Ann. Occup. Hyg., Vol. 52, No. 6, pp. 413–417, 2008

doi:10.1093/annhyg/men034

Commentary

Measurement Uncertainty

DAVID BARTLEY[1]* and GÖRAN LIDÉN[2]

[1]Assistant Editor, Annals of Occupational Hygiene, 3904 Pocahontas Avenue, Cincinnati, OH 45227, USA; [2]Department of Applied Environmental Science, Stockholm University, S-106 91 Stockholm, Sweden

Received 19 May 2008; in final form 22 May 2008; published online 23 June 2008

The reporting of measurement uncertainty has recently undergone a major harmonization whereby characteristics of a measurement method obtained during establishment and application are combined componentwise. For example, the sometimes-pesky systematic error is included. A bias component of uncertainty can be often easily established as the uncertainty in the bias. However, beyond simply arriving at a value for uncertainty, meaning to this uncertainty if needed can sometimes be developed in terms of prediction confidence in uncertainty-based intervals covering what is to be measured. To this end, a link between concepts of accuracy and uncertainty is established through a simple yet accurate approximation to a random variable known as the non-central Student's *t*-distribution.

> *Without a measureless and perpetual uncertainty, the drama of human life would be destroyed.*
> Winston Churchill

Keywords: accuracy; bias; non-central Student's *t*-distribution; systematic error; uncertainty

---

Uncertainty may be generally acknowledged as perpetual, yet likewise are attempts at its measurement. Wrangling over experimental data often leads to befuddlement over what is to be done about systematic error. One often hears confession that 'I don't know anything about measurement bias; so how do I ascribe a figure of uncertainty to the data?' Then, there is a contrary question: 'I know almost exactly how biased my measurement method is; so should I fold this bias into the computed uncertainty if I do not correct bias?'

The answer to the first question is 'The uncertainty is large (infinite) if *nothing* is known about bias'. The answer to the second question is 'No!' Correctible bias refers to what is known about the measurement, and thus does not enter into uncertainty, even though a biased measurement is 'inaccurate'.

Realistically, experimental data are generally not so dire that 'nothing' is known of the bias. Usually, one may state that the bias ranges almost certainly somewhere between $\pm\Delta_{\max}$, where $\Delta_{\max}$is some number. Alternatively, the distribution of possible bias about zero may be known or easily approximated. Clearly, replicating the measurement using a method set up under the same conditions, so it has the same bias every time, will not give a variation in bias; the variation in bias appears when the experiment is repeated in different ways or under different conditions, so that the practical range of possible biases can occur.

In any case, the short answer to what is to be done about characterizing systematic error is that bias uncertainty is treated just like variation in the data, even though the randomness occurs at different regimes of experiment. Then there is nothing very complicated about arriving at a bias uncertainty 'component'. The component is an estimated standard deviation in the bias, variation occurring upon a method's setup or establishment. A *combined uncertainty* u_c is then computed as the root sum of squared bias and random components. Finally, an 'expanded uncertainty' U may be reported, multiplying the combined uncertainty by a 'coverage factor' k, now almost traditionally taken to equal the number two.

Though it is reasonably simple to arrive at a figure of uncertainty accounting for bias, the meaning is not so simple. Often only a qualitative meaning is sufficient. Intervals constructed from U and the measurement value cover what is to be measured with

---

*Author to whom correspondence should be addressed. Tel: +1 513 652 4949; e-mail: dbartley@eos.net
The free full text of this article can be found in the online version of this issue.

reasonable certainty. Also, U increases with both bias and random uncertainty.

This issue includes a paper (Bartley, 2008), which aims to provide a more precise meaning when needed. The basic idea is that often, when the bias uncertainty does not dominate the combined uncertainty, a level of confidence (e.g. 95%) in the experimental setup can be specified, so that intervals constructed from U and measurement values taken under constant conditions enclose unknown true values at greater than a specified probability. This is done by adjusting the coverage factor k from two. Alternatively, the confidence level may sometimes be computed for k fixed at two.

This approach relates to the concept of tolerance or prediction intervals (Aichison and Dunsmore, 1975; Kenny and Lidén, 1993). One of the earliest examples of tolerance intervals concerns establishing a level of confidence in predicting, at chosen probability, a limiting range about a future number drawn from a normal distribution, after first having drawn several numbers. The initial draw is analogous to a measurement method evaluation during method establishment, and the subsequent draw to the measurement itself.

Another feature of Bartley (2008) is a quantitative link between the concepts of accuracy and uncertainty. This is accomplished from the point of a random variable known as the 'non-central Student's *t*-distribution'. A simplifying though accurate approximation of this somewhat complicated variable directly provides the link in the case that bias is corrected leaving the bias uncertainty sufficiently controlled.

This link relates to a move over the past decade toward harmonizing the characterization of uncertainty. Up until the present time, researchers had employed several different ways of describing how far off a given measurement may be. For instance, within CEN TC 137—Technical Committee on the Assessment of Workplace Exposure (see the early edition, EN 482, 1994), a quantity known as the 'overall uncertainty' U_{overall} has been used. If a measurement system has relative bias Δ (systematic error) and 'imprecision' relative standard deviation (RSD), then U_{overall} is defined as

$$U_{\text{overall}} \equiv |\Delta| + 2 \times \text{RSD}. \tag{1}$$

This quantity has the desired property of increasing with either bias magnitude $|\Delta|$ or with RSD and thus is a specific quantification of inaccuracy where 'accuracy' might be defined (ISO GUM, 1995) qualitatively by

> accuracy is the closeness of agreement
> between the result of a measurement (2)
> and a true value.

U_{overall} clearly increases as this closeness declines. The function U_{overall} also has the useful feature of being extremely easy to calculate.

Peculiarly, the word accuracy is generally used in the opposite sense, meaning actually 'inaccuracy'. For instance, within the US National Institute for Occupational Safety and Health (NIOSH), a quantity known as the 'symmetric accuracy range' A has been applied for many years (Busch, 1977; Kennedy *et al.*, 1995; NIOSH, 2003). The quantity A is defined (in some respects, arbitrarily) as follows:

> A is the (relative) range symmetric
> about true values such that the range (3)
> covers 95% of measurements taken.

Like U_{overall}, equation (3) implies that the function A increases with bias magnitude $|\Delta|$ or with RSD, but is more complicated than U_{overall}. In the not-too-distant past, $A[\Delta, \text{TRSD}]$ as a function of Δ and TRSD (the 'true relative standard deviation' defined relative to true values) required a computer algorithm to calculate, even in the case that the measurements are normally distributed though biased relative to true values. However, several years ago, A was found (Bartley, 2001) to be strangely accurately approximated as

$$A[\Delta, \text{TRSD}] = \begin{cases} 1.960 \times \sqrt{\Delta^2 + \text{TRSD}^2}, & \text{if} |\Delta| < \text{TRSD}/1.645, \\ |\Delta| + 1.645 \times \text{TRSD}, & \text{if} |\Delta| > \text{TRSD}/1.645, \end{cases} \tag{4}$$

See Fig. 1 for a graphical representation of the symmetric accuracy range A. This expression is nearly as simply calculated as U_{overall}; in fact, the second line is very similar. Note that U_{overall} can be generally considered a conservative expression of A, which lends more meaning [via equation (2)] to the overall uncertainty beyond simply monotonically decreasing with 'closeness'.

Very recently, an exact expression has been discovered (K. Krishnamoorthy, T. Mathew, in preparation) for the function $A[\Delta, \mathit{TRSD}]$ in terms of a probabilistic limit (a *quantile*) on the noncentral chi-square random variable, a standard statistical item. This discovery is expected to lead to a fresh fundamental understanding of the concept of accuracy as well as new means of calculating, for instance, confidence limits on the symmetric accuracy range. At the same time, the small-$|\Delta|$ approximation in equation (4) will remain as a link between A and current means of expressing measurement uncertainty.

Another way of characterizing measurement, as done traditionally by ASTM International (ASTM D 3670), is a listing of bias magnitude $|\Delta|$ and imprecision TRSD separately, rather than combined as in either equation (1) or (3). Specifically, a 'precision and bias statement' has been required of ASTM International 'test methods', which are the most

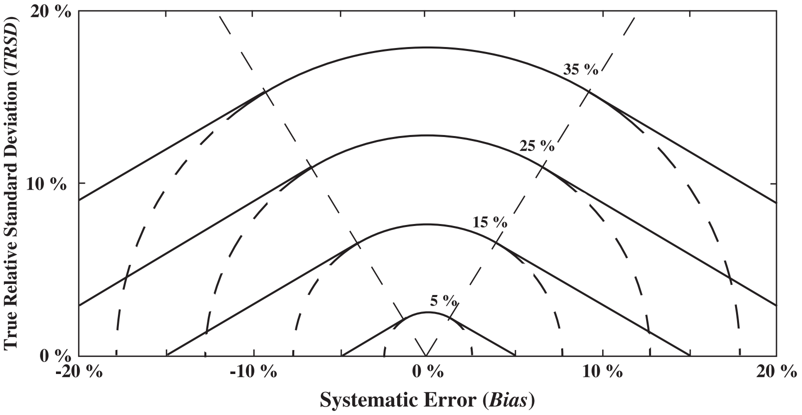

Fig. 1. The solid lines are contours of constant accuracy range *A* versus bias (Δ) and TRSD for $\alpha = 95\%$ coverage. The dashed lines indicate the boundary between the two branches of equation (4); the dashed curves were drawn merely to accentuate the circular nature of the symmetric accuracy range at bias magnitude small relative to TRSD.

carefully documented measurement methods published. The expression, 'precision and accuracy' is also heard, accuracy in this case taken as bias magnitude rather than with the meaning of equation (2).

CHANGES

More recently, many organizations have chosen harmonization in adopting procedures outlined in the *ISO Guide to the Expression of Uncertainty in Measurement* (Taylor and Kuyatt, 1994; ISO GUM, 1995). The European Union has taken a decision regarding ambient air assessment to include uncertainty characterization (EC, 2007). Furthermore, the European Committee for Standardization (CEN) has produced a guidance document (CEN BOSS, 2004) recommending uncertainty characterization. CEN TC 137—Assessment of Workplace Exposure has revised EN 482 (1994) accordingly, replacing the concept of overall uncertainty with the concept expanded uncertainty. The several working groups beneath TC 137 are currently revising existing standards in order to implement the revised EN 482 (2006) for specific situations and especially expressing the measurement uncertainty as the expanded uncertainty.

Within the International Organization for Standardization (ISO) itself, changes are also evident. At a plenary meeting in Ankara, Turkey, 2000, of ISO TC 146—Air Quality Technical Committee, a resolution was passed requiring all future ISO standards to include an uncertainty statement consistent with ISO GUM. To this end, ISO 20988 (2007) was developed as a guide. Several standards (DIS 15767, 2008; ISO 16107, 2007) have recently been revised in compliance with this resolution.

ASTM International has also recently adopted a policy whereby new test methods may optionally include a precision and bias statement or an uncertainty characterization. A relevant guide, ASTM D 7740 (2008), is currently in press. Several recently developed standards are also in compliance.

Beyond specifying uncertainty in practical measurements, ISO GUM has also been selected for evaluating analytical laboratory quality assurance. EN ISO/IEC 17025 (2005) has been adopted for this purpose in accrediting analytical laboratories. In this case, the sampling aspects (EURACHEM/CITAC GUIDE, 2007a) of a method play no role.

Aspects specific to sampling in the actual occupational setting cannot be ignored in expressing the uncertainty of an actual measurement. Within occupational hygiene, the spatiotemporal variations in air quality characteristics are generally large, precluding evaluation of a method during application through the use of replicate measurements (e.g. see Vaughan *et al.*, 1990). In this case, often an initial single-method evaluation is undertaken with the purpose of determining uncertainty present in subsequent applications of the method. Prediction confidence in such an evaluation can sometimes be specified. A related subject is measurement system control. The measurement system must remain in a state of statistical control if an introductory evaluation is to characterize later practical applications of the method. Measurement system control is evaluated using an ongoing quality control program, testing critical performance aspects for detecting problems which may develop in the method.

EXPANDED UNCERTAINTY: ISO GUM

What then are the main defining features of the expanded uncertainty *U*? As mentioned above, *U* is given by

$$U = k \times u_c, \tag{5}$$

in terms of combined uncertainty u_c and coverage factor *k* (traditionally equal to two). The combined uncertainty is the root sum of squared uncertainty components, if independent. ISO GUM is sometimes termed a 'bottom-up' approach because of analysis in terms of independent components, rather than a global measurement of the combined variation for several circumstances of intended application.

The expanded uncertainty *U* may often be easily computed for a given measurement method. Often a confidence level for the method's initial establishment may be computed. That this confidence level is not 100% generally relates to uncertainty denoted in ISO GUM as "Type B", meaning not measured at method application. "Type A" uncertainty refers to that which may be determined concurrent with the practical measurement. Type B uncertainty may refer to a lack of knowledge of the method variability. Often, however, Type B uncertainty relates to uncertain bias or systematic error.

This brings up another point emphasized in the ISO GUM approach. As mentioned earlier, uncertainty isolates what is unknown about a method and its measurement results. Therefore, correctible or known bias is not part of the uncertainty. For expressing uncertainty, bias is corrected as much as knowledge permits, and then the uncertainty in whatever residual bias remains becomes an uncertainty component.

This is the main distinction between uncertainty and accuracy. The symmetric accuracy range function *A* can refer to a method with bias measured, as during evaluation. A criterion can be established for method acceptance based on such an accuracy value. For example, NIOSH has traditionally required methods with uncorrected bias to have the 95% confidence limit on *A* to be <25% as evaluated under laboratory conditions. The interesting point, however, is that because of the fundamental definition in equation (2), if bias is corrected as much as possible, a confidence limit (e.g. at level equal to 95%) on *A* can be taken as the (relative) expanded uncertainty.

There are several common practical cases to consider. Perhaps unknown bias is normally distributed about zero. Then, the uncertainty component is the estimate of the distribution's standard deviation; in this case, the component is the square root of estimated bias squared since the expected value of the bias (as often positive as negative) is zero.

On the other hand, suppose the bias magnitude is known (or thought to be) less than a positive number $\Delta_{\max}$, but that the expected bias (as often positive as negative) is otherwise zero. Then, ISO GUM generally suggests treating the bias magnitude as uniformly distributed up to $\Delta_{\max}$. This leads to

$$\text{Variance}[\Delta] = \frac{1}{3}\Delta_{\max}^2, \qquad (6)$$

whose square root become the uncertainty component. Note again that this variation refers not to method application, but to method establishment; for example, application of a particular standard calibration material, or in the case of aerosol sampling, the selection of a site with unknown but somewhat constant particle size distribution.

A specific important example of bias in occupational hygiene is presented by inhalable aerosol sampling. A particular sampler generally cannot sample exactly according to the ideal sampling convention (ISO 7708, 1995; EN 481; Kenny *et al.*, 1997; Bartley, 1998). When used, the sampler measurements will then be biased in sampling the inhalable fraction of a specific aerosol with unknown size distribution.

The approach to this problem taken in EN 13205 (2001) is as follows. A representative set of theoretical particle size distributions is selected. Then, knowledge of the sampler collection efficiency allows calculation of the bias for each distribution. In theory, bias could then be corrected in the mean over the set of size distributions. Generally, in part for historical reasons, no such correction factors are employed when calculating the inhalable concentration from the sampled concentration. This means that different countries' samplers will on average be biased relative to each other in a partially known way. Nevertheless, the bias uncertainty is taken to equal the variance in sampling any of the representative distributions at random. Calculated as such, the bias uncertainty is independent of whether a correction is made for an average bias. The square root becomes the bias component of the measurement uncertainty.

This brings up two final issues about bias. First, some users of a particular sampler may not wish to adopt a calibration factor for correcting bias. In this case, the sampler must necessarily be characterized by its uncertainty together with the known correctible bias value or equivalently its correction factor for a representative range of particle size distributions. Secondly, if the size distribution of the sampled aerosol is known within some limits, it could be possible to estimate a much narrower range for the sampler bias (see CEN/TR 15547, 2007).

WHAT IS THE POINT?

Why harmonize uncertainty characterization according to ISO GUM? First of all, analysis of uncertainty by source components can in some cases isolate problem areas of a given measurement method. Perhaps, for example, a diffusive sampler is found to be particularly wind sensitive. Then, design steps may be undertaken to improve the sampler. Secondly, unlike a situation-dependent global or 'top-down' evaluation, component analysis can often allow prediction of sampler performance in situations not imagined during evaluation. Finally, adoption of a common way of assessing and reporting uncertainty leads to a simple understanding of reported measurements and may also bring about quality measurements within epidemiological research, in the setting of permissible workplace limits, and for making compliance (see EURACHEM/CITAC GUIDE, 2007b) judgments, all required for making rational decisions based on measurement data. Probably the dominant facilitation for this approach of component analysis results from the simplicity in the description of systematic and random errors in parallel terms.

Out of intense complexities intense simplicities emerge. Winston Churchill

REFERENCES

Aichison J, Dunsmore IR. (1975) Statistical prediction analysis. Cambridge, England: Cambridge University Press.

ASTM D 3670. (2007) Guide for determination of precision and bias of methods of committee D-22. West Conshohocken, PA: ASTM International.

ASTM D 7740. (2008) Standard practice for characterizing uncertainty in air quality measurements. West Conshohocken, PA: ASTM International.

Bartley DL. (1998) Inhalable aerosol samplers. Appl Occup Environ Hyg; 13: 274–78.

Bartley DL. (2001) Definition and assessment of sampling and analytical accuracy. Ann occup Hyg; 45: 357–64.

Bartley DL. (2008) The meaning of the bias uncertainty measure. Ann occup Hyg;52: 519–25.

Busch KA. (1977) Standards completion program statistical protocol. In: Taylor DG, Kupel RE and Bryant JM, editors. Documentation of the NIOSH Validation Tests. Cincinnati, OH: U.S. Department of Health, Education, and Welfare (DHEW), Centers for Disease Control (CDC), National Institute for Occupational Safety and Health (NIOSH) [DHEW (NIOSH) Publ. No. 77–185], p. 7–12.

CEN BOSS. (2004) Guidance—uncertainty of measurement concept in European Standards. Available at http://www.cen.eu/boss/supporting/guidance+documents/gd063+-+uncertainty+of+measurements/index.asp. Accessed 10 June 2008.

CEN/TR 15547. (2007) Workplace atmospheres—calculation of the health-related aerosol fraction concentration from the concentration measured by a sampler with known performance characteristics. Brussels, Belgium: Comité Européen de Normalisation.

DIS 15767. (2008) Workplace atmospheres—controlling and characterizing errors in weighing collected aerosol. Geneva, Switzerland: International Organization for Standardization.

EC. (2007) Common Position (EC) No 13/2007 of 25 June 2007 adopted by the Council, acting in accordance with the procedure referred to in Article 251 of the Treaty establishing the European Community, with a view to the adoption of a Directive of the European Parliament and of the Council on ambient air quality and cleaner air for Europe. Available at http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52007AG0013:EN:NOT. Accessed 22 May 2008.

EN 481. (1993) Workplace atmospheres—specification for conventions for measurement of suspended matter in workplace atmospheres. Brussels, Belgium: Comité Européen de Normalisation.

EN 482. (1994) Workplace atmospheres—general requirements for the performance of procedures for the measurement of chemical agents. Brussels, Belgium: Comité Européen de Normalisation.

EN 482. (2006) Workplace atmospheres—general requirements for the performance of procedures for the measurement of chemical agents. Brussels, Belgium: Comité Européen de Normalisation.

EN 13205. (2001) Workplace atmospheres—assessment of performance of instruments for measurement of airborne particle concentrations. Brussels, Belgium: Comité Européen de Normalisation.

EN ISO/IEC 17025. (2005) General requirements for the competence of testing and calibration laboratories. Geneva, Switzerland: International Organization for Standardization.

EURACHEM/CITAC GUIDE. (2007a) Measurement uncertainty arising from sampling, 1st edn (http://www.eurachem.org/). Accessed 10 June 2008.

EURACHEM/CITAC GUIDE. (2007b) Use of uncertainty information in compliance assessment, 1st edn (http://www.eurachem.org/). Accessed 10 June 2008.

ISO GUM. (1995) Guide to the expression of uncertainty in measurement, ISO Guide 98. Geneva, Switzerland: International Organization for Standardization. (See Taylor and Kuyatt, 1994, giving initial publication.).

ISO 7708. (1995) Air quality—particle size fraction definitions for health-related sampling. Geneva, Switzerland: International Organization for Standardization.

ISO 16107. (2007) Workplace atmospheres—protocol for evaluating the performance of diffusive samplers. Geneva, Switzerland: International Organization for Standardization.

ISO 20988. (2007) Air quality—guidelines for estimating measurement uncertainty. Geneva, Switzerland: International Organization for Standardization.

Kennedy ER, Fischbach TJ, Song R *et al.* (1995) Guidelines for air sampling and analytical method development and evaluation. Cincinnati, OH: US Department of Health and Human Services (DHHS), CDC/NIOSH [DHHS Publ. No. 95–117] p. 1–104.

Kenny LC, Aitken R, Chalmers C *et al.* (1997) Outcome of a collaborative European study of personal inhalable sampler performance. Ann occup Hyg; 41: 135–54.

Kenny LC, Lidén G. (1993) The application of performance standards to personal airborne dust samplers. Ann occup Hyg; 33: 289–300.

Krishnamoorthy K, Mathew T. (2008) Inference on the symmetric accuracy range. Private communication; (manuscript under preparation).

NIOSH. (2003) Measurement uncertainty and NIOSH method accuracy range, Chapter P, NIOSH manual of analytical methods. In: Schlecht PC and O'Connor PF, editors. 4th edn, 3rd Suppl, DHHS (NIOSH) Publication No. 2003-154. Washington, DC: US Government Printing Office, p. 208–228.

Taylor BN, Kuyatt CE. (1994) Guidelines for evaluating and expressing the uncertainty of NIST measurement results, NIST Technical Note 1297. Washington, DC: US Government Printing Office.

Vaughan NP, Chalmers CP, Botham RA. (1990) Field comparison of personal samplers for inhalable dust. Ann occup Hyg; 34: 553.