The Limit of Measurement of the Polychlorinated Contaminants (Biphenyls, Dioxins, and Furans)

William Horwitz and Richard Albert
Center for Food Safety and Applied Nutrition, HFS-500, Food and Drug Administration, Washington, DC 20204, USA

1. Introduction

The polychlorinated contaminants (PCCs), i.e., biphenyls, dioxins, and furans, are a class of chemical compounds routinely determined below ng/g (ppb, 10^{-9}) levels, and sometimes, even below pg/g (pptr, 10^{-12}) levels. A recent notice from the U.S. Food and Drug Administration[1] refers to a survey of bleached food-contact paper products to determine the degree of compliance with a voluntary specification of 2 pptr of 2,3,7,8-tetrachlorodibenzodioxin.

The purpose of this paper is to continue a series of examinations of interlaboratory precision of various specialties of analytical chemistry, all recalculated on a uniform basis by the harmonized IUPAC-1987 protocol[2]. Fields that have been covered to date include pesticide formulations; pharmaceutical preparations; major constituents, major elements, and mycotoxins in foods; geological and standard reference materials (SRMs); trace elements; and dairy products. These areas cover the concentration levels from 100% to ppb. The present review is intended to extend the concentration region examined to pptr and below.

2. The Database

The performance parameters of the 34 available interlaboratory studies of the determination of PCCs (biphenyls, dioxins, and furans) were recalculated on a uniform basis by the IUPAC-1987 harmonized protocol[2]. The database contained 1052 test samples, each from a minimum of 4 laboratories, 56 analytes, 19 matrices, and 2 types of detectors (electron capture and mass spectrometers). The mean and the within-laboratory and among-laboratories relative standard deviations, RSD_r and RSD_R, respectively, were calculated for each data set. The expected RSD_R, was also calculated from the Horwitz formula[3]:

$$RSD_R\ (\%) = 2^{(1-0.5 \log_{10} C)} \approx 2C^{-0.1505},$$

A ratio, designated as HORRAT, of the RSD_R actually found to the RSD_R calculated from the experimentally found concentration, C, was compared with those HORRATs determined for the various other specialties of analytical chemistry. About a dozen previous papers from this laboratory attest to the validity of the general relationship between RSD_R and concentration down to about parts per US billion ($C = 10^{-9}$) levels[3].

3. Results

On the basis of historical data, HORRAT, which removes the predominant effect of concentration, should center about 1.0. Method performance studies providing HORRAT values >2.0 are considered unsatisfactory. Laboratories that provide values far removed from a consensus value that shows acceptable precision should examine their operations to determine the cause of the discrepancies.

In general, the HORRAT values for the biphenyls at approximately the ppm (10^{-6}) level follow the historical pattern, but the HORRAT values for the dioxins and furans at the ppb (10^{-9}) and pptr (10^{-12}) levels are considerably better (i.e., lower). Nevertheless, even though the variabilities found are lower than anticipated, they are still very high at these low concentrations. As shown in Table 1, a considerable number of the low concentration data have RSD_R values >50%, a variability for which statistical control begins to be lost and where quantitative estimates start to lose validity. A process is said to be in statistical control when it produces a repeatable, consistent pattern of data. This point of losing statistical control appears to be at a concentration level of about 0.1-1 ppb. This loss of control is also characterized by the erratic appearance (1) of zeros, (2) of "less thans," and (3) of negative values, as lower concentration levels are measured. However, this evidence is invisible to the individual chemist because a single result from a test sample is unequivocal. Only when it is compared with the result from another operator is it seen that the results do not agree. From the point of view of the parent population, however, the disparate results from different chemists in different laboratories are all members of the same parent population that merely exhibits an inherently high variability.

Table 1. Fraction of RSD_R values greater than specified values, by chemical groups.

Chemical Group	Fraction of RSD_R values		
	>30%	>50%	>100%
Biphenyls	0.30	0.10	0
Dioxins	0.50	0.20	0.10
Furans	0.50	0.45	0.15

4. Discussion

The explanation suggested for the superior performance found is that supplying common reference calibrating solutions, as was done in many of these studies, does not reflect realistic operating conditions. Furthermore, the ability to repeat, discuss, and reassess aberrant reported values also results in underestimating the true RSD_R.

The use of common, identical calibration solutions by all participants does not permit the systematic errors introduced in the preparation of the individual calibration standards to become the random error of the group of participating laboratories. Only if all laboratories subsequently use common calibrating standards does RSD_R obtained from such studies reflect routine precision. If routine practice requires that each laboratory prepare its own calibration solution, this practice must also be required in any interlaboratory study designed to determine the expected RSD_R of the group.

Other factors that may introduce variability are (1) calibration standards that do not correspond to their stated values; (2) instability and drift of the detection systems, although the more modern instrumentation appears to be considerably more stable; (3) operational details such as failure to extract the contaminants completely from the matrix; and (4) failure to perform adequate quality control operations and to conform to good laboratory practices.

5. The Practical Limits of Measurement

Although a number of interlaboratory studies achieved an overall objective of an RSD_R of 20-30%, such values as an overall mean imply that individual data sets may be roughly twice as large, i.e., 40-60%, for 95% confidence. Analytical chemists are not ordinarily aware of the actual variability that large two-digit RSD_R values imply. For example, if we assume a dataset from 6 laboratories with a mean of 0.5 ppb, an RSD_R of 50%, and a normal distribution, a surrogate set of data with these statistics is 0.16, 0.30, 0.43, 0.57, 0.70, and 0.83 ppb, which has a range factor of 5. If the data set were lognormal, the range would be even broader. The only feasible method of reducing variability is through replication. This approach is limited by the cost of the analysis, which is now about $1000 per assay.

An even more important limiting factor is the appearance of nonnumerical, "less than" values, false positive and false negative values, and nonfunctional zeros (indistinguishable from zero but not necessarily absent), and the substitution of arbitrary values for low-level measurements (i.e., 1/2 an undefined detection limit and zero are common). This factor too is invisible to individual operators. Only when results are available from a number of laboratories for the same test material is the evidence apparent that the system lacks control and that it has reached or exceeded its limit.

Figure 1 shows a statistical basis for the appearance of false negative results. For a normal distribution with a mean, $\bar{x}$, say, of 100 and a standard deviation, s, of 100 (RSD_R = 100%), $\bar{x} - s = 0$. But 1s contains only about 34% of the left half of the distribution, leaving 16% of the values below zero (the false negatives). If the true $\bar{x}$ is exactly zero (a true blank), half of the found values should be below zero; otherwise a biased positive value results. If the distribution is lognormal, negative values do not exist by definition. We know negative values can exist (e.g., random excursions of the blank are greater than the random excursions of the measurement) so a lognormal distribution is not a physically appropriate model. The data might be modeled by other nonsymmetrical distributions, but then a calculated RSD_R would be greater than the true RSD_R. With multiple zeros the distribution is at least bimodal and consequently uninterpretable, unless expensive data are discarded.

We can agree that 16% false negatives is too great a number to tolerate, but what is acceptable? Perhaps Table 2 may be of assistance. If we take as our default an acceptable limit of 5% false negatives, and translate RSD_R to concentration through the Horwitz curve, the limit of measurement is somewhere between 0.1 and 1 ppb. What is truly interesting about this conclusion is that we found the same region to be our limit of measurement when we examined the interlaboratory studies of the determination of mycotoxins in foods[4]. The predominant assay methods in the mycotoxin case are thin-layer and high-performance liquid chromatography.

Table 2. Expected percentage of false negative values as a function of concentration, assuming a normal distribution and the Horwitz curve.

Concentration	RSD_R, %	False negatives, %
1×10^{-12} (1 pptr)	130	22
1×10^{-11}	90	13
1×10^{-10}	64	5.9
1×10^{-9} (1 ppb)	45	1.3
1×10^{-8}	32	0.09
1×10^{-7}	23	0.001
1×10^{-6} (1 ppm)	16	0.000

An objection can be raised that the variability exhibited by these results far exceeds the variability experienced by most individual laboratories. This is true, but the implied comparison is invalid. The experience of most laboratories involves primarily their own work, which generates a within-laboratory variability that is incorrectly used as the basis for comparison with the among-laboratories variabilities calculated from interlaboratory studies. This within-laboratory variability should be compared only with the within-laboratory variabilities of other laboratories performing comparable analyses. Although their variabilities may be compared, their absolute magnitudes may not be compared unless the data are linked to a certified reference material. Even then, such traceability will be confounded by the RSD_R.

6. Conclusion

The interlaboratory studies on PCCs as well as on aflatoxins at the ng/g (ppb) and pg/g (pptr) levels indicate that the conservative assumption of the exponential, historical Horwitz curve breaks down at the current limits of analytical chemistry. This model collapses when the analytical measurements are no longer reproducible, a point where RSD_R exceeds about 50%. When methods are operated near their limits, the negative controls (commodity or field blanks) show false positive values and at $RSD_R \approx 50\%$, the low-level positive controls begin to show intolerable levels of false negative or "less than" values. This phenomenon occurs at concentrations of about 10^{-9} (ng/g; ppb). Much of the superb labor that has gone into the

statistically sound theory of limits of measurement fades into the chaos of uncertainty, as these limits are approached. Tomas Hirschfeld[5] pointed out 20 years ago that we may be setting goals that involve "beating the Heisenberg's principle, Shannon's limit, or the second law of thermodynamics, all of which can take rather surprising forms under extreme conditions." Unfortunately, the inevitable uncertainty in low-level concentration estimates is invisible to those who do not look at equally legitimate results from outside their own laboratory.

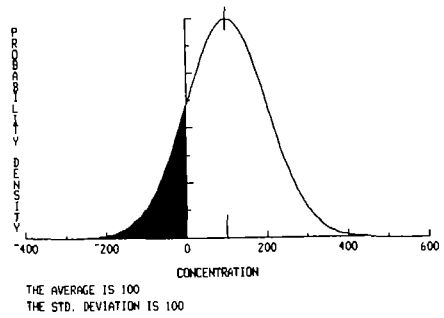

Figure 1. A normal distribution with a mean of 100 (say pptr) and a standard deviation of 100 (pptr). The shaded area, which contains the false negatives, comprises about 16% of the total area of the distribution.

7. References

1) Food and Drug Admin. (Apr. 12, 1994) Fed. Regist. 59, 17384-17389
2) Horwitz, W. (1988) Pure Appl. Chem. 60, 855-864
3) Peeler, J.T., Horwitz, W., and Albert, R. (1989) J. Assoc. Off. Anal. Chem. 72, 784-806
4) Horwitz, W., Albert, R., and Nesheim, S. (1993) J. Assoc. Off. Anal. Chem. 76, 461-491
5) Hirschfeld, T. (1976) Anal. Chem. 48, 16A-31A