# HOW TO DEAL WITH PROFICIENCY TEST RESULTS FOR DIOXINS, FURANS AND D-L PCBs AT ppt AND sub-ppt LEVELS

Gauthier Eppe[1], Edwin De Pauw [1]

[1]CART-University of Liege, Liege

## Introduction

Proficiency testing (PT) is a useful method for assessing the accuracy of laboratories performing particular measurements. In analytical chemistry, most of PT schemes use the z-score system as recommended in the Harmonized Protocol[1]. The participant's results are converted into a z-score by using the following equation: $z = (x-X)/\sigma$. Where x is the participant's result, X is the assigned value and $\sigma$ is the target standard deviation. Thus, the z-score obtained in a PT depends on the $\sigma$ value. Different ways to assess $\sigma$ can be used: a relationship between $\sigma$ and the concentration is generally applied but the setting of a target standard deviation is also frequently used. In fact, $\sigma$ should represent the uncertainty on the measurement that the analyst estimates to be tolerable to fit for purpose. The most famous function to assess $\sigma$ in PT is the Horwitz function, $\sigma_H = 0.02c^{0.8495}$, where c is expressed in dimensionless mass ratios e.g. $1\text{ppb} \equiv 10^{-9}$. This equation provides an estimate of the inter-laboratory reproducibility precision. However, the relationship is not appropriate at low levels (below 0.1ppm) because it gives unacceptable high $\sigma$ values. More recently, Thompson et al[2] presented a modification of the Horwitz function based on collaborative trials involving analytes at levels below 10 ppb. The Thompson function is $\sigma = 0.22c$.

As already mentioned, another possibility to calculate the z-score is to set the reproducibility standard deviation. Generally, the establishment of criteria for $\sigma$ is based on legislation recommendations, on the analyst background for a target compound or even $\sigma$ can describe the end user's requirements.

For dioxins, furans (PCDD/Fs) and dioxin-like (D-L) PCBs in foodstuffs, levels are varying from sub-ppt levels for some PCDD/Fs congeners to ppb levels for some D-L PCBs depending on the matrix analyzed. Most of the PT organized in that field set $\sigma$ values to assess the laboratory performances using the z-score. For instance, the international PT called FOOD, organized by the Norwegian Institute of Public Health, established a $\sigma$ value expressed in RSD at 20% for z-score TEQ assessment. But how to deal with the performances for all the individual congeners varying from sub-ppt levels to ppb levels. We developed recently[3,4] a function for PCDD/Fs and D-L PCBs in the range from $5\text{x}10^{-14} \leq c \leq 10^{-8}$. The relationship has been estimated from a robust statistical evaluation of data from a collaborative trial for PCDD/Fs and D-L PCBs in animal feedingstuffs[3] by the reference HRGC/HRMS method. The function is $\sigma = 0.0169 + 0.119c$ where c is expressed in ppt units. The function can be applied in the range between $0.05 \text{ ppt} \leq c \leq 10000 \text{ ppt}$ (see figure 1). Table 1 summarizes different values of RSD (%) that can be obtained using the different functions for the dioxin range levels in foodstuffs.

**Table 1:** RSD (%) calculated using different functions at different levels.

| Levels ppt | Thompson RSD (%) | Horwitz RSD (%) | Function presented here RSD (%) | Pre-established value RSD (%) |
|---|---|---|---|---|
| 0.05 | 22.0 | 200.8 | 45.7 | 20.0 |
| 0.1 | 22.0 | 180.9 | 28.8 | 20.0 |
| 1 | 22.0 | 127.9 | 13.6 | 20.0 |
| 10 | 22.0 | 90.5 | 12.1 | 20.0 |
| 100 | 22.0 | 64.0 | 11.9 | 20.0 |

In the present study, all the results obtained from our laboratory during a Proficiency Test (FOOD) were treated according to our function and compared with the results obtained using a pre-established $\sigma$ value.

## Results and Discussion

The function ($\sigma = 0.0169 + 0.119c$) has been converted into RSD (%) = $1.69c^{-1} + 11.9$ and represented in figure 1. The horizontal dashes line represents a pre-established recommended RSD of 20%. The relationship and the 20% RSD line intersect at 0.2 ppt. The figure indicates for congeners at levels below 0.2 ppt that a pre-established RSD value of 20% is certainly too severe and can lead to high z-score. On the other hand, for congeners at levels beyond 0.2 ppt and especially for congeners close to ppb levels (D-L PCBs), the pre-established criterion underestimates the z-score. In the range below 0.1 ppt and close to LOD/LOQ of the HRGC/HRMS reference method, it is normal to select $\sigma$ values corresponding to RSD values higher than 50% for z-score assessment.

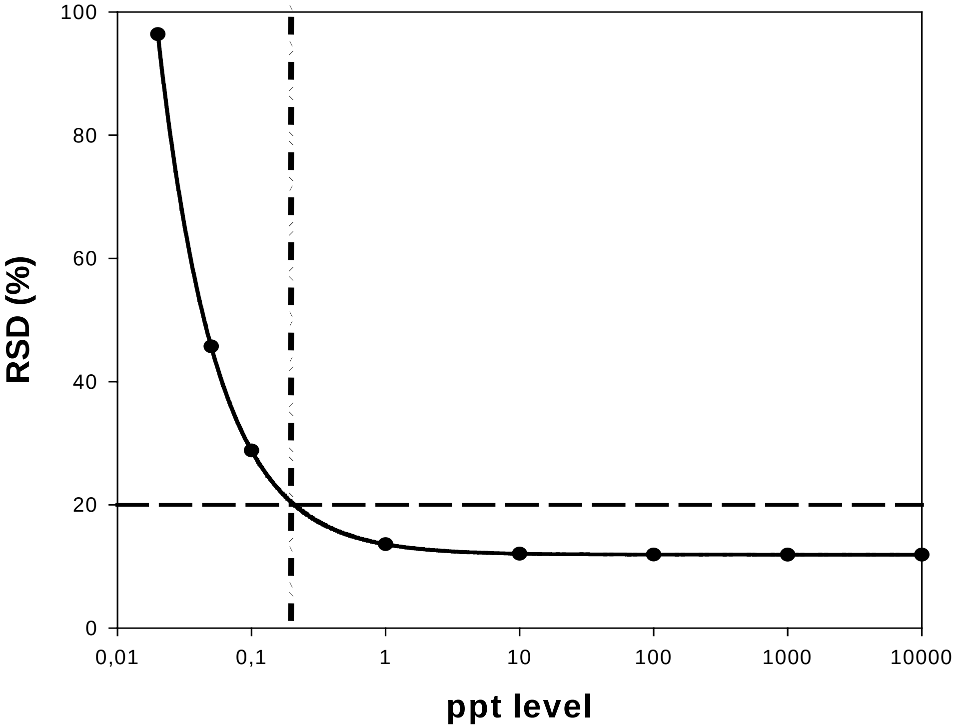


**Figure 1** : RSD (%) versus dioxins, furans and D-L PCBs congeners ppt levels.

The international PT FOOD organized in 2003 supplied three different matrices for PCDD/Fs and D-L PCBs measurement.

Turkey, salmon and cheese were shipped to participants. The statistical treatment of data provided together assigned values for all the congeners after removal of obvious outliers and TEQ values. The laboratories can evaluate their performances in TEQ by z-scores calculated with a pre-established RSD of 20%. In fresh weight, all the TEQ z-scores obtained by our laboratory were between ± 1 for the three matrices. This indicates good performances in TEQ but what about the performances for each individual congener? The statistical treatment of data showed that 4 obvious outliers were detected for two congeners and for those ones, investigations can be easily triggered to find out the reasons. They have been removed from the statistical evaluation. For the others congeners, we provided, in fresh weight, a total of 79 results for PCDD/Fs and D-L PCBs for the three matrices. Among the matrices, levels in turkey and cheese were very low and many congeners were close to LOD/LOQ. In fact, 84% of PCDD/Fs consensus means values in fresh weight are below the cut off value of 0.2 ppt (see figure 1) while 86% of D-L PCBs are above the 0.2 ppt value. In order to evaluate our performances, the function presented in figure 1 as well as the criterion of 20% were applied to calculate the z-scores. Figure 2A shows the distribution of the 79 z-scores using the model while figure 2B shows the distribution obtained using the RSD set at 20%.

If z-scores between ± 2 are considered as satisfactory, then the z-scores derived from our function (figure 2A) suggested that 94% of our results fall within the limits. Two values are questionable in the range between +2 and +3 and three values are not satisfactory in the range between ± 3 and ± 4. It is also important to note that the z-score are normally distributed indicating no systematic bias or error. There is no tendency for our method to under- or overestimate the PCDD/Fs and the D-L PCBs levels. The mean z-score value for the 79 z-scores calculated is 0.2.

Figure 2B represents the distribution of the 79 z-scores calculated with a pre-established criterion of RSD = 20%. As can be seen, the percentage of satisfactory z-scores (± 2) falls to 72%. The distribution of the z-scores is also wider (± 7). The distribution here is just a little asymmetric with a higher proportion of z-scores between 0 and 1.5 even if the mean z-scores value merges with the target 0 value.

If normality can be assumed, the superposition of the two distributions is graphically represented in figure 3. The dashed curve represents the z-score distribution obtained using our model and the continuous curve the distribution calculated with the 20% RSD. This clearly shows that the way to deal with the same results can lead to different interpretations and conclusions.

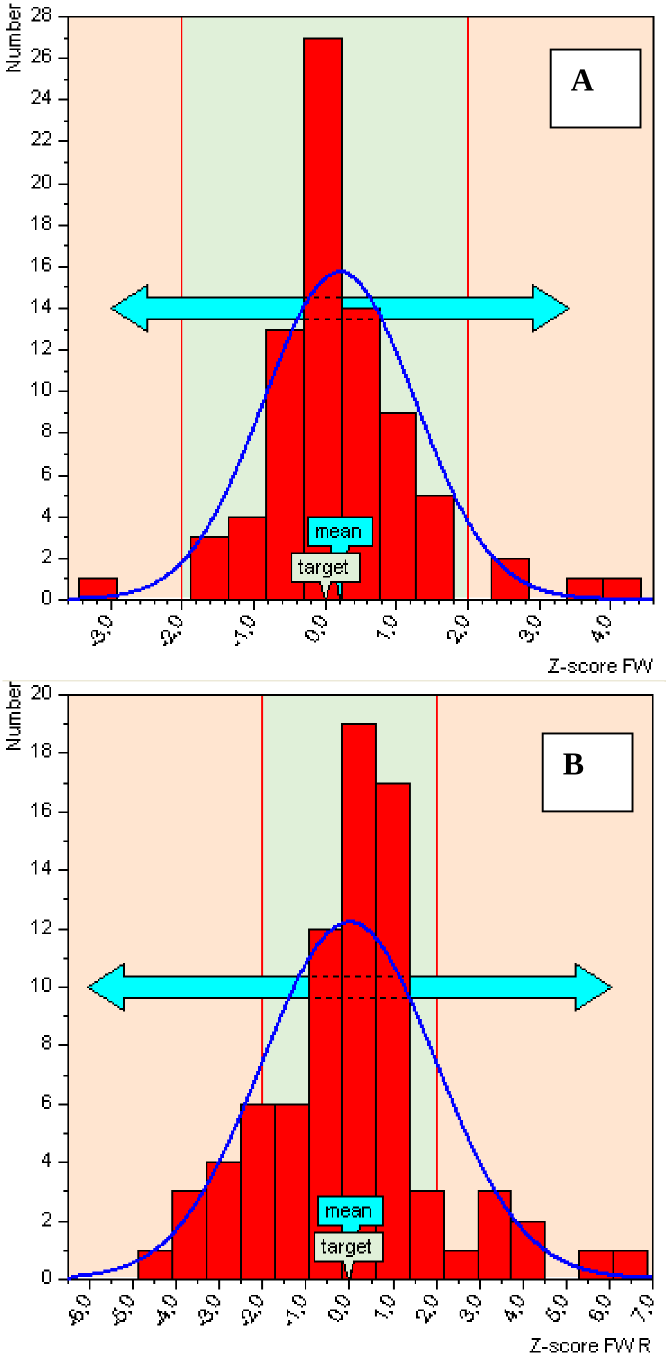


Figure 2: Distribution of the 79 z-scores obtained in the PT FOOD 2003 using (A) the function developed by Eppe et al. and (B) using a pre-established (RSD 20%) σ value.

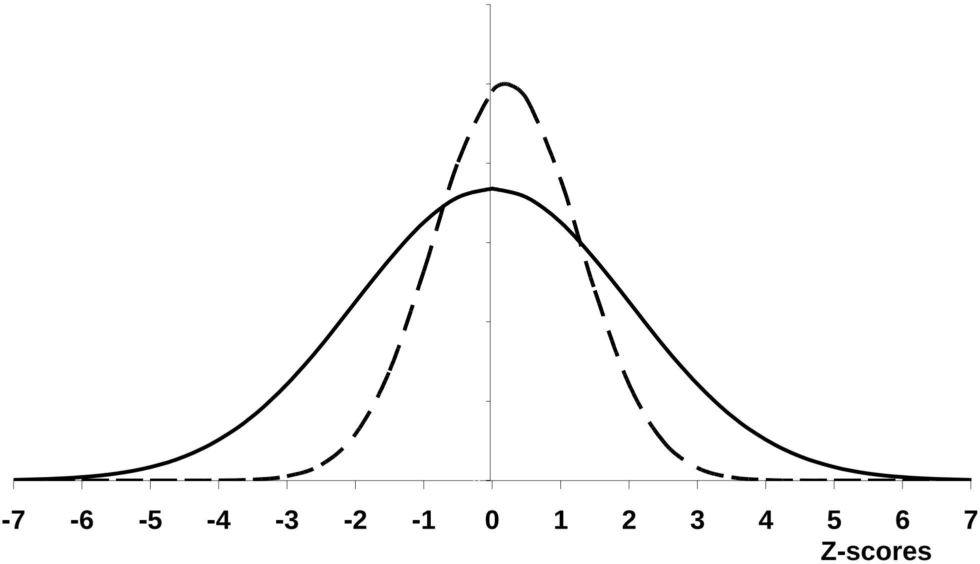


**Figure 3**: Superposition of the two distributions A and B.

## Conclusions

There are two main conclusions to be drawn from these findings. First, the 20% RSD used to calculate the z-score is certainly well appropriate to evaluate the performances in TEQ. For individual congener's results, this criterion should probably be applied in a restricted range of concentrations and it is not realistic to apply it below 0.1 ppt ; this is particularly true when the laboratory's performances have to be assessed in foodstuffs matrices close to background contamination. Second, the model we developed seems to be more appropriate when dealing with performances assessment at background levels. It can cover 5 orders of magnitude for PCDD/Fs and D-L PCBs levels. The function provided here z-scores of which about 94% were satisfactory and it is realistic to consider that a result at 0.1 ppt is satisfactory between ± 2 even if it corresponds to roughly a difference of 60% or more between the laboratory's result and the assigned mean value. On the other hand, at higher levels, our model demands a smaller standard deviation than 20% for z-score assessment

## Reference

1. Thompson M., Wood R., Pure Appl. Chem., 1993, 65, 2123
2. Thompson M., The analyst, 2000, 125, 38386
3. Eppe G., Cofino W.P., De Pauw E., Anal. Chim. Acta, in press
4. Eppe G., Maghuin-Rogister. G., De Pauw E., Anal. Chim. Acta, in press