The Analyst
COMMUNICATION

Recent trends in inter-laboratory precision at ppb and sub-ppb concentrations in relation to fitness for purpose criteria in proficiency testing

Michael Thompson

School of Biological and Chemical Sciences, Birkbeck College (University of London), Gordon House, 29 Gordon Square, London, UK WC1H 0PP

Received 12th January 2000, Accepted 31st January 2000

Recently conducted collaborative trials in which the analyte concentration was below 100 ppb provided reproducibility standard deviations that were systematically lower than the predictions of the Horwitz function. This study shows that such statistics are better represented by a model with a constant relative standard deviation. A modified function is suggested as suitable for use (with due caution) as a fitness-for-purpose criterion.

Introduction

The Horwitz function, $\sigma_H = 0.02c^{0.8495}$, is a useful generalisation about the reproducibility standard deviation (expressing inter-laboratory precision) expected in a collaborative trial.[1,2] In the equation both the expected standard deviation σ_H and the concentration c are expressed in dimensionless mass ratios (for example, 1 ppm $\equiv 10^{-6}$). This relationship is so widely recognised that it is used both as a benchmark to judge the efficacy of collaborative trials[3] and as a fitness-for-purpose criterion in proficiency testing in the food and other sectors.[4–6]

In 1996, however, Horwitz[7] reported that, at the low concentrations of analyte encountered in the analysis of pesticides, estimates of the reproducibility standard deviation (σ_R) were consistently lower than σ_H. Furthermore, the same tendency was reported and discussed in 1997 in a study of the experimental basis of the Horwitz function.[2] In the latter study the data remarkably showed laboratories achieving reproducibility standard deviations that clustered around a trend, $\sigma_R = c/3$, that could be regarded as a definition of the 'reproducibility detection limit'. The new trend was therefore attributed to the practical requirement that the reproducibility precision must be no worse than that associated with the reproducibility detection limit of the method, if the method was to be usable. In other words, at concentrations below 10 ppb, the Horwitz function predicted inter-laboratory precisions so poor that, if they were realised in practice, there would be doubt about the presence or absence of the analyte. Laboratories, when they needed to, could agree with each other more closely than predicted by the Horwitz function.

It was also clear that there was a more restricted deviation from the Horwitz function at higher concentrations: at concentrations greater than about 10^{-1} (10% m/m) the reproducibility precision was on average again somewhat smaller than σ_H. The trend of the data at these concentrations could be represented as $\sigma = 0.01c^{0.5}$. This line intersects the Horwitz function at a concentration of about $10^{-0.86}$, that is, 13.8% m/m.

These facts have implications for the use of σ_H as a fitness-for-purpose criterion in proficiency tests. Previously it has been argued that the Horwitz function was an appropriate criterion, at least down to 10^{-8}, because analytical methods tend to evolve towards fitness for purpose by a kind of natural selection.[6] However, it was clear that the function should not used in that context for proficiency tests at the very low concentrations appropriate for analytes such as mycotoxins, *etc*.

In the present study data from recent (post-1997) collaborative trials, all involving analytes at concentrations below 10 ppb, were examined to see if the previously noted trend was being maintained, and whether a modification of the Horwitz function could be formulated to serve as an objective fitness-for-purpose criterion. The trials all related to the determination of mycotoxins and, in all, 47 different trial materials were analysed in nine separate studies.

Results and discussion

The results of the collaborative trials are given in Fig. 1, which shows the $\log_{10}$ reproducibility standard deviation plotted against $\log_{10}$ concentration estimated as the mean result. On such a plot, if A and B are constants, any functional relationship of the form $\sigma = Ac^B$ appears as a straight line of slope B. On the plot nearly all of the points fall below the Horwitz function. The trend of the data seems to be linear and, estimated by a robust procedure, shows a slope of unity on the plot and corresponds with a relationship $\sigma_R = 0.22c$. The scatter of points around the trend line is (three outliers aside) about that expected for estimating σ_R from a small number of data. This line intersects the Horwitz function at a concentration of $10^{-6.92}$, about 1.2×10^{-7} or 120 ppb.

There are two main conclusions to be drawn from these findings. First, the deviation from the Horwitz function is more marked in the current data than in the 1997 study. It is not clear that this trend towards better precision has stabilised, although that would be a reasonable assumption for the moment. Second, the deviations can be represented well by a simple generalisation relating precision with concentration, apparently without lack of fit apart from a few outliers. The generalisation is a

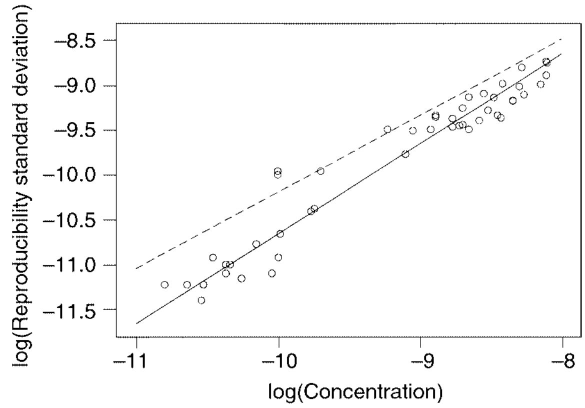

Fig. 1 Results from recent collaborative trials of methods for the determination of mycotoxins, showing the trend of the data (solid line) and the Horwitz function (dashed line).

DOI: 10.1039/b000282h

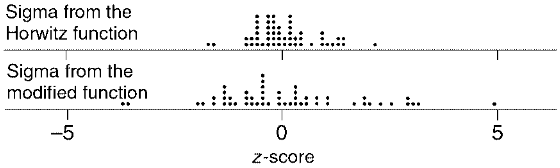

Fig. 2 z-Scores for aflatoxin M_1, calculated from results in FAPAS Round 0423, by using sigma values from both the Horwitz function and the modified function.

better guide to true behaviour than individual results, because errors would be smaller.

The following function is therefore suggested as a contemporary model for reproducibility standard deviation:

$$\sigma = \begin{cases} 0.22c & \text{if } c < 1.2\times10^{-7} \\ 0.02c^{0.8495} & \text{if } 1.2\times10^{-7} \le c \le 0.138 \\ 0.01c^{0.5} & \text{if } c > 0.138 \end{cases}$$

This function could be used as a fitness for purpose criterion where appropriate, subject to review when further data have accumulated. As an example of this suggested use, Fig. 2 represents the *z*-scores calculated from the equation $z = (x - x_{ass})/\sigma$, where x is the participant's result, x_{ass} is the assigned value, and σ-values are derived from both the original Horwitz function and the modified function. (Two extreme outliers are not represented on the plots.) The results were taken from a proficiency test (FAPAS Round 0423) requiring the determination of aflatoxin M_1 in milk. At the assigned value of 0.572 ppb, σ-values derived from the Horwitz and modified functions were 0.281 and 0.126 ppb, respectively. (For comparison, a robust standard deviation of the participants' results was 0.19 ppb.) *z*-Scores derived from the Horwitz function suggest that all but one of the results were comfortably in the 'satisfactory' class. The modified function provided *z*-scores of which, more realistically, about 81% were 'satisfactory'.

References

1 W. Horwitz, L. R. Kamps and K. W. Boyer, *J. Assoc. Off. Anal. Chem.*, 1980, **63**, 1344.
2 M. Thompson and P. J. Lowthian, *J. AOAC Int.*, 1997, **80**, 676.
3 J. T. Peeler, W. Horwitz and R. Albert, *J. Assoc. Off. Anal. Chem.*, 1989, **72**, 784.
4 *Protocol for the Food Analysis Performance Assessment Scheme (FAPAS)*, FAPAS Secretariat, CSL, Sand Hutton, York YO41 1LZ, UK, 5th edn., 1997.
5 M. Thompson, P. Potts, P. Webb and J. Kane, *Analyst*, 1997, **122**, 1249.
6 M. Thompson, *Analyst*, 1999, **124**, 991.
7 W. Horwitz and R. Albert, *J. AOAC Int.*, 1996, **79**, 589.

Paper b000282h