Cluster analysis

Hierarchical clustering (HC) was used to identify compounds with similar patterns in mother's milk levels based on correlations between levels of single compounds. For this purpose, a subset of the total dataset was compiled which included the maximum number of samples with the largest number of measured compounds. PCB 77, 123789-HxCDF and 1234789-HpCDF were not included in the cluster analysis because of a large proportion of samples with levels below LOQ (limit of quantification) (63-82%). The compiled dataset included results for 184 mother's milk samples, each with recorded concentrations of 29 compounds. In previous studies, we have identified that both mother's age and sampling year have large impact on POP concentrations in mother's milk.[1] Therefore all concentrations were adjusted for these two factors prior cluster analysis. Cluster analysis was performed in the MATLAB™ programming environment (The MathWorks Inc., Natick, MA). Two separate techniques for HC were employed: standard binary *agglomerative* HC and a flexible multi-branching *divisive* HC algorithm. In the former algorithm, series of data partitions, P_n, P_{n-1}, ... , P_1 are iteratively produced, where P_n consists of all n compounds as n single clusters and the P_1 consists of a single group containing all n compounds. At each particular stage the two nearest clusters are joined. Conversely, in the latter procedure all compounds are at the initial point a single cluster, which iteratively is divided into sub-clusters. This divisive HC method employs the classical *k*-means algorithm to achieve cluster representations at each branching point and the likewise classical silhouette measure for automatic selection of the most suitable number of sub-clusters.[2-4] Both HC-procedures were conducted using the same dissimilarity measure (d):

$$d\,(X, Y)$$

$$d\,(X, Y) = 1 - r_{S(XY)}$$

where $r_{S(XY)}$ is the Spearman rank correlation coefficient, between two compounds *X* and *Y* (or rather sub-clusters containing either one or several compounds). In the agglomerative HC, *complete linkage (furthest neighbor)* was used, wherein the greatest distance of cluster members defines the cluster distance. The hierarchic structure was depicted in a dendrogram where all levels are graphically shown; from the point where one compound equals one cluster until the level where one cluster contains all compounds. In the divisive HC, the maximal number of sub-clusters (k) allowed at each partitioning was set to 10. Since standard dendrograms only allows binary splits at each hierarchical level, the structure was visualized in a specialized dendrogram, featuring a flexible number of multiple branches.[5]

Table S1. Concentrations of POP in mother's milk from primiparas in Uppsala, Sweden.

Substance	N	Mean ± SD	Range	Unit
PCB 28	325	2.8 ± 3.9	<0.50 – 31	ng/g lipid
PCB 153	325	58 ± 28	12 – 186	ng/g lipid
di-*ortho* PCB[a]	325	114 ± 53	24 – 363	ng/g lipid
mono-*ortho* PCB TEQ[b,c]	325	0.55 ± 0.31	0.13 – 2.7	pg/g lipid
non-*ortho* PCB TEQ[b,d]	220	5.1 ± 2.4	1.3 – 14	pg/g lipid
PCDD TEQ[b]	184	4.7 ± 1.8	1.4 – 12	pg/g lipid
PCDF TEQ[b]	184	2.3 ± 1.0	0.81 – 7.0	pg/g lipid
Total-TEQ[b,e]	183	13 ± 5.1	4.0 – 31	pg/g lipid
BDE-47	276	1.9 ± 1.7	<0.40 – 16	ng/g lipid
BDE-99	276	0.45 ± 0.51	<0.12 – 5.2	ng/g lipid
BDE-100	276	0.36 ± 0.40	<0.10 – 5.1	ng/g lipid
BDE-153	276	0.64 ± 0.45	<0.23 – 4.6	ng/g lipid
sumPBDE[f]	276	3.4 ± 2.7	0.86 – 28	ng/g lipid

[a]sum of PCB 138, PCB 153 and PCB 180 [b]TEQs calculated with TEFs from 2005 (Van den Berg et al., 2006).[6] [c]sum of PCB 105, PCB 118, PCB 156 and PCB 167 TEQs. [d]sum of PCB 77, PCB 126 and PCB 169 TEQs. [e]sum of mono-*ortho* PCB TEQ, non-*ortho* PCB TEQ, PCDD TEQ and PCDF TEQ. [f]sum of BDE-47, BDE-99, BDE-100 and BDE-153. Levels <LOQ (limit of quantification) were set to ½LOQ in the calculations.

References

1 S. Lignell, M. Aune, P. O. Darnerud, S. Cnattingius and A. Glynn, *Environ. Res.*, 2009, **109**, 760-767.
2 S. Lamrous and M. Taileb, International Conference on Computational Intelligence for Modelling, Control and Automation, and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, Sydney, 2006.

3 T. Hastie, R. Tibshirani and J. Friedman, *The elements of statistical learning: data mining, inference and prediction*, Springer-Verlag, New York, 2001.
4 P. J. Rousseeuw, *J. Comp. App. Math.*, 1987, **20**, 53-65.
5 D. Soeria-Atmadja, A. Onell and A. Borga, *J. Allergy Clin. Immunol.*, 2010, **125**, 1379-1386 e1.
6 M. Van den Berg, L. S. Birnbaum, M. Denison, M. De Vito, W. Farland, M. Feeley, H. Fiedler, H. Hakansson, A. Hanberg, L. Haws, M. Rose, S. Safe, D. Schrenk, C. Tohyama, A. Tritscher, J. Tuomisto, M. Tysklind, N. Walker and R. E. Peterson, *Toxicol. Sci.*, 2006, **93**, 223-241.