TECHNICAL COMMUNICATIONS

The Certainty of Uncertainty

WILLIAM HORWITZ
whorwitz@aoac.org

The official definition of measurement uncertainty (from the NIST website http://physics.nist.gov/cuu/Uncertainty/glossary.html) (emphasis added) is:

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

Considerable confusion about this term will be swept away immediately if you note that *the term "UNCERTAINTY" is attached to a RESULT, not to a method*; i.e., *measurement uncertainty* is being discussed, not *method uncertainty*. We will see how the method gets into the discussion later.

The introductory chapters to practically every textbook of quantitative analysis discusses the variability of analytical results and often advises reporting results in terms of the mean of a series of replicates and an interval within which you expect most (i.e., 95%) of your future results to fall if the future analyses were conducted in an identical manner. However, the economics of chemical analysis dictates that only a few analyses are conducted on a test sample ("the results are usually good enough for government work") so this theoretical admonition has been largely ignored until recently. Now, for accreditation purposes, laboratories are required to attach a statement of *measurement uncertainty* to their analytical results.

To obtain that halo of uncertainty surrounding your reported result you have essentially four options:

(*1*) The option of calculating the equivalent of a confidence interval from the "t" factor applied to the standard deviation of replicates.

(*2*) The theoretical "bottom-up" approach recommended by the bible on uncertainty rubber stamped by nine international organizations (1).

(*3*) The practical "top-down" approach from the relative standard deviation derived from an interlaboratory study by the Harmonized IUPAC/AOAC protocol or ISO 5725.

(*4*) The estimate obtained by applying the Horwitz formula relating the relative standard deviation to concentration, as a mass fraction, $RSD_R = 2C^{(-0.15)}$, which is based upon a review of over 10 000 interlaboratory results, primarily published in the *Journal of AOAC INTERNATIONAL*.

OPTION 1

Run sufficient replicates on the specific test sample under consideration to obtain a fairly good idea of how the results will scatter in routine work. If you manufacture a product to a specification of 20% fat day in and day out, with the help of a statistician, you would soon be able to know the typical uncertainty of the fat content of the product, of the sampling, and of the analysis. But if you are called upon to provide an *estimate of uncertainty* from a set of duplicates from a material you will never see again, you will have to multiply the standard deviation calculated from that pair of results by a factor of 12! Such an estimate is essentially useless because experience shows that future analysis from even a moderately experienced analyst will rarely approach the expected extreme.

Incidentally, running more replicates will not change the "true value" of the mean or of the standard deviation. More replicates provide more confidence in the interval estimate bracketing the true concentration and the true standard deviation.

OPTION 2

Sit down and think about everything that might possibly affect the result and estimate the expected variation that each factor will contribute to the final value. These will include uncertainties, expressed as standard deviations, from:

- standard weight corrections
- buoyancy corrections (temperature, pressure)
- volumetric flask corrections (calibration, temperature)
- pipet volume corrections (calibration, temperature)
- reference material content uncertainty
- concentration of calibrant uncertainty
- signal measurement uncertainty
- time measurement uncertainty

Received October 29, 2002. Accepted by RR December 18, 2002.

- extraction variability (volume, temperature, and solubility effects)
- reaction or separation variability
- effect of interferences which may or may not be present
- etc.
- etc.

When you have thought of everything that might possibly influence your reaction, separation, and measurement, and assigned a standard deviation to each factor, calculate the square root of the linear combination of the variances to obtain the final standard deviation that you attach to your measurement as the measurement uncertainty. Then multiply this final standard deviation by a coverage factor (k) of 2 to ensure a probability of 95%, i.e., only a 5% chance that the true value lies outside the expanded uncertainty limits. Incidentally, do not forget lot and analytical sampling, which is unique for every lot and which, therefore, requires individual estimation by replication of these components for completeness. "Practical" examples will be found in a EURACHEM guide (2).

This is known as the bottom-up approach. You can come back later and add in those factors that you initially overlooked or which are pointed out to you by your colleagues or by your friendly assessor months after the report has been delivered and forgotten.

This absurd and budget-busting approach (for analytical chemistry) arose from metrological chemists taking over in entirety the concepts developed by metrologists for physical processes measured with 5–9 significant figures (gravitational constant, speed of light, etc.) and applying them to analytical chemistry measurements with 2 or 3 significant figures. This approach also ignores the fact that some chemical methods are influenced by numerous factors, some positive and some negative, that tend to cancel out, and that often other chemical methods are influenced by a few factors that overwhelm the weight and volume uncertainty calculations presented in the published examples.

OPTION 3

The approach, which is becoming generally accepted in Europe, is to conduct an interlaboratory study utilizing the Harmonized IUPAC/AOAC or ISO 5725 protocol (which utilize an identical statistical model except for outlier removal). These protocols require a sample of at least 8 typical laboratories analyzing a minimum set of 5 matrices covering the range of materials of interest. Then relate the standard deviation among laboratories (S_R) as being proportional to measurement uncertainty. This is known as the top-down approach. By utilizing a sample of presumably typical laboratories operating in different environments on at least 5 materials covering the range of interest, it is very likely that most of the potential error factors that are likely to be encountered in practice will have been introduced. Therefore if we equate this S_R to measurement uncertainty and call it standard measurement uncertainty (standard uncertainty for short) we are at least about 70% certain that our result plus and minus S_R will encompass the "true" value. If we multiply S_R by a coverage factor of 2 we obtain the "expanded measurement uncertainty" (expanded uncertainty for short) we are now at least 95% certain that our result plus and minus $2S_R$ will encompass the "true" value.

When using this collaborative study approach, which results in a "standard method" as used by ISO 17025, be sure that all of the important variables are specified or understood (*see Definition of Terms and Explanatory Notes* section of the *Official Methods of AOAC INTERNATIONAL*) with assigned limits. Weights are assumed to be within ±10% (but use the actual weight for calculations), volumetric glassware are assumed to have their assigned volume with negligible uncertainty when used with instrumental methods (but not when used in titrations), graduates are assumed to deliver the volume read from their scale, temperatures are set to be within ±2°, pHs are within ±0.05 unit, times are followed to within 5%, and instrument scales, dials, and markers are estimated to their finest degree, then Clause 5.4.6.2 Note 2 in ISO 17025 reading, "In those cases where a well-recognized test method specifies limits to the values of the major sources of uncertainty of measurement and specify the form of presentation of the calculated results, the laboratory is considered to have satisfied this clause by following the test method and reporting instructions." Under such conditions, S_R derived from the supporting collaborative study in the same units as the reported result with the accompanying number of significant figures, usually 2 or 3, may be used as the standard uncertainty, assuming the laboratory has demonstrated that it operates within the performance limits for that method.

OPTION 4 OR 0

As a last resort, or even before you start any analyses, you can make a rough calculation to determine if the expected uncertainty at the expected concentration will be fit for the intended purpose. Apply the Horwitz formula (or a suitably adjusted version of the Horwitz formula to account for special circumstances such as a single laboratory) to the anticipated concentration to obtain a within-laboratory S_r and multiply it by 2 to obtain the expanded uncertainty. The Horwitz formula as initially applied to among-laboratory reproducibility parameters in %, and with C expressed as a mass fraction, is

$$RSD_R \text{ (in \%)} = 2C\text{^}(-0.15)$$

or as a standard deviation

$$S_R = 0.02C\text{^}(0.85).$$

To apply to within-laboratory repeatability parameters, divide by 2 and equate this to estimated standard uncertainty:

$$S_r = 0.01C\text{^}(0.85);$$

To obtain the expanded (repeatability) uncertainty, multiply by 2:

$$S_r = 0.02C\text{^}(0.85).$$

For example, if we are dealing with a pure compendial material, C expressed as a mass fraction is 1, so the

anticipated expanded uncertainty, $2 \cdot S_r$, is 0.04 or 4%. This is interpreted as 95% of anticipated results will fall between 96 and 104%. You can "improve" your uncertainty by running independent replicates. "Independent" means as a minimum "non-simultaneous," but again economics would not permit it, so the improvement would be considerably less than theoretical.

Summary: The Horwitz formula will tell you if your anticipated uncertainty is such that you will be within the limits of the ballpark with a typical method. The maximum spread obtained by the top-down approach will encompass the "true value" in almost all practical cases. It is usually easier to let nature slip in all the unanticipatable tricks that can befall even the most careful analysts than to valiantly attempt to foresee them beforehand by the budget approach. This is how the uncertainty of the method becomes entangled with the uncertainty of the measurement.

Note 1: Some of these "unanticipatable tricks" are chaotic, like dropping the thermometer or missing a decimal point. They are not subject to statistical description. Such adventitious flaws are handled by quality control but they cannot be predicted in any quantitative way. Such flaws are not intrinsic to the method.

Note 2: The uncertainty of a method, its bias and variability, is revealed by the spread of the individual measurements, i.e., by the average and standard deviation of the set of measurements. The theory envisions an infinite set of concentration estimates is obtained for each true concentration but the hapless finite chemist is forced just to take a sampling from this infinite set at the given concentration, usually just one or two estimates. Outlier tests are applied to remove clearly extrinsic interferences with the proper application of the chemical method. Note also that the uncertainty components, both bias and variability, are functions of the true concentration, though variability is usually observed to be more concentration dependent than the bias.

If a method is to be corrected for recovery (bias) the method will usually so indicate. Many regulatory methods do not require such a correction because the specification (tolerance) was established by the same method so the recovery is "built into" the specification.

Note 3: The analytical chemist usually ignores sampling uncertainty primarily because typically little or no information accompanies the laboratory sample as to whether or not the laboratory sample truly reflects the lot. It is usually left to "management" to coordinate the analytical information with the sampling information. However, if the sample has been collected according to statistical principles (which usually requires a very large number of increments) and if these increments have been analyzed to provide the basis for an estimate of sampling uncertainty, then propagation of error considerations can provide an overall "sampling + analysis" uncertainty.

Note 4: We have deliberately omitted mentioning the problem of expressing measurement and method uncertainties of microbiological examinations where the target analyte is intentionally diluted to the point of producing "true" false positives and "true" false negatives for comparison of the results from a test method to those from a reference method.

(1) *Guide to the Expression of Uncertainty in Measurement.* ISO, Geneva Switzerland.

(2) EURACHEM "Quantifying Uncertainty in Analytical Measurement" 2nd Edition (2000), http://www.measurement uncertainty.org.