DETERMINATION OF CONSENSUS VALUES IN INTERLABORATORY COMPARISON STUDIES ON LEVELS OF DIOXINS IN LOW CONTAMINATED FOOD SAMPLES

Cathrine Thomsen[1], Line Småstuen Haug[1], Tove Nicolaysen[1], Gunilla Lindström[2], Georg Becher[1]

[1]Norwegian Institute of Public Health, P.O.Box 4404 Nydalen, 0403 Oslo, Norway
[2]Man-Technology-Environment Research Centre, Örebro University, 701 82 Örebro, Sweden

Introduction

Recent episodes of accidental contamination of commercial food with PCDD/Fs and PCBs, introduction of maximum contaminant levels for PCDD/Fs in food by the EU, as well as recurrent risk assessments of human health effects of dioxin-like compounds have created a world-wide demand for highly reliable and comparable data on dioxins in food. It is well recognised that there can be much variation between laboratories' results despite the use of validated methods of analysis and wide-spread accreditation by recognised bodies. In recognition of the need to have high analytical quality in trace analysis of dioxin-like compounds, various interlaboratory comparison exercises or proficiency testing schemes have been initiated. However, so far only few studies had been performed on consumable food with background contamination. Starting in 2000, the Norwegian Institute of Public Health (NIPH) has organised three world-wide interlaboratory comparison exercises ("Dioxins in Food") on determination of PCDD/Fs and dioxin-like PCBs in three frequently consumed foods for each round[1-3].

Recent reports from several countries show that levels of dioxin-like compounds have declined considerably during the last decade, e.g., by 40-75 % in the Netherlands from 1991 to 1999[4]. Nevertheless, 8 % of the Dutch population is exposed to intake levels above the Tolerable Daily Intake (TDI) of 2 pg TEQ/kg body weight, as recently derived by the EU. When determining intake of dioxin-like compounds from various low contaminated but high consumption foods and performing cumulative exposure assessments, it is of course better to work with measurable, though low levels, than to handle the uncertainties associated with "less than" values. This implies that there is an increasing requirement to the laboratories to improve sensitivity of the analytical methods in order to be able to report measured values in foods of decreasing contamination with dioxin-like compounds.

In the context of the interlaboratory comparison "Dioxins in Food 2001" organised by the NIPH, the question arose of how to treat non-detected congeners (NDs) when determining the consensus from participant laboratories and calculating the toxic equivalent levels. It was clear that the relative standard deviation of the consensus mean could be unacceptably high and spoil the trueness of the assigned value when including data of insufficient quality.

We here describe the effects of different methods for treatment of NDs and for exclusion of reported PCDD/F concentrations from the statistical calculation of the consensus TEQ value of a low contaminated beef sample analysed in the "Dioxins in Food 2001" exercise.

Methods

The design of the interlaboratory comparison "Dioxins in Food 2001" is described elsewhere[2]. The statistical analysis involved calculations of congener-by-congener mean (mean1) and standard deviation (SD1) using the reported detection limit as concentration for NDs. Obvious outliers were defined as values reported outside a range of ±4 x SD1, which were removed from the data set. The

mean (mean2) and standard deviation (SD2) were re-calculated from the remaining data and now values outside a range of 2 x SD2 were removed. The mean of this final data set was called consensus value and the standard deviation SD3.

The TEQ consensus was calculated from the individual consensus values of the PCDD/Fs congeners using WHO_{98} TEFs. Z-scores were determined for each laboratory using 20% as a target value for the standard deviation: $Z= (TEQ_{laboratory} – TEQ_{consensus})/ 0.2 \times TEQ_{consensus}$.

Problems were encountered for the low contaminated beef sample (14 % fat), which showed a very skewed Z-score distribution, a large difference between mean and median TEQ and a large relative standard deviation (RSD) of the mean (see Table 2). This indicated that the consensus mean was not a good estimate of the true value. We therefore re-calculated the mean and median TEQ consensus using the concentrations of the five PCDD/F congeners giving the largest contribution to TEQ, i.e. 2,3,7,8-TCDD, 1,2,3,7,8-PeCDD, 2,3,7,8-TCDF, 2,3,4,7,8-PeCDF and 1,2,3,4,7,8-HxCDF, together 80 % of PCDD/F TEQ. The different methods evaluated for calculation of the consensus value are presented in Table 1. Outliers were excluded in all methods except method A.

Table 1. Different methods evaluated for calculation of the consensus values. NDs were assigned the value of the reported detection limit.

Method	In calculation of the consensus value -
A	- all reported values were included
B	- outliers were excluded[a]
C	- all NDs were excluded
D	- all NDs above 150% of the median (all values) were excluded
E	- all laboratories with Z-score >5 were excluded

[a]This method was used in the “Dioxins in Food 2001” exercise.

Results and Discussion

A selection regarding the statistics of the three samples included in the “Dioxins in Food 2001” study are summarised in Table 2, which clearly expresses the difficulties encountered with the beef sample.

Table 2. Major results from “Dioxins in Food 2001”.

	Consensus mean, TEQ pg/g fw[a]	Consensus median, TEQ pg/g fw[a]	RSD % TEQ	RSD % Range for congeners	Range of the laboratories’ individual TEQ, pg/g fw[a]
Beef	0.084	0.058	32	56-169	0.021-12
Breast milk	0.14	0.14	14	10-105	0.072-4.5
Cod liver	3.1	3.1	13	15-100	0.83-8.1

[a] fw: fresh weight

Besides evaluating internationally the analytical performance of laboratories with respect to dioxin analysis of food, another objective of these interlaboratory comparison studies is to establish a good

estimate for the true content of PCDD/Fs, for the participating laboratories to use in their quality assurance. The consensus mean and median values for beef, resulting from the different calculation methods described above, are presented in Figure 1 and discussed below.

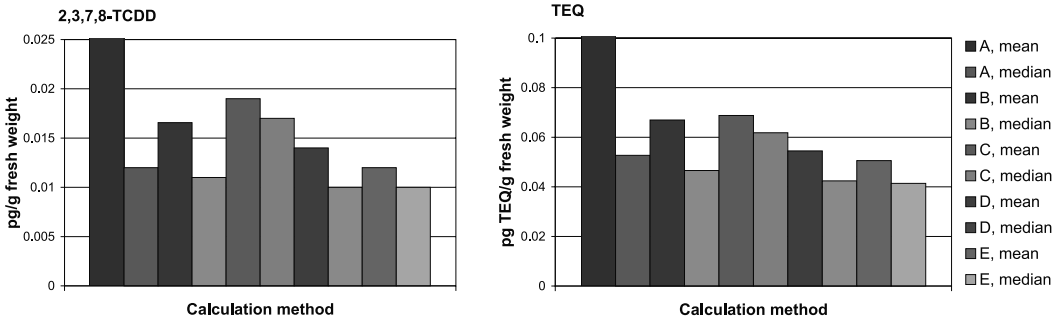

Figure 1. The consensus concentration of 2,3,7,8-TCDD and consensus TEQ (5 congeners) obtained with different methods used for calculation.

A) all reported values were included

This is a simple approach, which expresses the consensus of all the values that were actually reported. The largest difference between mean and median was obtained using this method, and this mean is not a good estimate for the true PCDD/F content. Also, obviously wrong values (outliers), which have large impact on the statistics, should be removed in general.

B) outliers were excluded

This method, which was used in the current study[2], is not subjected to obviously wrong values and expresses the consensus of the values actually reported. However, values reported for NDs from laboratories whose determinations suffer from insufficient sensitivity, will lead to an overestimation of the true value.

C) all NDs were excluded

Any value assigned the NDs is inevitably subjected to a higher uncertainty compared to the concentrations of the detected compounds, it is therefore reasonable to exclude NDs. This method shows the least deviation between the mean and the median consensus value. However, the resulting TEQ consensus value is most probably an overestimate of the true value, due to many low NDs. Furthermore, the consensus value may be based on a very small number of measured values when samples containing low PCDD/F levels are to be assessed.

D) all NDs above 150 % of the median were excluded

By removal of high NDs, congeners that have been determined with insufficient sensitivity have no influence on the consensus value. The low NDs are kept in the data set, because at such low levels, the reported congeners are not *proven* to be present in the sample, they could actually be a result of sample contamination.

E) all laboratories with Z-score >5 were excluded

The rationale for this calculation method is that laboratories with a Z-score above 5 do not meet the requirements of sufficient analytical performance to be incorporated in the consensus. However, a high Z-score may possibly be due to a gross error in only a few congeners important with respect to the TEQ, and is not necessarily representative for the whole analysis. Vice versa, laboratories with acceptable Z-score might have determined these congeners properly, but failed on other congeners with lower TEF.

Using these different approaches, the resulting consensus TEQ for the beef sample ranges from 0.041 to 0.069 pg TEQ/g fresh weight (mean of method A excluded). The corresponding RSD is 19 %, thus the variation in the consensus TEQ calculated with different methods is about the same as the applied target value for the standard deviation used in the Z-score (20 % TEQ).

Concluding remark

After having performed the different calculations on the data set of the beef sample from "Dioxins in Food 2001", we consider the process of removing NDs greater than a certain percentage of the median and further excluding values exceeding ±2SD, as most suitable for establishing a best estimate for the true PCDD/F content (method D). This approach is reasonable with respect to analytical quality criteria, and will in our opinion be fair to the participating laboratories as only single values are excluded. However, a challenge encountered in this approach is to decide which values for the NDs are acceptable for the individual congeners and samples.

References

1. Lindström G., Haug L.S., Nicolaysen T. (2000) *Intercalibration on dioxins in food.* Report 2000:9, National Institute of Public Health, Oslo, Norway, ISBN 82-7364-167-8
2. Becher G., Lindström G., Nicolaysen T., Thomsen C. (2001) *Interlaboratory comparison on dioxins in food.* Report 2001:4, National Institute of Public Health, Oslo, Norway, ISBN 82-7364-178-3
3. Lindström G., Haug L.S., Nicolaysen T., Dybing E. (2002) *Chemosphere* 47, 139
4. Freijer J.I., Hoggerbrugge R., van Klaveren J.D., Tragg W.A., Hoggenboom L.A.P., Liem A.K.D. (2001) RIVM report 639102 022, National Institute for Public Health and the Environment, Bilthoven, The Netherlands