

amc technical brief

Editor: M Thompson AMC Technical Brief No.17 July 2004 © Royal Society of Chemistry 2004

The amazing Horwitz function

Collaborative trials

For many years Dr William Horwitz has been well known as an advocate of the collaborative trial or, using the more correct IUPAC terminology, the interlaboratory method performance study. In collaborative trials, the organiser distributes a duplicated set of test materials to the participant laboratories, which analyse them blind by a strictly defined method. The results are returned to the organiser, who calculates the estimates (s_r and s_R) of the repeatability and reproducibility (between laboratory) standard deviations. These statistics are taken as measures of the performance of the analytical method. Thousands of analytical methods (mostly in food analysis) have been subjected to a collaborative trial and Bill Horwitz made a close study of the results.

The Horwitz 'trumpet'

With so many results to hand, he noticed a striking pattern in the relative standard deviations. As the concentration of the analyte decreased over two orders of magnitude, the relative standard deviation of reproducibility (RSDR) increased by a factor of two. So at 100% concentration of analyte the RSDR was about 2%, at 1% the RSDR was about 4%, and at 0.01% (100 ppm) the RSDR was about 8%. This pattern persisted at least down to sub-ppm levels. These findings gave rise to the famous 'Horwitz Trumpet'¹, which depicts the relationship expressed as a two-sided one-sigma confidence interval (Figure 1).

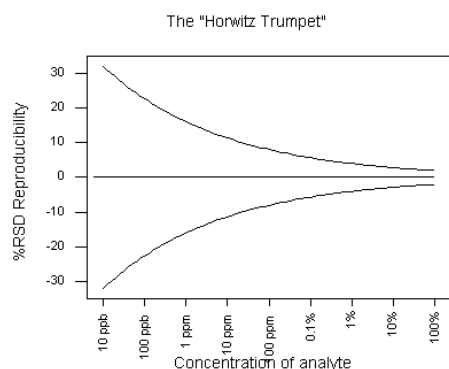


Fig 1. The "Horwitz Trumpet". Purists may object to negative RSD values—the lines are best regarded as confidence boundaries.

A remarkable aspect of this was that the results of collaborative trials seemed to obey this law regardless of the nature of the analyte and the test material, or the physical principle underlying the measurement method.

Moreover, the precision did not improve with time, despite the enormous strides in analytical technology: counter-intuitively, it was found that collaborative trials conducted in the 1920s gave results falling on the same curve as those conducted in 1990s.

Collaborative trial data

Of course, the results from all of these RSDR values did not all fall exactly on the implied mathematical line. There are a number of obvious reasons for that. First each value of RSDR was estimated from small samples of results (the typical 10-20 participants is 'small' by statistical standards) and had a correspondingly large standard error. An estimated RSDR could easily vary by $\pm 30\%$ relative. This factor alone accounts for about a half of the scatter around the mathematical line. Second, RSDR values vary somewhat *within* a single method, especially at concentrations less than about 50 times the detection limit. Finally, some methods have inherently higher between-laboratory precision than others by a small degree. Nevertheless, when this large dataset was considered as a whole, the median trend was extraordinarily close to Bill Horwitz's very simple law.

The mathematical form of the function

The functional form of the Horwitz relationship is more easily perceived if the traditional trumpet is replaced by the mathematically equivalent relationship between predicted reproducibility standard deviation σ_H and concentration c , namely

$$\sigma_H = 0.02c^{0.8495}$$

or, in logarithmic form, the linear equation

$$\log_{10} \sigma_H = 0.8495 \log_{10} c - 1.6990$$

where σ_H and c are mass ratios. This is a simple power law but with a very strange exponent of 0.8495 ($= 1 - \log_{10} 2/2$).

The linearised Horwitz function as expressed above suggests a useful way to look at analytical systems empirically. Applied to Horwitz's compilations of collaborative trial data up to 1996 (over 4000 results), it shows that the function is slightly pessimistic at high concentrations (above 10% m/m) and more noticeably so at low trace concentrations. Below about 10 ppb, we see a tendency for an invariant RSD of about 20-25%. This is because a method with a higher RSD would hardly provide any useful quantitative information: results would tend to be below the detection limit.²

Moreover, the empirical exponent for the region between 10 ppb and 10% m/m is not exactly as given in the Horwitz function but closer to 0.824. But despite these small deviations, the Horwitz function is still impressive, as can be seen in Figure 2.

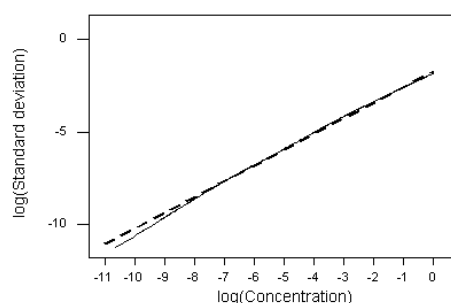


Fig 2. Trend of data from collaborative trials (shown as a lowest fit, solid line) compared with the Horwitz function (dashed line). The systematic deviation below about 10 ppb is apparent. Units are mass fractions (e.g., 1% = 0.01, 1 ppm = 10^{-6}).

Compilations of data from proficiency tests show similar functions. For example, early data from FAPAS (a foodstuffs proficiency test scheme) gave an excellent fit to a Horwitz-style function³, of the form $\sigma = 0.023c^{0.826}$. This indicates a slightly lower precision than collaborative trials, but that is hardly surprising: proficiency test data include uncertainty due to variation in analytical method, obviously not present in collaborative trials.

A benchmark

The Horwitz function is now widely used as a benchmark for the performance of analytical methods, via a measure called the 'Horrat' which is defined as $\text{Horrat} = s_R / \sigma_H$.

An analytical method that during collaborative trial gives Horrats that are substantially worse than unity is regarded as flawed and requiring improvement or rejection. The function also became a benchmark for performance in some important proficiency tests, by equating the Horwitz reproducibility standard deviation with the sigma-value used to calculate z-scores. The rationale for this latter decision is that the Horwitz law describes a fitness for purpose criterion in many application areas.⁴

Generality

While it is thus widely useful, it would be unreasonable to expect the Horwitz function to cover every contingency. Applications where very high accuracy is required readily spring to mind, and there is evidence that laboratories can fulfil the enhanced requirement. Never the less, the function often seems applicable to areas other than food analysis. A startling instance of this generality comes from a recent interlaboratory study of the analysis of a volcanic glass by microprobe methods (laser ablation-inductively coupled plasma mass spectrometry and electron probe).⁵ This test

material, and analytical method employed, could hardly be more remote from the materials and methods that provided the original Horwitz data, especially as the mass of material analysed in LA-ICP-MS is only a few microgrammes. The data (Figure 3) conform with the Horwitz function to a remarkable degree.

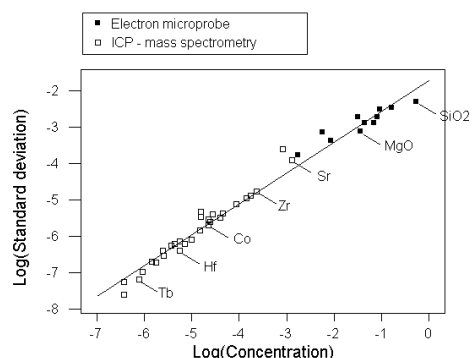


Fig 3. Reproducibility (between laboratory) standard deviation vs. concentration obtained by microprobe methods (points), compared with the Horwitz function (line). Each point is a different element.

An explanation of the function?

As well as being useful, the Horwitz Trumpet is a feature of considerable theoretical interest. It is hard to avoid the assumption that a simple mathematical law that describes the behaviour of large numbers of methods over at least six orders of magnitude of analyte concentration must have some inherent meaning and deserves serious consideration. So far, though, nobody has managed to explain the strange empirical exponent from basic principles, although several people have made conjectures. Are we seeing the manifestation of a physical law here, or is there a psychological basis, perhaps to do with our perception of fitness for purpose? There is a sure-fire paper in *Nature* waiting for somebody!

Biographical

Bill Horwitz has now retired after 57 years [*sic*] with the FDA. He was given a unique personal award by AOAC International in 1995, and the Boyle Medal by the Analytical Division of the RSC in 2000. See *Chemistry International*, 2000, **22** (No 6 November) for further biographical details.

References

1. W Horwitz, L R Kamps, K W Boyer, *J Assoc Off Anal Chem*, 1980, **63**, 1344.
2. M Thompson and P J Lowthian, *JAOAC Internat*, 1997, **80**, 676-679.
3. M Thompson and P J Lowthian, *Analyst*, **120**, 271-272.
4. M Thompson, *Analyst*, 1999, **124**, 991.
5. P J Potts, M Thompson, and S Wilson, *Geostandards Newsletter*, 2002, **26**, 197-235.

This Technical Brief was prepared by the Statistical Subcommittee of the AMC.