MEASUREMENT UNCERTAINTY FOR PERSITENT ORGANIC POLLUTANTS BY ISOTOPE-DILUTION MASS SPECTROMETRY

Gauthier Eppe[1], Gianfranco Diletti[2], Alwyn Fernandes[3], Johannes Haedrich[4], Jerry Hart[5], Helge Hove[6], Anna Laura Iamiceli[7], Alexander Kotz[5], Rainer Malisch[5], Philippe Marchand[8], Wolfgang Moche[9], Georges Scholl[1], Giampiero Scortichini[2], Thorsten Bernsmann[10], Yves Tondeur[11], Wim Traag[12]

[1] CART, Chemistry Department, University of Liege, Belgium
[2] Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise "G. Caporale, Teramo, Italy
[3] The Food and Environment Research Agency (FERA), York, YO41 1LZ, United Kingdom
[4] European Union Reference Laboratory for Dioxins and PCBs in Feed and Food, Freiburg, Germany
[5] SGS Environmental Services, Wilmington, NC, USA
[6] National Institute of Nutrition and Seafood research, Bergen, Norway
[7] Istituto Superiore di Sanitá (ISS), Roma, Italy
[8] LABoratoire d'Étude des Résidus et Contaminants dans les Aliments (LABERCA), Nantes, France
[9] Federal Environment Agency Ltd, Austria
[10] Chemical and Veterinary Analytical Institute Münsterland-Emscher-Lippe, Münster, Germany
[11] It's About Purpose, LLC, Carolina Beach, NC, USA
[12] RIKILT, Wageningen, The Netherlands

Introduction

An analytical result cannot be properly interpreted without knowledge about the uncertainty inherent in its measurement. Estimation of measurement uncertainty (MU) is not only a requirement of ISO 17025 for testing laboratories but it is generally acknowledged that the fitness-for-purpose of analytical results cannot be correctly assessed without MU. For the purposes of regulating dioxins and PCBs in food and feed, legislation referring to maximum levels should always address how analytical results should be expressed and interpreted. All analytical results actually take the form of an interval 'x ± U', e.g. (3.5 ± 0.5) pg-WHO-TEQ/g fat, where x is the analytical result (the best estimate of the true value) and U the expanded measurement uncertainty, at a specified level of confidence (e.g. 95% or 99%). x ± U is the range within which the unknown true value of the real sample analyzed is most likely to fall, at a stated level of confidence.

The Guide to the expression of Uncertainty in Measurement (GUM), published by ISO[1], establishes general rules for evaluating MU. The guide has been interpreted for analytical chemistry by EURACHEM[2] and in other guidelines[3,4,5]. In this paper, we describe the general approaches proposed by a working group established within the network of the European Union Reference laboratories network (EU-RL) and the National Reference Laboratories (NRLs) of the European member states for estimation of MU for persistent organic pollutants (POPs) quantified by isotope dilution-mass spectrometry based techniques. The various approaches proposed here, only consider the uncertainty associated with the analytical measurement procedure. Uncertainties related to sampling, homogeneity or stability of the sample also contribute to the total uncertainty but they are not discussed here.

Measurement Uncertainty in the context of EU legislation for dioxin and PCBs in food and feed

In the context of EU legislation where maximum and action limits were set for various kinds of food and feeding stuffs, MU is used for compliance assessment. The interpretation of results is depicted in Figure 1. In practice, if we consider a maximum value in legislation, the analyst will measure the TEQ level in the sample and estimate the associated MU in TEQ at that level, subtract the uncertainty from the reported concentration (x-U) and use that value to assess compliance. Only if the (x-U) value is greater than the legislation limit, confirmed by a second independent analysis, will the control analyst be sure 'beyond reasonable doubt' that the sample concentration exceeds the limit prescribed by food and feed legislation. Thus, according to that definition, in Figure 1, only situation 4 is non-compliant beyond reasonable doubt.

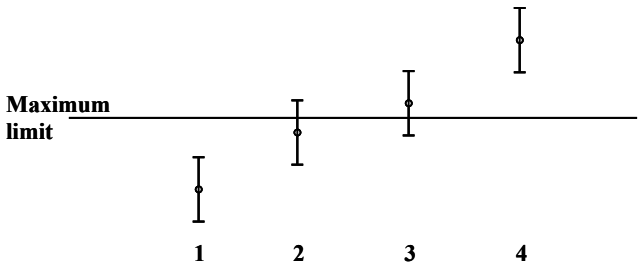

Figure 1: Interpretation of results for compliance assessment of PCDD/Fs and PCBs levels in food and feed.

Working group on Measurement Uncertainty within the network of EU-RL and NRLs

In conclusion to a report on results (including data on MU) from various accredited laboratories having participated in proficiency tests (PTs) organized by the EU-RL, , an expert group was established in 2012. It consisted of a number of representatives from European NRLs and two experts from the US. The main task was to evaluate a harmonized approach for estimation of MU for POPs based on measurement by isotope dilution mass spectrometry.

Table 1 summarizes the expanded uncertainty values associated with laboratory results for salmon filet within the scope of an EU-RL PT in 2012.

Table 1: Reported expanded uncertainties (U %) for results for salmon filet (EU-RL PT organized in 2012)

Salmon filet	Number of labs. Reporting MU (%)	Mean U(%)	Median U(%)	Minimum U(%)	Maximum U(%)
WHO-PCDD/F-PCB TEQ	28	20	20	12	31
WHO-PCDD/F-TEQ	30	20	20	7	30
WHO-PCB-TEQ	30	21	20	10	39
Sum NDL-PCBs	42	22	21	6	40

Table 1 shows that, although median or mean values reported by laboratories are consistently around 20%, the range between minimum and maximum U values is quite large, depending on the approach used by laboratories. According to the manner in which MU is taken into account for compliance assessment in the EU legislation, it is evident that similar TEQ results reported by two laboratories with high discrepancies between their MU estimates could lead to contradictory decisions.

Proposed approaches

International bodies, ISO or Eurachem, recognize inter-laboratory study and validation processes as a valid basis for MU in analytical work[2,3]. Both sources are appropriate to ultra-trace analysis of POPs, mainly due to the complex sample preparation and measurement steps required. In this way, MU becomes a natural extension of the validation process, which is, for analytical chemists, an easier concept to understand compared to other approaches. The main disadvantage of this approach is that it relies on data that has been recorded during a validation study and the MU calculated does not necessarily reflect the uncertainty associated with results obtained in daily routine analyses. In addition, it gives little insight as to where the major sources of uncertainty lie because it is based on a global (i.e. top-down) approach. The other approach that could be used is to estimate MU derived according to the GUM component-by-component approach (i.e. bottom-up)[1]. The latter approach consists of an exhaustive dissection of the analytical system, i.e. by decomposition of errors, followed by determination of all the relevant sources of uncertainties and finally combining all the relevant individual contributions to an overall uncertainty. This approach gives more insight into the analytical process but also on uncertainties associated with daily or routine sample measurement, but it might lead to over or underestimation of MU as some sources can easily be overlooked, doubly accounted or wrongly quantified, especially in the case of a long and complex analytical process.

The working group proposed to develop a concept including both approaches as they reflect the procedures implemented and used by laboratories in the network but also because both approaches gave comparable results. In order to take advantage of each method but also to overcome the drawbacks highlighted, an on-going top-down approach is suggested that integrates historical data, reflecting the precision and trueness of the analytical method but also additional tools such as matrix-based quality control (QC), PTs results generated several times a year, the daily limits of quantification (LOQs), matrix and procedural blank effects, and the variability of relative response factors (RRFs) in order to take into account as much as possible the impact of daily performance in the MU assessment.

According to Barwick and Ellison[5], two sets of experiments can be carried out, a precision study and a trueness study, which will provide the information required for an estimate of the combined uncertainty of the method. They should be planned in a way that as many sources of uncertainty identified in a cause and effect diagram as possible, are covered. The experimental design consists of a long-term precision study for looking at repeatability and between run variability by using naturally contaminated quality control (QC) samples. The QC samples must match matrix and levels of interest (i.e. assessing MU at 0.5x, 1x and 2x the maximum limit would be consistent with the current EU legislation). The total range will cover 0.5x the lowest maximum limit, to 2x the highest maximum limit. QC samples should be treated in exactly the same way, covering the whole analytical procedure.

The trueness study, based on relevant matrix Certified Reference Materials (CRMs), or if CRMs are not available, matrix-matched spiking experiments, and inter-laboratory studies or PTs, provides estimation of the laboratory bias plus the method bias. If an appropriate certified reference material (CRM) is available, a single laboratory test allows a laboratory to assess laboratory bias and method bias in combination, by analyzing the

certified reference material over a number of times (n≥6). If CRMs are not available, participation in PTs is a good alternative in order to cover a full range of food and feed matrices for trueness assessment. In order to rationalize the diversity in matrices, the working group decided to differentiate matrices for PTs according to the method of analysis (e.g. fat based matrices, feed, dry food,…) and suggests 6 participations in PTs resulting on average in a moving time window of 3 years if laboratories participate twice a year at PTs. Thus, any new PT result would be added to the root mean square bias (RMS_{bias})[4] calculation while the oldest PT result would be removed from the dataset to maintain 6 PT results in the calculation. This approach gives enough weight to the latest PT result in the assessment of the bias contribution (U_{bias}) to combined uncertainty, reflecting more accurately, current laboratory performance. Priority is given first in calculating U_{bias} in TEQ (WHO-TEQ PCDD/Fs, WHO-TEQ DL-PCBs and WHO-TEQ PCDD/Fs+DL-PCBs) but also for each individual congener although all of them are not necessarily given an assigned value in each PT. To correctly estimate U_{bias}, the uncertainty associated with the assigned value from the PTs, i.e. $u(C_{ref})$, is required. PT providers are urged to assess the uncertainty on the assigned values and provide all related statistical information in order to facilitate the retrieval of this information for laboratories participating in the PTs, and intending to apply the proposed approach. Generally $u(C_{ref})$ is higher than $u(C_{CRM})$ because both PTs and the certification process in CRMs are not designed for the same purposes. Consequently, the contribution of $u(C_{ref})$ might have a significant impact on the laboratory combined uncertainty, while it is not directly related to its performance. A performance criterion, based on robust standard deviation[6], is applied on $u(C_{ref})$ to decide whether or not the PT study can be used for MU assessment. The proposed top-down approach will allow reassessment of the MU on a regular basis over the time.

However, MU is associated with a result and not with a method. To take this fact into account, the working group proposes to include the contribution of daily performances in the combined uncertainty by integrating a sensitivity parameter such as dynamic limits of quantification (LOQs), recalculated for each set/batch from the blank but also the impact of daily calibration and RRF responses if significant variation in these is observed.

For the bottom-up approach, the working group adopted the EURACHEM guidelines providing that uncertainty components significantly contributing to the overall uncertainty should be considered[2]. The approach distinguishes Type A components derived from the statistical distribution of the quantity values from series of measurements; e.g. intra-laboratory reproducibility, calibration curve, and Type B components evaluated from probability density functions based on experience or other information; e.g. standard solution concentration, volume of internal standard added to the sample. Overall bias (best estimated by repeated analysis of relevant CRMs) and data obtained from proficiency testing need also to be included in the uncertainty estimation. Other factors subjected to daily variation, such as mass response linearity (variation of RRFs), calibration curve drift and analyte-specific detection limit in the actual sample, should be taken into account for a comprehensive MU evaluation.

Combined uncertainty in TEQ from individual congeners

By definition and principle, MU is associated with a measurand (e.g. a congener concentration in pg/g) and not the TEQ. However, the EU legislation expressed maximum limits in TEQ units and MU must therefore be expressed in TEQ for decision-making and compliance assessment. Since the congener-specific isotope-dilution mass spectrometry confirmatory method is mandatory for compliance assessment in the context of EU legislation, it involves evaluating MU in toxic equivalent units. The question arises as to how the MU estimated for each individual congener could be propagated to provide a MU expressed in TEQ?

Four different approaches were compared based on quality control data provided by members of the working group. The four approaches were the following:

1) The square root of the sum of squares (RSS): $u_c(TEQ) = \sqrt{\sum_{i=1}^{29} (TEF_i * u_{ci})^2}$

2) The SUM: $u_c(TEQ) = \sum_{i=1}^{29} (TEF_i * u_{ci})$

3) The average of $(u_{ci} * TEF_i)_{i=congener}$

4) The median of $(u_{ci} * TEF_i)_{i=congener}$

Statistical parameters were calculated from analytical results derived from 13 different sample matrices involving more than 16000 individual congeners. Plotting the empirical standard deviation in TEQ against the standard deviation calculated according to each of the four different approaches mentioned above (Figure 2) reveals that the average (3) and median (4) approaches clearly underestimate the total SD for the sum of PCDD/Fs and DL-PCBs compared to the empirical values. The sum approach (2) clearly overestimates the empirical value while the square root of the sum of squares (1) slightly underestimates the experimental SD in TEQ as can be seen by comparing the different slopes of regression curves.

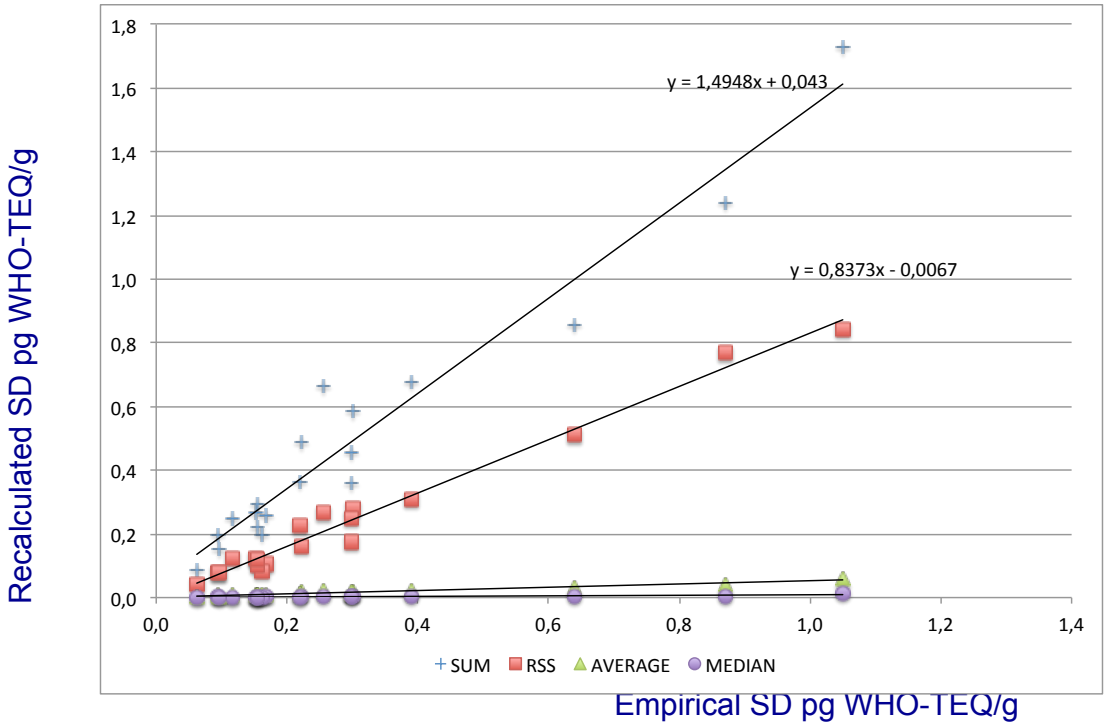

Figure 2: Empirical SD in TEQ versus recalculated SD in TEQ by four approaches for PCDD/Fs + DL-PCBs by approaches 1 through 4.

Of the different method used to propagate uncertainty to the TEQ, it appears that the square root of the sum of squares (RSS) provides the most realistic estimate.

The working group will continue to consolidate the above findings and use examples of real data to finalize the approaches to estimating MU associated with the determination of POPs by isotope dilution mass spectrometry.

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 the scientific contribution.

References:

1. Guide to the expression of uncertainty in measurement (GUM), ISO, 1993.
2. Quantifying uncertainty in analytical chemistry, EURACHEM/CITAC guide third edition 2012.
3. ISO/TS 21748:2004(E), Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation.
4. The Nordtest report, TR 537, 2003-05
5. Barwick V.J., Ellison S.L.R. Development and harmonisation of measurement uncertainty principles. Part (d): Protocol for uncertainty evaluation from validation data, VAM Project 3.2.1.
6. ISO13528: 2005(E), Statistical methods for use in proficiency testing by interlaboratory comparisons