EU-RL PROFICIENCY TESTS ON DETERMINATION OF PCDD/Fs AND PCBs IN FEED AND FOOD – EVALUATION OF DATA AND SCORING OF RESULTS

Kotz A[1*], Hädrich J[1], Wahl K[1], Malisch R[1]

[1] European Union Reference Laboratory (EU-RL) for Dioxins and PCBs in Feed and Food, State Institute for Chemical and Veterinary Analysis (CVUA), Freiburg, Germany

Introduction

The European Union Reference Laboratory (EU-RL) for Dioxins and PCBs in Feed and Food, located at the State Institute for Chemical and Veterinary Analysis (CVUA Freiburg, Germany), organizes interlaboratory studies and proficiency tests (PTs) on the determination of PCDD/Fs and PCBs (dioxin-like PCBs and indicator PCBs) in food and feed matrices for National Reference Laboratories (NRLs) of EU member states regularly twice a year. These interlaboratory studies and PTs are also open for official laboratories of these member states, NRLs from other countries and in certain cases also for commercial laboratories worldwide. Since 2006, 13 interlaboratory studies and proficiency tests covering various food and feed matrices were organized by the EU-RL, some of these in cooperation with the Norwegian Institute of Public Health (2007) and the RIKILT – Institute of Food Safety (2010).

Objective of the interlaboratory tests is to assess the analytical performance of participating laboratories and the interlaboratory comparability of results from analyses of the relevant parameters: 17 PCDD/Fs, 12 dioxin-like PCBs and 6 indicator PCBs. For assessment of the analytical performance, the determination of the assigned value and the scoring of the results are of profound importance and are therefore based on the requirements of ISO/IEC 17043[1], ISO 13528[2] and the IUPAC technical report on proficiency testing[3]. Beyond, the network of EU-RL and NRLs of EU Member States for Dioxins and PCBs in Feed and Food developed additional criteria for an overall assessment of PT test samples and for results from application of bioanalytical screening methods.

Structure of the proficiency tests

EU-RL proficiency tests comprise the determination of PCDD/Fs, dioxin-like PCBs and indicator PCBs in feed and food samples applying at least one of the following methods:

- GC-HRMS-methods for PCDD/Fs and dioxin-like PCBs
- GC-MS/MS (or other alternative methods for GC/HRMS) for PCDD/Fs and dioxin-like PCBs
- Bioanalytical screening methods for PCDD/Fs and dioxin-like PCBs
- any kind of method for indicator PCBs

For reporting of results, laboratories applying physico-chemical methods are asked to report, besides the analytes of interest and the sum parameters, if or if not the test sample exceeds respective EU maximum or action levels layed down as sum parameters for regulated feed and food matrices beyond reasonable doubt, taking into account the measurement uncertainty and the applied expanded measurement uncertainty. Laboratories applying bioanalytical screening methods are requested to report if the sample is suspected to be noncompliant with EU legal limits and confirmation is required, and, if applicable, PCDD/F and/or PCB results in bioanalytical equivalents (BEQ).

Test material

Test materials are prepared from regular market food/feed or samples from contamination incidents are used. In some cases test samples are fortified with the analytes of interest. Selection and/or fortification of the test materials is performed in a way that the concentration of at least one of the sum parameters WHO-PCDD/F-PCB-TEQ, WHO-PCDD/F-TEQ, WHO-PCB-TEQ and sum of six indicator PCBs covers the range of the level of interest (= maximum or action levels as laid down in Commission Regulations (EU) No 277/2012[4] and No 1259/2011[5] and in Commission Recommendation 2011/516/EU[6]).

Tests for sufficient homogeneity are performed for the sum parameters WHO-PCDD/F-PCB-TEQ, WHO-PCDD/F-TEQ, WHO-PCB-TEQ, the sum of six indicator PCBs and individual congeners.

Calculation of assigned values

Statistical evaluation of the PT results is performed according to ISO 13528[2] and the IUPAC protocol[3]. Assigned values for the test samples are determined by estimating the consensus value of participants' GC-MS/GC-ECD results (including LOQ of individual congeners). The Huber robust mean[9] is taken as assigned value after excluding extreme outliers (outside the range of ± 50 % of the median of all reported results) and examination of the distribution of the remaining results using histogram and kernel density estimation, if necessary. The proportion of participants' results contributing to the assigned value is calculated and must be higher than 2/3 of all reported results. The Huber robust mean is additionally compared with the median of all values. For individual congeners (including LOQs) assigned values are only then calculated according to the above mentioned procedure, if less than 1/3 of all reported concentrations (including LOQ) are outside the range of ± 50 % of the median of all reported results. For other congeners only the median of all reported values (including LOQs) is calculated. Robust standard deviation and standard uncertainty on the assigned value are calculated according to IUPAC[3]. Assigned values are calculated for WHO-PCDD/F-PCB-TEQ, WHO-PCDD/F-TEQ, WHO-PCB-TEQ, the sum of six indicator PCBs and individual PCDD/F and PCB congeners. In addition, TEQ-based results are re-calculated using the WHO-TEFs of 2005.

Scoring of results

Physico-chemical methods:

Criteria for successful participation of laboratories using physico-chemical methods are based on the evaluation of the results of the sum parameters WHO-PCDD/F-TEQ, WHO-PCB-TEQ, WHO-PCDD/F-PCB-TEQ and the sum of six indicator PCBs and evaluated individual congeners. The criteria are applicable for sum parameter concentrations in the range (about 0.5 to 4 times) of the level of interest (maximum or action level).

For evaluation of results, z-scores are calculated as $z = (x - x_a) / \sigma_p$, with x_a = assigned value, x = participants result and σ_p = target deviation (fitness-for-purpose-based "standard deviation for proficiency assessment").

For WHO-PCDD/F-TEQ, WHO-PCB-TEQ and WHO-PCDD/F-PCB-TEQ the standard deviation for proficiency assessment σ_p is defined as being 10 %, for the sum of six indicator PCBs (PCB #28, 52, 101, 138, 153, 180) as 15 % and for evaluated individual PCDD/F and PCB congeners as 20 %.

Acceptable z-scores are between - 2 and + 2; not acceptable are z-scores outside the range of - 3 to + 3.

A "positive scoring" system including results for sum parameters and congeners has been developed within the EU-RL/NRL network. This new scoring system yields an assessment for one PT sample covering all relevant sum parameters and congeners and will be included in the evaluation in future PTs.

The total score is calculated according to some general principles:

- Calculation of z-scores for sum parameters and evaluated individual congeners
- Calculation of the positive scores for individual congeners based on their contribution to the sum parameters (TEQ, sum indicator PCBs) and z-score according to table 1.

Positive scoring system	|z-score| ≤ 2	2 < |z-score| ≤ 3	|z-score| > 3
Individual congeners	Positive score	Positive score	Positive score
Contribution to sum parameter* > 10 %	12	6	0
Contribution to sum parameter* 3 – 10 %	8	4	0
Contribution to sum parameter* < 3 %	6	3	0
Not evaluated congeners	0	0	0

*separately for the respective sum parameters WHO-PCDD/F-TEQ, WHO-PCB-TEQ and the sum of six indicator PCBs

Table 1: Calculation of positives scores for individual congeners

The participant's positive scores for the different groups (17 PCDD/Fs, 12 DL-PCBs and 6 indicator PCBs) are compared with the maximum achievable scores for these groups and the achieved scoring percentage is calculated:

- Calculation of maximum achievable scores ($|z\text{-score}| \leq 2$) for PCDD/F and DL-PCB and indicator PCB congeners separately:

  $$\text{Maximum score} = \Sigma\text{max. score}_{(> 10\ \%)} + \Sigma\text{max. score}_{(3\text{-}10\ \%)} + \Sigma\text{max. score}_{(< 3\ \%)}$$

- Calculation of the participant's scores for PCDD/F and DL-PCB and indicator PCB congeners separately:

  $$\text{Participant's score} = \Sigma\text{score}_{(> 10\ \%)} + \Sigma\text{score}_{(3\text{-}10\ \%)} + \Sigma\text{score}_{(< 3\ \%)}$$

- Calculation of achieved scoring percentage for each participant:

  $$\text{Participant's scoring percentage} = \text{Participant's score} / \text{Maximum score} \times 100$$

- Criteria for successful participation:

Sum parameters:	≤ 1 parameter with z-score > $\lvert 2 \rvert$, no parameter with z-score > $\lvert 3 \rvert$
PCDD/F congeners:	≥ 75 % of maximum score
DL-PCB congeners:	≥ 75 % of maximum score
Indicator PCB congeners:	≥ 75 % of maximum score

The assessment based on the positive scoring system is performed for each PT test sample. A laboratory participates successfully for a PT test sample if all above mentioned criteria for the reported analytes are met.

Bioanalytical screening methods:
According to Commission Regulations (EC) No 278/2012[7] and 252/2012[8], "a screening method in principle classifies a sample as compliant or suspected to be non-compliant. For this, the calculated BEQ level is compared to the cut-off value […]. Samples below the cut-off value are declared compliant, samples equal or above the cut-off value as suspected to be non-compliant, requiring analysis by a confirmatory method." Therefore, the main criterion for evaluation of results from bioanalytical screening methods is their ability to reliably identify compliant samples and samples suspected to be non-compliant with established legal limits. For further evaluation of the performance of bioanalytical screening methods, bioassay-scores were applied: The reported BEQ-values derived from bioanalytical screening methods are compared to the WHO-TEQ consensus values calculated on basis of the results of physical-chemical methods for the concentration range of 0.5 to 2 times the level of interest. Due to the focus of bioanalytical screening methods on the decision over compliance or potential non-compliance of a sample, direct comparison of bioassay-scores and z-scores is not possible. However, bioassay scores may serve as a tool to assess method performance within the scope of external quality control measures of the respective laboratory.
Bioassay-scores are calculated as: bioassay-score $=(x - x_a) / \sigma_{bioassay}$, with x_a = assigned value (physical-chemical methods), x = participants result (BEQ from bioanalytical screening method), and $\sigma_{bioassay}$ = bioassay target deviation.
For PCDD/F-BEQ, PCB-BEQ and PCDD/F-PCB-BEQ, the bioassay target deviation $\sigma_{Bioassay}$ is defined as being 20 %.

Assessment of analytical results and measurement uncertainty
Reported concentrations and measurement uncertainties applied are compared with the respective action and maximum levels for the sum parameters. In previous PTs, the assessment of analytical results and the application of the expanded measurement uncertainty showed that comparable concentrations could be assessed either compliant or non-compliant with the established maximum or action levels, especially if the concentrations of PT test samples are about 20 % above maximum or action levels. For sum parameters with assigned values clearly above or below maximum or action levels, more than 90 % of the results were reported as exceeding or below maximum or action levels, respectively.

To find out whether the applied expanded measurement uncertainty is realistic the analytical result including expanded measurement uncertainty is compared with the assigned value including uncertainty. According to Commission Regulations (EC) No. 278/2012[7] and 252/2012[8] measurement uncertainty may be taken into account by calculating the expanded uncertainty, using a coverage factor of 2 providing a level of confidence of approximately 95 %. In figure 1, participants' results including expanded measurement uncertainty are compared with the assigned value. The assigned value should be covered by the range of the analytical result including expanded measurement uncertainty, if the applied expanded measurement uncertainty is realistic.

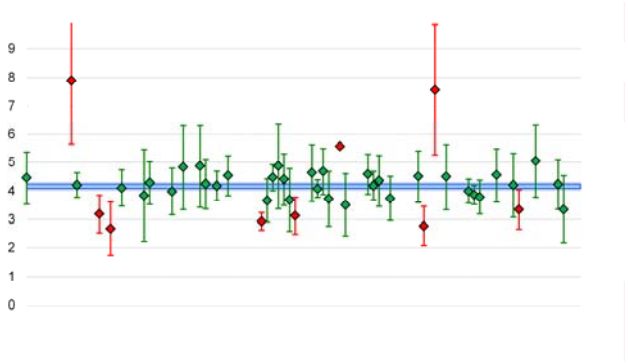

Figure 1: Comparison of analytical results including expanded measurement uncertainty with assigned value (Green dots/error bars: assigned value within range, red dots/error bars: assigned value outside range, blue line: assigned value incl. uncertainty)

Other tools for assessing the expanded measurement uncertainty within the framework of the evaluation of PT results include for example the calculation of E_n-numbers and ζ(zeta)-scores.

Summary and conclusions

Evaluation of results from EU-RL proficiency tests is based on international standards and on the respective IUPAC protocol. For a more comprehensive overview on the performance of participants, and of NRLs in particular, evaluation of EU-RL PT results not only covers deviation of participants' results from assigned values, but also includes the assessment of analytical results of physico-chemical and bioanalytical screening methods and the application of the expanded measurement uncertainty.

Comparison of calculated z-scores derived from EU-RL PTs with the acceptable deviations of results based on criteria for trueness and precision as laid down in relevant Commission Regulations[7,8] shows that the criteria defined for the fitness-for-purpose-based standard deviation for proficiency assessment are stricter than the analytical criteria laid down in these Regulations. This approach strongly supports the objective of EU-RL PTs to demonstrate and maintain the required high analytical quality of European NRLs.

Acknowledgements

We would like to thank the European Commission for the financial support of the work of the European Union Reference Laboratory for Dioxins and PCBs in Feed and Food (EU-RL), Freiburg, Germany, and the network of EU-RL and National Reference Laboratories for Dioxins and PCBs in feed and food (NRLs) for their scientific contribution.

References

1. ISO/IEC 17043:2010, Conformity assessment - General requirements for proficiency testing
2. ISO 13528:2005, Statistical methods for use in proficiency testing by interlaboratory comparisons
3. International Harmonized Protocol For The Proficiency Testing Of Analytical Chemistry Laboratories , IUPAC Technical Report, Pure Appl. Chem, Vol. 78, No. 1, pp-145-196, 2006
4. Commission Regulation (EU) No 277/2012 of 28 March 2012 amending Annexes I and II to Directive 2002/32/EC (OJ L 91, 29.3.2012, p. 1–7)
5. Commission Regulation (EU) No 1259/2011 of 2 December 2011 amending Regulation (EC) No 1881/2006 (OJ L 320, 3.12.2011, p. 18–23)
6. 2011/516/EU: Commission Recommendation of 23 August 2011 (OJ L 218, 24.8.2011, p. 23–25)
7. Commission Regulation (EC) No 278/2012 of 28 March 2012 amending Regulation (EC) No 152/2009 (OJ L 91, 29.3.2012, p. 8–22)
8. Commission Regulation (EU) No 252/2012 of 21 March 2012 repealing Regulation (EC) No 1883/2006 (OJ L 84, 23.3.2012, p. 1–22)
9. Huber, P. J. (1981) *Robust Statistics*, Wiley, New York (1981)