case that the estimates are independent, a combined uncertainty u_c may be computed (through the propagation of uncertainty approximation) as:

$$u_c = \sqrt{u_1^2 + u_2^2 + \cdots} \quad \textbf{(4)}$$

Through a coverage factor k, often taken as equal to 2, an expanded uncertainty U may be computed as:

$$U = k \times u_c \quad \textbf{(5)}$$

The purpose of the expanded uncertainty U is to provide tight intervals that generally contain measurand values. In particular, given a concentration estimate $\hat{c}$, the measurand value C is bracketed at better than 95% confidence by intervals of the type:

$$\hat{c} - U < C < \hat{c} + U \quad \textbf{(6)}$$

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.

EXAMPLE: RESIDUAL BIAS

Many solid sorbent tube methods for sampling and analyzing organic vapors are common.[12] Air to be assessed is drawn through the sorbent (e.g., activated charcoal) using a personal sampling pump. The compound of interest is later desorbed in a laboratory, generally for subsequent chromatographic analysis. As desorption is usually not perfect, each sorbent batch must be evaluated so that poor desorption efficiency may be compensated through a correction factor. As the factor can never be known exactly, all the samples within a given batch retain an unknown residual bias.

More complicated cases exist and are considered here. Some compounds are so problematic that the desorption efficiency is strongly loading-dependent. For instance, data on chlorobenzene indicate that the desorption efficiency varies from 70.5 to 93.0% over the working range of the sorbent tested.[12] Similar desorption effects occur with benzyl chloride, tetrachlorethylene, and carbon tetrachloride.

Suppose then that many future measurements of chlorobenzene tied to a single sorbent batch are to be compared, and that an estimate of measurement uncertainty is required based on the single batch. In a recent validation experiment seven responses $\hat{m}$ to each of four known (reference) spiked masses M were measured,[12] ranging over the expected application. It turns out that the variation in desorption efficiency can be modeled successfully as:

$$\hat{m} = M \cdot (1 + \Delta) + \delta + M\hat{\varepsilon} \quad \textbf{(7)}$$

where $\hat{\varepsilon}$ is normally distributed about zero with constant analytical variance $TRSD_a^2$, the analytical component (squared) of the combined uncertainty. Ordinary linear regression with the $n = 4 \times 7 = 28$ values provides estimates $\hat{\Delta} = -5.1\%$, $\hat{\delta} = -14.7\ \mu g$, and $T\hat{R}SD_a^2 = 3.8\%$ with $n - 2 = 26$ degrees of freedom (ignoring reference mass effects). Goodness of fit and constancy of $TRSD_a^2$ over the range of reference masses M in the validation experiment for chlorobenzene are indicated graphically in Figure 1.

Suppose that in future measurements of unknown mass X, the same model holds. Corrected estimates $\hat{x}'$ may then be calculated from responses $\hat{x}$ by:

$$\hat{x}' \equiv (\hat{x} - \hat{\delta})/(1 + \hat{\Delta}) \quad \textbf{(8)}$$

i.e., with an $\hat{x}$-dependent correction factor:

$$\text{correction factor} = \frac{(\hat{x} - \hat{\delta})}{\hat{x} \cdot (1 + \hat{\Delta})} \quad \textbf{(9)}$$

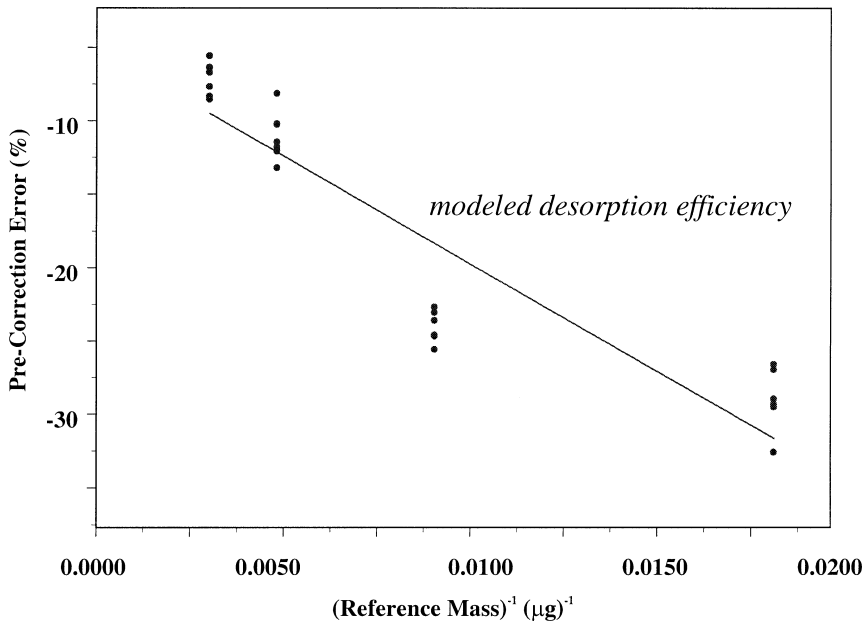


FIGURE 1. Uncorrected error $((\hat{m} - M)/M)$ vs M^{-1} (inverse spiked mass) and regression line from extraction efficiency data on chlorobenzene

Note that the estimates $\hat{\Delta}$ and $\hat{\delta}$ are constant for all samples analyzed with the same sorbent batch. Therefore, within each batch, correction leaves a small residual uncorrectable (i.e., unknown) relative bias. Because this bias is unknown and is an input into the (small-bias) version of the accuracy range A (Equation (2)), A itself is unknown. Nevertheless, a 95%-confidence limit on A can be found by noting that since the expected value of the bias is zero, an estimate of the bias squared (for A in Equation (2)) is given by the estimated variance of the bias (sorbent batch-to-batch). The result for

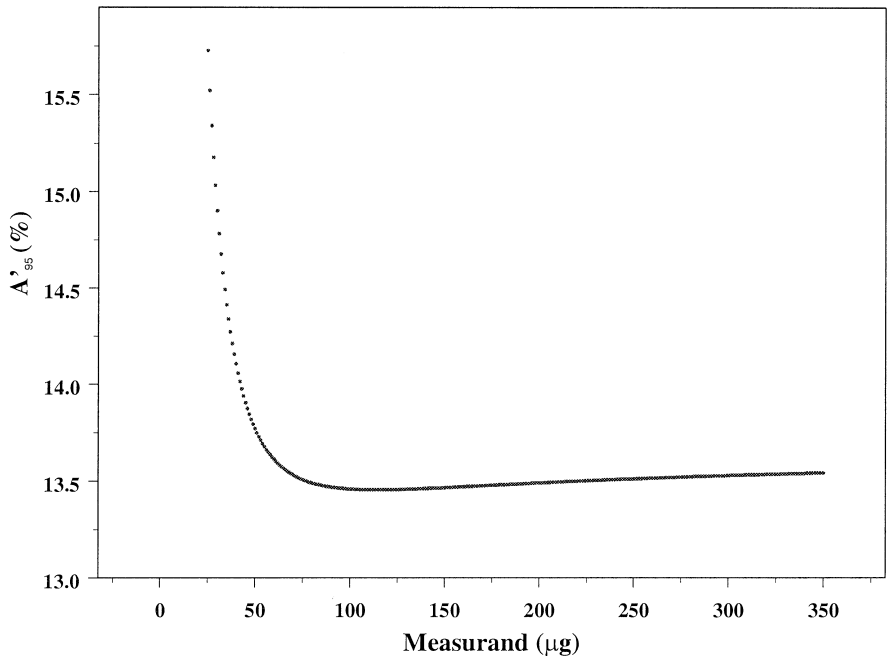


FIGURE 2. The confidence limit $A'_{95\%}$ on the symmetric accuracy range for corrected estimates of chlorobenzene loading

TABLE I. Uncertainty Budget: Batch Evaluation Confidence = 95% for Chlorobenzene

Uncertainty Components	Estimates	Computed Via
Bias correction (uncertainty in)	<3.9%	$\frac{T\hat{R}SD_a}{\sqrt{n}} \times \left(1 + \frac{(X^{-1} - \bar{M}^{-1})^2}{Var[M^{-1}]}\right)^{1/2}$
Analytical	3.8%	$T\hat{R}SD_a$
Sampling pump	5%	5%, Type B
Combined uncertainty u_c:	<7.4%	Equation 10
Expanded uncertainty U:	<15.5%	$k = 2.1$, degrees of freedom $\upsilon \approx 199$

the estimated accuracy range can then be shown to be[11]

$$\hat{A}'^2/1.960^2 = [\hat{bias}'^2 + T\hat{R}SD'^2] \approx TRSD^2_{pump} + T\hat{R}SD^2_a \times \left(1 + \frac{1}{n} + \frac{1}{n}\frac{(X^{-1} - \bar{M}^{-1})^2}{Var[M^{-1}]}\right) \quad \textbf{(10)}$$

and for the 95%-confidence limit:

$$A'_{95\%} \approx \sqrt{\frac{\upsilon}{\chi^2_{\upsilon,0.05}}}\hat{A}' \quad \textbf{(11)}$$

where the prime symbols (′) indicate that bias-corrected estimates are to be made; $TRSD^2_{pump}$ is the sampling pump uncertainty, traditionally taken to equal 5%. The quantity υ is an effective number of degrees of freedom, determined (as in Ref. 11) by approximating $\hat{A}'^2$ as chi-square distributed. The chlorobenzene data, together with an assumed infinite number of degrees of freedom in the pump standard deviation estimate, give $\upsilon = 199$. The quantity $\chi^2_{\upsilon,0.05}$ is a chi-square quantile obtainable from most statistical packages (see Ref. 11). The confidence limit is depicted in Figure 2 as dependent on the measurand X per Equation (10).

RECONCILIATION

The above description leads to equivalence of the ISO/GUM and traditional accuracy approaches, with the provisos of details specific to measurements in the workplace. If bias is negligible, then the estimated true relative standard deviation $TR\hat{S}D$ and combined uncertainty u_c are equivalent.

$$\hat{c} \times TR\hat{S}D \leftrightarrow u_c \quad \textbf{(12)}$$

In the case of potentially non-negligible, residual bias, the intervals of Equations (3) and (6) bracketing true concentrations indicate equivalence of the accuracy range confidence limit $A_{95\%}$ and expanded uncertainty U:

$$\hat{c} \times A_{95\%} \leftrightarrow U, \text{ with coverage factor } k \approx 2 - 3 \quad \textbf{(13)}$$

For example, an uncertainty budget for the expanded uncertainty U may be presented as comprised of the various components of Equation (10). A summary is given in Table I for 25 μg < M < 350 μg, where uncertainty components, combined uncertainty, and expanded uncertainty are reported relative to any future concentration estimate.

It must be remembered, of course, that with the difficulties specific to workplace measurements, the expanded uncertainty in this case quantifies confidence at 95% in (single) method validations. In the several similar cases of residual bias (as with diffusive sampling without continual calibration, and with aerosol sampling relative to an accepted size-selective convention), it can be shown[11] that if bias is corrected through a single validation-like experiment, then the bias uncertainty is simply pooled in along with other uncertainty components, as in Equations (4) and (10).

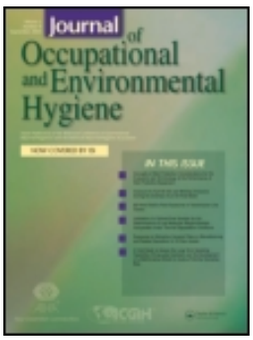
The presence of double confidence limits directly relates to the statistical theory of tolerance or prediction intervals.[13] Requiring 95% confidence in the validation experiment results in a coverage factor k involving the chi-square, rather than the student-t quantile described in the ISO/GUM. The student-t quantile provides validation confidence only in the mean (rather than the 95% confidence usually required in workplace air assessment methods).

In conclusion, the traditional accuracy assessment of industrial hygiene measurement methods carries over easily to the ISO/GUM approach now finding nearly universal favor. Comparisons of different statistical analyses of data are found to be simple. Adoption of a standard way of expressing measurement uncertainty is expected to permit a clear understanding of data and justification of its use, while maintaining a strong link with earlier, sound analysis.

REFERENCES

1. **International Organization for Standardization (ISO):** *Guide to the Expression of Uncertainty in Measurement (GUM)*. Geneva: ISO, 1993.
2. **International Organization for Standardization (ISO):** *Air Quality—Determination of the Uncertainty of the Time Average of Air Quality Measurements*. Geneva: ISO, 2002.
3. **S.L.R. Ellison, M. Rosslein, and A. Williams (eds.):** *Quantifying Uncertainty in Analytical Measurement*, 2nd ed. Caparica, Portugal: Eurachem, 2000.
4. **International Organization for Standardization (ISO):** *General Requirements for the Competence of Testing and Calibration Laboratories*. Geneva: ISO, 1999.
5. **International Organization for Standardization (ISO):** *Workplace Atmospheres—Controlling and Characterizing Errors in Weighing Collected Aerosols*. Geneva: ISO, 2003.
6. **Busch, K.A.:** Standards completion program statistical protocol. In *Documentation of the NIOSH Validation Tests*, D.G. Taylor, R.E.

Kupel, J.M. Bryant (eds.). Cincinnati, Ohio: NIOSH (DHEW), 1977.
7. **Kennedy, E.R., T.J. Fischbach, R. Song, P.M. Eller, and S.A. Shulman:** *Guidelines for Air Sampling and Analytical Method Development and Evaluation* (DHHS Pub. No. 95-117). Cincinnati, Ohio: NIOSH (DHHS), 1995.
8. **Horwitz, W., and R. Albert:** The concept of uncertainty as applied to chemical measurements. *Analyst 122*:615–617 (1997).
9. **Vaughn, N.P., C.P. Chalmers, and R.A. Botham:** Field comparison of personal samplers for inhalable dust. *Ann. Occup. Hyg. 34*:553–573 (1990).
10. **Bartley, D.L.:** Definition and assessment of sampling and analytical accuracy. *Ann. Occup. Hyg. 45*:357–364 (2001).
11. **Bartley, D.L., S.A. Shulman, and P.C. Schlecht:** Measurement uncertainty and NIOSH method accuracy range, Chapter P. In: *NIOSH Manual of Analytical Methods*, 4th ed., 3rd Suppl., DHHS (NIOSH) Publication No. 2003-154. Cincinnati, Ohio: NIOSH (CDC), 2003.
12. **National Institute for Occupational Safety and Health (NIOSH):** Method 1003. In *NIOSH Manual of Analytical Methods*, 4th ed. Cincinnati, Ohio: NIOSH, 1994.
13. **Aitchison, J., and I.R. Dunsmore:** *Statistical Prediction Analysis*. Cambridge, U.K.: Cambridge University Press, 1975.



Journal of Occupational and Environmental Hygiene

ISSN: 1545-9624 (Print) 1545-9632 (Online) Journal homepage: http://www.tandfonline.com/loi/uoeh20

Analytical Performance Criteria

Kevin Ashley & David L. Bartley

To cite this article: Kevin Ashley & David L. Bartley (2004) Analytical Performance Criteria, Journal of Occupational and Environmental Hygiene, 1:4, D37-D41, DOI: 10.1080/15459620490432079

To link to this article: https://doi.org/10.1080/15459620490432079