HANDLING OF NON-DETECT DATA: A CASE STUDY OF PCDD/Fs IN FISH SPECIES FROM SOUTH KOREA

Barghi M[1], Kwon H-Y[1], Lee Y-S[1], Choi S-D[1]*, Chang Y-S[2]

[1] School of Urban and Environmental Engineering, Ulsan National Institute of Science and Technology (UNIST), UNIST-gil 50, Ulsan, 689-798, Republic of Korea; [2] School of Environmental Science and Engineering, Pohang University of Science and Technology (POSTECH), 77 Cheongam-ro, Pohang, 790-784, Republic of Korea

Introduction

The ultra-trace analysis of persistent organic pollutants (POPs) using mass spectrometry is essential to understand the fate of POPs in ecosystems. In spite of the development of analytical instruments, extremely low levels of POPs are frequently non-detected. Data below the detection limit (DL) are reported only as "non-detect" or "< DL", and no value is provided[1].This type of data distribution is called "left-censored"[2].

One of the methods commonly used for summing or averaging data that include nondetects is to assign one-half the detection limit to each non-detect. This method was introduced as a simple method for calculating mean values with relatively small bias, but data with high detection limits often have a strong influence on the resulting mean values. Several previous studies concluded that methods replacing all non-detects with a single value (substitution methods) were frequently inferior[3,4].

The other recommended method for handling non-detects are statistically intensive approaches such as Kaplan-Meier (KM), Maximum Likelihood Estimation (MLE), and Robust Regression on Order Statistics (ROS)[5]. Among them, the KM method, commonly used to compute means for censored data, can be used to calculate a sum of congeners that include nondetects. This nonparametric method relies only on ranks of data and makes no assumptions about the statistical distribution of the data, which is strongly recommended by US-EPA[5].

In this study, we collected fish samples in South Korea and analyzed polychlorinated dibenzo-*p*-dioxins/furans (PCDD/Fs). Using this dataset, we compared the concentrations of 17 toxic congeners and toxic equivalents (TEQ) with different treatment methods for non-detects. The aim of this study is to know whether the handling of non-detects with statistical methods can be an important issue for PCDD/F monitoring studies or not.

Materials and methods

Sampling and instrumental analysis

A total of 240 fish samples (36 species) was acquired in 2012 from traditional markets and department stores of four cities (Busan, Gangneung, Gunsan, and Gwangju) in South Korea. The fish samples were randomly purchased independently on their geographical origin.

To analyze the 17 toxic congeners of PCDD/Fs (pg/g wet weight), fish samples were freeze-dried and homogenized before being Soxhlet extracted. Multi-silicagel columns were used for clean up, and the final samples were analyzed using a HRGC/HRMS (Autospec Premier, Waters). Identification of PCDD/Fs was based on the criteria reported by US-EPA (method 1613), and quantification was carried out by the isotopic dilution method. Detection limits (S/N ratio=3) varied depending on congeners, and concentrations below the limits of detection (LOD) were considered as non-detects (NDs). Toxic equivalents (TEQ) of the analyzed PCDD/Fs were calculated using the 2005 WHO-TEF values.

Data handling

The levels of each PCDD/F congener were calculated with both statistical and substitution methods regarding NDs in datasets. Also, the sum of 17 toxic congeners (TEQ and concentration) in each sample was calculated with the KM and substitution methods; NDs were substituted with 0, ½ DL, and DL. Afterwards, sum results with over 70% NDs were removed, and remaining data were classified into five groups as representative datasets: (a) top 50% highest concentration (n=42), (b) top 30% highest concentration (n=24), (c) samples with 60−70% NDs in their datasets (n=57), (d) samples with 30−60% NDs in their datasets (n=29), and (e) using S/N ratio=10 instead of S/N ratio=3 as LOD (n=34, 53−65% NDs). The data analysis in this study was conducted using the statistical software ProUCL 4.1 for environmental applications for data sets with and without nondetect observations, which was downloaded from US-EPA website (http://www.epa.gov/osp/hstl/tsc/software.htm).

Results and discussion

Levels of PCDD/F congeners in fish species

The mean, median, range, and ND percent of 17 PCDD/F congeners in 240 fish samples are listed in Table 1. OCDF was a predominant congener in fish samples, while with respect to TEQ, this role corresponded to 2378-TCDF. 123478-HxCDD and 123789-HxCDD were not detected in any of the samples. The mean concentration of $_{17}$PCDD/F was 0.14 pg WHO-TEQ/g ww, and it was lower than the permissible limit for sum of PCDD/Fs (4 pg WHO-TEQ/g ww) in fish set forth by the EU Council Regulation.

Table 1. Concentration (pg/g ww) of PCDD/Fs in fish samples (n=240)

Congener	Mean	Median	Range	ND%
2378-TCDD	0.002	ND	ND–0.16	97
12378-PeCDD	0.002	ND	ND–0.37	99
123478-HxCDD	ND	ND	ND	100
123678-HxCDD	0.001	ND	ND–0.16	99
123789-HxCDD	ND	ND	ND	100
1234678-HpCDD	0.31	0.27	ND–1.11	47
OCDD	0.82	0.74	ND–10.1	35
2378-TCDF	0.30	ND	ND–3.46	57
12378-PeCDF	0.074	ND	ND–4.52	95
23478-PeCDF	0.10	ND	ND–2.92	88
123478-HxCDF	0.14	ND	ND–1.46	70
123678-HxCDF	0.17	ND	ND–3.55	82
234678-HxCDF	0.007	ND	ND–0.95	99
123789-HxCDF	0.21	0.22	ND–0.66	43
1234678-HpCDF	1.44	1.43	ND–6.06	17
1234789-HpCDF	0.001	ND	ND–0.20	99
OCDF	1.89	1.96	ND–8.72	22
Σ_{17} PCDD/Fs*	5.46	5.12	ND–16.4	
WHO-TEQ*	0.14	0.07	ND–1.38	

*Nondetected concentrations were assumed to be equal to zero (ND=0).

Calculating the sum of PCDD/Fs in fish samples and comparing fish species

The total TEQ levels of PCDD/Fs in each fish sample were calculated using the KM method regarding NDs[3]. In this calculation, KM derived means were multiplied by the number of congeners, 17. Among 36 fish species, the highest TEQ levels were found in *Clupea pallasii* (0.55 pg WHO-TEQ/g ww) (n=4), followed by *Scomber japonicas* (0.51 pg WHO-TEQ/g ww) (n=8), and *Arctoscopus japonicas* (0.38 pg WHO-TEQ/g ww) (n=7), respectively. In contrast, the lowest TEQ levels were detected in *Anguilla japonica* (0.027 pg WHO-TEQ/g ww) (n=5), *Lophiomus setigerus* (0.041 pg WHO-TEQ/g ww) (n=7), and *Cynoglossus joyneri* (0.055 pg WHO-TEQ/g ww) (n=5) respectively.

In a similar study, the concentrations of PCDD/Fs in 24 fish species from Korean coastal waters varied from 0.03 to 1.63 pg WHO-TEQ/g ww[8]. Also, in China, the PCDD/F concentrations of 10 marine fish species were measured between 0.308 and 0.947 pg WHO-TEQ/g ww[9]. In this study, no sample exceeded the maximum permissible level (4 pg WHO-TEQ/g ww).

Figure 1 demonstrates the profiles of 17 PCDD/F congeners in six marine fish species showing the maximum and minimum TEQ levels. Even though the TEQ concentrations differ a lot among these fish species, the congener profiles were rather similar. The average concentrations of Σ_{17} PCDD/Fs in *Clupea pallasii*, *Scomber japonicas*, and *Arctoscopus japonicas* with the highest TEQ concentrations among 36 fish species were 5.2, 3.6, and 4.4 pg/g ww while in *Anguilla japonica*, *Lophiomus setigerus*, and *Cynoglossus joyneri* which contained the lowest TEQ, the concentrations of 17 congeners were measured 5.4, 6.3, and 3.0 pg/g ww, respectively.

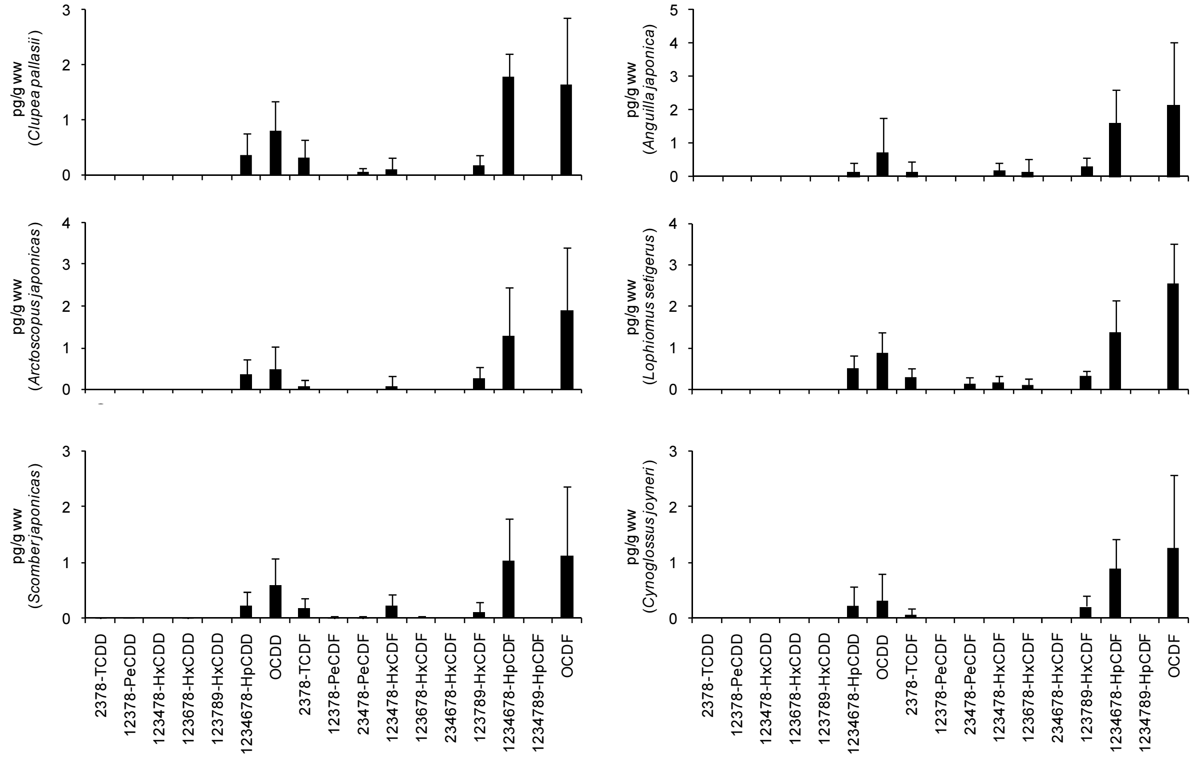

Figure 1. Average concentrations of 17 PCDD/F congeners in six marine fish species showing the maximum and minimum TEQ levels from South Korea.

Comparing representative groups

Table 2 shows the TEQ and PCDD/F concentrations of five representative data groups. PCDD/F concentrations were calculated with three substitution methods and the KM method. All groups showed the same trend: NDs=0 < KM method < NDs=½ LOD < NDs=LOD. Therefore, the KM results positioned somewhere between those of NDs=0 and NDs=½ LOD. Also, with the exception of group (b) with the top 30% concentrations, all groups showed the similar levels of total concentrations. Selecting the top 30% concentration of the total dataset has a significant effect on the average TEQ concentration in all methods, while other representative groups showed the same range of average TEQ, indicating that difference in ND percentages and DL values did not have a large influence on TEQ concentrations in this study. However, a comparison between group (c) and (d) demonstrates that the range of average concentrations in group (c) with a higher percentage of NDs is larger than that of group (d).

Table 2. Average TEQ and PCDD/F concentrations for five representative groups calculated by the three substitution and KM methods (pg/g ww)

		average TEQ				average concentration			
		(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
a	Top 50% concentration	0.23	0.25	0.29	0.35	7.71	8.02	8.21	8.72
b	Top 30% concentration	0.44	0.45	0.50	0.56	10.8	11.1	11.3	11.8
c	60–70% NDs	0.24	0.25	0.30	0.36	7.11	7.43	7.64	8.18
d	30–60% NDs	0.25	0.27	0.30	0.36	7.44	7.75	7.91	8.39
e	S/N=10	0.27	0.32	0.36	0.41	7.34	7.73	8.10	8.89

(1) NDs=0, (2) KM, (3) NDs=½ LOD, (4) NDs=LOD

In order to find differences in average concentrations among four methods, the percentages of these values were calculated regarding the average concentration in method (4) as 100%. Results of average TEQ indicated that group (c) has the widest range (66–100%) among the other representative datasets, which could be due to the high percentage of nondetects in this dataset. Several studies also demonstrated that the bias caused by substitution increased dramatically as the percent of censored observations increases[6]. However, for the average PCDD/F concentration, group (e) (S/N=10) led to the widest range of results (83–100%).

For an easier explanation, we subtracted the percentages which were calculated to find the differences of results among four methods (Table 3). Methods (4) and (3) have the highest difference of both TEQ and PCDD/F concentrations, whereas methods (1) and (2) and methods (2) and (3) have the least differences in TEQ and PCDD/F concentration, respectively. In other words, when ND data were substituted with LOD, the mean of PCDD/F concentrations showed noticeable differences with KM results.

According to the results of this study, the substitution methods are not appropriate for the calculation of the sum of PCDD/Fs especially for TEQ concentrations. If we consider KM derived data as reliable statistical mean values, substituting NDs with 0 would be more suitable than other substitution methods for both TEQ and concentrations. As the comparison result for TEQ inthis study is dependent on PCDD/F profiles, reliable methods for calculating the sum of PCDD/Fs can be specific for different types of samples.

Table 3. Percentage differences among three substitution and KM methods

	average TEQ concentration			average concentration		
	(2) - (1)	(3) - (2)	(4) - (3)	(2) - (1)	(3) - (2)	(4) - (3)
50% highest Con.	4	13	17	4	2	6
30% highest Con.	3	8	11	2	2	4
60–70% ND	4	13	17	4	3	6
30–60% ND	5	10	15	4	2	6
SN=10	11	9	14	4	4	9

(1) NDs=0, (2) KM, (3) NDs=½ LOD, (4) NDs=LOD

Acknowledgements
This work was supported by the KoreaFood andDrug Administration (KFDA) (No.12162KFDA015).

References

1. Helsel D. (2005); *Environ Sci Technol.* 39(20): 419A-423A
2. Leith K, Bowerman W, Wierda M, Best D, Grubb T, Sikarske J. (2010); *Chemosphere* 80(1): 7-12
3. Helsel D. (2012); Statistics for censored environmental data using Minitab and R
4. Eastoe E, Halsall C, Heffernan J, Hung H. (2006); *Atmos Environ.* 40(34): 6528-6540
5. US-EPA (2013); Preparing Data for Analysis, http://www.epa.gov/ttnamti1/files/ambient/airtox/workbook/AirtoxWkbk4Preparingdataforanalysis.pdf
6. Helsel D. (2010); *Integr Environ Assess Manag*. 6(3): 361-366
7. She N. (1997); *J Am Water Resour As*. 33(3): 615-624
8. Moon HB, Ok G. (2006); *Chemosphere* 62(7): 1142–52
9. Wei X, Leung KS, Wong MH, Giesy J, Cai ZW, WongChris KC. (2011); *Mar Pollut Bull* 63(5-12): 166-171