amc technical brief

Editor: M Thompson Analytical Methods Committee AMCTB No 2 Revised December 2005

The z_L-score—combining your proficiency test results with your own fitness-for-purpose criterion

The 2006 revision of the Harmonised Protocol [1] encourages providers of proficiency testing schemes to base their scoring method on fitness for purpose. The resulting z-score relates the deviation of the participant's score from the assigned value to an uncertainty, determined by the provider, that is widely regarded as fit for purpose in the specific application. In some circumstances, however, a participant may wish to apply a different criterion to the result, to represent fitness for a purpose different to that envisaged by the scheme provider. These recalculated 'z_L-scores' are appropriate for criteria agreed between laboratories and their customers.

Background

Proficiency testing is a method for regularly assessing the accuracy of results in laboratories conducting particular measurements. In analytical chemistry, proficiency testing usually comprises the distribution of effectively identical portions of the test material to each participant for analysis as an unknown. The laboratories conduct the test under routine conditions, and report the result to the organiser by a deadline. The organiser then converts the result to a score which helps the participant assess the accuracy of the result in relation to a fitness for purpose criterion defined by the scheme provider. (See Technical Briefs Nos 11 and 16, and Background Paper No 2 for more details [2-4].)

The primary purpose of the proficiency test (PT) is to allow the participants to confirm that they are complying with the external criterion or, failing that, to detect unexpected errors in their results. Unexpected errors should trigger an investigation of causes of the problem and, if necessary, remedial activity. Proficiency tests have also acquired secondary purposes beyond the original self-help ethos. Accreditation agencies usually require that candidate laboratories (a) participate in appropriate proficiency tests where available, (b) perform satisfactorily overall, and (c) have a procedure for investigating exceptional errors when they occur. Moreover, laboratories are increasingly using PT results to demonstrate competence in their bids for contract analytical work.

Scoring systems

Most proficiency testing schemes in analytical chemistry use the scoring system recommended in the Harmonised Protocol. In this system, the participant's result x is converted into a z-score given by the equation:

$$z = (x - x_a)/\sigma_p,$$

where x_a is the assigned value, the provider's best estimate of the true value, and σ_p is the standard deviation for proficiency assessment (previously called the 'target value of standard deviation'). $(x - x_a)$ is the estimate of the error in the result, and z is the same error scaled to σ_p.

In an ideal PT scheme, the value given to σ_p is determined by fitness for purpose: it represents the amount of uncertainty in the result that is tolerable in relation to the purpose of the analysis. Notice that here σ_p describes the end-user's requirements, not the data.

σ_p describes the end-user's requirements, not the data

It is then simple for the participant to define limits, based on the value of z, to initiate investigatory or remedial action. If the participants all complied exactly with the criterion we would expect z for a particular round to be the random normal deviate, that is, a random normal variable with a mean of zero and a variance of unity $N(0,1)$. That is why many providers and participants regard a value of z falling between ±2 as indicating satisfactory performance and a value $|z| > 3$ as requiring investigation.

Participants only rarely perform exactly in line with the fitness criterion, so sets of z-scores show a

variety of deviations from $N(0,1)$. Never-the-less, except in exceptional circumstances, the properties of the normal distribution are still useful for defining action limits when σ_p represents fitness for purpose. If, however, you make σ_p describe the *data* (rather than the requirements), for example by setting the value to the robust standard deviation of the participants' results, the z-scores will have a standard deviation close to unity and slightly less than 95% of them will fall into the 'satisfactory' category, irrespective of whether or not the results are fit for purpose.

Do-it-yourself scoring

A problem sometimes encountered by participants is that their customer's fitness-for purpose criterion differs from that of the PT scheme that they are using. This can easily happen: the PT scheme sets its criterion for the sector of analysis in general, while the participant deals with specialised applications.

If the scheme provider's σ_p criterion were smaller than required for the specialist application, the participant could get a poor z-score in the PT scheme, but would do better if the target value were compatible with the customer's requirements.

The recommended course of action is for the participant to calculate an auxiliary score called the 'z_L-score', given by:

$z_L = (x - x_a)/u_f$.

In this equation, u_f is the specification, agreed between the participant and the end-user, of an uncertainty that is fit for purpose. The laboratory should take the PT scheme's assigned value for the calculation. The z_L-score is therefore a customised z-score that applies to the participant's individual circumstances. For accreditation or contractual purposes, the participant can list the z_L-scores obtained and show the u_f values on which they are based. The value of u_f would have to be demonstrably justifiable and, of course, may need to vary with the concentration of the analyte.

u_f is the specification, agreed between the participant and the end-user, of an uncertainty that is fit for purpose

(*Note*: the z_L-score was called the 'zeta-score' in the original version of this document and in other documents. The name had to be changed because of the subsequent use of 'zeta-score' for other purposes in a normative document.)

The z_L-score and the role of the scheme provider

It would be ineffectual for proficiency test providers to carry out z_L-score calculations based on submitted uncertainties. The organisers would have no control over whether the uncertainties were appropriate, and would therefore be unable to attribute any meaning to the scores based on them. In addition, participants might have several different fitness-for-purpose criteria for different customers, each of which could generate a different score. It is therefore more appropriate for the individual participants to calculate and record their own z_L-scores, in consultation with their customers.

References

1 M Thompson, S L R Ellison, R Wood, 'International Harmonised Protocol for the Proficiency Testing of Analytical Chemistry Laboratories', *Pure Appl. Chem.*, 2006, **78**, 145-196.

2 AMC Technical Brief No 11 (2002), can be downloaded from www.rsc.org/amc/

3 AMC Technical Brief No 16 (2004), can be downloaded from www.rsc.org/amc/

4 AMC Background Paper No 2 (2005), can be downloaded from www.rsc.org/amc/

This Technical Brief was prepared and revised for the Analytical Methods Committee by the Statistical Subcommittee (Chairman M Thompson)

AMC Technical Briefs (Editor M Thompson) are informal but authoritative bulletins on technical matters of interest to the analytical community. They are prepared by the Analytical Methods Committee of the Analytical Division of the RSC, and are carefully scrutinised for accuracy. Correspondence should be addressed to: The Analytical Methods Committee, The Royal Society of Chemistry, Burlington House, Piccadilly, London W1V 0BN.