amc technical brief

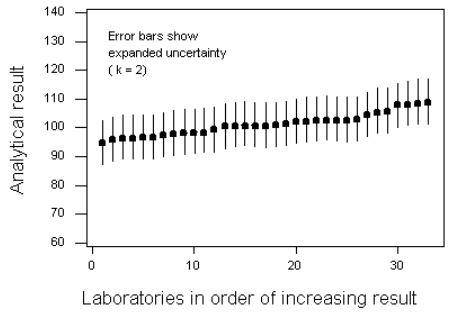
Analytical Method Committee No. 15 Dec 2003

Is my uncertainty estimate realistic?

Analysts' estimates of the uncertainty of their results are often somewhat low. How do we know? By looking at the results of interlaboratory studies such as collaborative trials and proficiency tests. These studies are designed to make explicit any latent contributions to uncertainty. The results can be helpful in assessing the validity of our uncertainty estimates.

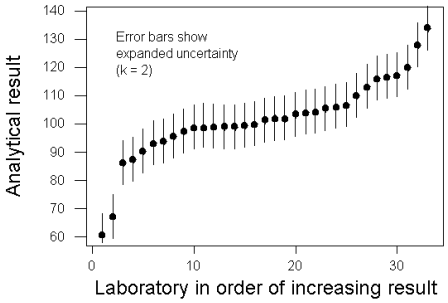
Consider a hypothetical example. Suppose we have group of laboratories, each of which analyses the same homogeneous material for a minor constituent present at a concentration of 100 ppm. Suppose also, to keep things simple, that they all report the same estimated standard uncertainty of measurement, namely 4 ppm. Under these conditions, we ought to see a set of results that look like those in Figure 1. There the between-laboratory standard deviation $(\sigma_R = 4)$ is explained fully by the uncertainties in the individual measurements.

Figure 1. Results of concordant laboratories



What we actually see is more like Figure 2: there are clear differences among the laboratories that are not explicable in terms of the estimated uncertainties.

Figure 2. Results of non-concordant laboratories



Figures 1 and 2 are not completely realistic in that we would expect the uncertainties from different laboratories to vary somewhat. Furthermore, diagrams showing individual uncertainties have not so far been common in routine proficiency tests. Nevertheless, there is good evidence to show that the underlying situation is very often exactly as shown.

We can draw two immediate conclusions from a situation such as that in Figure 2:

- There are sources of error contributing to the dispersion of results that many, perhaps most of the participants did not take into account in their uncertainty budgets.
- Until these additional sources of error are understood and properly incorporated into the individual uncertainty estimates, the estimates cannot be regarded as adequate or realistic.

Interlaboratory studies

A collaborative trial is designed to explore the performance of a particular analytical method applied to a specified type of test material. All of the participant laboratories apply the same closely defined analytical procedure to the same set of materials. The main outcome of the study is separate estimates of repeatability and reproducibility standard deviations (σ_r and σ_R respectively), which are regarded as characteristics of the method. Repeatability conditions are those prevailing within a single analytical run. A standard deviation based on repeated results obtained under repeatability conditions can never incorporate all the factors that are relevant to an uncertainty estimate. The reproducibility (or between-laboratory) standard deviation, however, also takes account of variation due to

- different interpretations of the method protocol in the various laboratories;
- different occasions (runs) when the method is used within a laboratory, perhaps due to different analysts, different equipment and new calibration curves.
- many other systematic errors of individual laboratories, such as long-term calibration differences, different reference (that is, calibration) material batches, permitted variation in ambient conditions etc.

Clearly all of these additional effects, where present, should contribute to the uncertainty estimate. For most laboratories, therefore, the reproducibility standard deviation provides a better estimate of the uncertainty introduced into the result than any estimate that does not make allowances for such effects.

How large are these additional effects?

On average, we find in collaborative trials for a single method that

Eq 1 $$\sigma_r \approx 0.5\sigma_R ,$$

which is an indication of the magnitude of the 'missing' uncertainty.

We can also estimate the possible biases associated with particular analytical methods. These can arise, for example, through variations in the recovery when the analyte is transferred from the test material into the test solution, and uncorrected interference effects.

Proficiency tests do not usually prescribe specific methods; as a result, inter-method effects appear as additional dispersion in the results. A robust statistical treatment of proficiency test data shows that the between-laboratory standard deviation $\sigma_{R'}$ under these conditions is on average

Eq 2 $$\sigma_{R'} \approx 1.5\sigma_R$$

In real life (as opposed to the specially designed studies considered above) there may be further sources of error that may need to be taken into account, but for the present purposes we can restrict ourselves to those manifested in collaborative trials and proficiency tests. These latent sources of error are present, not only in special studies but, in routine analytical results as well.

Consideration of the above enables us to assert the following: **unless the individual laboratory or analyst takes extraordinary precautions to avoid them and to confirm their absence, effects leading to method bias, laboratory bias and run bias are present at substantial levels in routine measurement, and therefore contribute to the uncertainty. If these contributions are omitted from the uncertainty budget, the uncertainty will be underestimated, often substantially.**
Note: Method bias is absent for empirical or 'operationally defined' analytes, where the analyte is defined in terms of response to a particular procedure.

How can we check that an uncertainty estimate is realistic?
The basic principle is to compare the uncertainty estimate with an estimate of reproducibility standard deviation; if the uncertainty estimate is much smaller, suspect that important contributions have been omitted. (Some laboratories, of course, may have good grounds for claiming smaller uncertainty smaller than σ_R, but that needs special justification.) Any of the following indicators would be helpful in judging an uncertainty estimate.

- ***Check against collaborative trial statistics.***
  Compare an estimate of standard uncertainty with a reproducibility standard deviation σ_R from a collaborative trial of the same method for the same analyte. The standard uncertainty should be at least as large as σ_R unless, as stated above, unusually stringent attempts have been made to minimise high-level errors. It may be necessary to interpolate between the collaborative trial results from different concentrations of the analyte to find a value for the uncertainty at an appropriate concentration. Note, too, that while a laboratory's standard uncertainty for routine analysis is unlikely to be less than the estimate σ_R from a collaborative trial, it is quite possible for it to be greater (and still, conceivably, fit for purpose).
  Note: A recent draft ISO Technical Specification provides detailed procedures for both checking and preparing uncertainty estimates using collaborative study data [1].

- ***Compare the uncertainty with $\hat{\sigma}_R$ estimated from available repeatability (within-run) precision statistics or from run-to-run statistics.***
  Obtain an estimate of $\hat{\sigma}_R$ from Eq 1 as $\hat{\sigma}_R \approx 2s_r$ where s_r is obtained from repeated results obtained over the duration of a typical run. It is important to carry out a full replication of the procedure from the point at which the test portion is weighed out from the laboratory sample, otherwise the estimate s_r will be much too small.

  An alternative estimate of $\hat{\sigma}_R$ could be obtained from standard deviations s_{run} obtained from run-to-run repeated results, such as those produced in routine internal quality control. In that case a reasonable estimate of $\hat{\sigma}_R = 1.5s_{run}$ can be found by use of the approximate relationship. In addition, if there is a known bias b in the results of the IQC material, make an extra allowance by using the modified relationship $\hat{\sigma}_R = \sqrt{(1.5s_{run})^2 + b^2}$.

  Note: This last equation provides a useful and simple method of checking an uncertainty estimate for realism, but there is currently no broad consensus on treatment of known but uncorrected bias in uncertainty estimates.

- ***Examine your proficiency test results* [2].**
  This method assumes that your z-scores represent routine analytical conditions, and that the scheme uses prescribed σ_p -values (the standard deviations used by the scheme to calculate z-scores, probably expressed as a function of concentration) to characterise the required uncertainty. If your collected z-scores for the determination in question over a recent period can be taken as zero-centred, and with a standard deviation of unity, then your real uncertainty will be consistent with the scheme's prescribed uncertainties over the relevant concentration range. If the mean z-score is significantly different from zero, or the standard deviation is significantly greater than unity, then your uncertainty is probably worse than the scheme's prescription.

How can we correct an unrealistic uncertainty estimate?
There are several approaches:

i) Identify the effects causing the problem and eliminate them through further method development or improved quality procedures.

ii) Identify the specific effects that cause the problem and either include a proper correction for the effect (as in correcting volume measurements for temperature effects), or include additional quantitative terms for each effect in the uncertainty budget.

Both i) and ii) are ideal approaches, recommended where practicable but seldom available in routine analysis using standard methods.

iii) Base the uncertainty estimate on interlaboratory study. With due attention to systematic effects, this approach is considered acceptable by both the Eurachem guide and ISO DTS 21748 and can be recommended for routine analysis. It is not currently considered acceptable for calibration laboratories.

iv) Increase the expansion factor k applied in order to obtain the 'expanded uncertainty'.

v) Add an arbitrary term to the uncertainty budget.

Approach iv) is appropriate where information on random or other effects is sparse; rounding k up to 3 instead of using a factor k=2 is eminently justified in such cases on the grounds that the effective degrees of freedom are genuinely low and k=2 provides inadequate coverage. Approach v) is not recommended because the origin of the effect has not been identified and it is consequently hard to provide any consistent rationale for the size of the additional term.

References

1. ISO DTS 21748: *Guide to the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation.*
2. AMC Technical Brief No. 11. *Understanding and acting on scores obtained in proficiency testing schemes.*