MULTIVARIATE STATISTICAL ANALYSIS OF DIOXIN PROFILES TO EXPLAIN SOURCE CONTRIBUTIONS TO SERUM DIOXINS

[1]Adriaens, P, [3]Towey, T, [3]Barabas, N, [1]Chang, S-C, [2]Hedgeman, E, [2]Zwica, L, [2]Garabrant, D,

[1]Department of Civil and Environmental Engineering, 1351 Beal Ave, The University of Michigan, Ann Arbor, MI 48109 USA, [2]School for Public Health, 1420 Washington Heights, The University of Michigan, Ann Arbor, MI 48109 USA, [3]Limno-Tech, Inc., 501 Avis Dr. Ann Arbor MI 48108 USA

Key words: Principal components analysis, dioxin patterns, polytopic vector analysis

Abstract

The major determinants of human dioxin body burden include the sources and routes of exposure and the toxico-kinetic processes occurring after uptake. This paper describes out approach and initial results of multivariate statistical analyses to identify contributing sources and modifying factors to explain serum dioxin patterns in the University of Michigan Dioxin Exposure Study (UMDES) cohort. The study was designed to quantify the differences in exposure between populations living in an industrially-impacted area, and a control group. Principal components analysis (PCA) and cluster analysis of soil (n = 2,098), blood (n = 945) and dust (n = 760) dioxin patterns indicated that (i) serum pattern clusters indicate some degree of differentiation with age, suggesting the need for comparing age-stratified subsets to account for dioxin elimination, (ii) soil patterns clustered along floodplain, combustion and combustion-like 'background patterns with regional characteristics, suggesting further resolution may be achieved after regionalization, (iii) dust patterns unmixed in combustion and floodplain-type sources, in addition to 'background' patterns, suggesting the need for further analysis after geographical bounding. Whereas PCA is exploratory to screen samples and cluster patterns, polytopic vector analysis (PVA) will be required to obtain physically-meaningful source patterns and to allow for cross-media correlations.

Introduction

The distribution of PCDD/F and PCBs in human serum samples collected from cohorts impacted by incinerators[1], living close to industrial areas[2,3], sport fishing[4,5], or exposed to background levels[6-8] is well-documented. While a number of large congener-specific datasets are now available, their analysis has primarily involved descriptive statistics and correlation studies. With these methods, limited success in evaluating congener patterns with respect to contributing sources and modifying factors has been achieved. Our previous regression analysis based on TEQ and TCDD alone indicated that soil and dust were not statistically-significant contributors to explaining the levels in serum (as based on contributions to the correlation coefficient, and using population weighted analysis). Whether the lack of contribution is real, or whether subpopulations exist that are masked by the weighted analysis is unclear at this time. A promising alternative approach to assessing the impact of various environmental (e.g. dust inhalation, soil contact) sources on serum loadings, is through application of multivariate exploratory data analysis that incorporate the entire pattern of contaminants. Such ''data-mining'' techniques can provide important insights into the underlying structure of a complex dataset. In particular, certain methods known as self-training receptor models (also called mixing models) allow contributing chemical fingerprints to be resolved in complex environmental systems without a priori assumption of contributing patterns and with minimal knowledge of source, toxicokinetics, or data structure distribution[9].

For example, a multivariate analysis using polytopic vector analysis of PCB patterns in blood serum identified five patterns in an adult Native American population living in an industrialized area, each of which was characterized by a unique mix of congeners[3]. One pattern observed in a limited number of Mohawks was similar to those reported for air sampled near contaminated sediment deposits at Akwesasne and for volatilized Aroclor 1248 and is hypothesized to reflect recent inhalation exposure in these subjects. A second pattern was consistent with unaltered Aroclor 1254. A third pattern, resembling Aroclor 1262 but without labile congeners, was correlated with age and is interpreted as

representing a lifetime PCB accumulation profile. The final two patterns were dominated by subsets of major persistent congeners and are hypothesized to reflect intermediate bioaccumulation profiles and/ or differences in individual toxicokinetics.

This paper integrates our previous results from PCA and cluster analysis on serum, soil and dust, and presents a strategy towards addressing pattern linkage across media.

Materials and Methods

Study Site and Available Dataset: The University of Michigan Dioxin Exposure Study (UMDES) addresses exposures from industrial and background sources, using an age-stratified cohort of residents living near an incinerator outfall, the Tittabawassee River floodplain area, outside of the floodplain, and in areas geographically removed from the Dow facility. The study has access to a database of the WHO29 congeners: 946 serum samples, 764 dust samples, and over 2000 soil samples (collected from 766 properties).

Multivariate Statistical Tools: The general approach to identify fingerprints includes a phased analysis of methods ranging from simple exploratory investigations to more complex targeted analysis. The simpler methods can indicate whether more complex methods are justified (i.e. if more than one discharge is indicated by the variations among samples).

(i) Principal components analysis to reduce the dimensionality of large numbers of correlated variables
(ii) Clustering of the scores to identify the number of independent sources;
(iii) Sensitivity, goodness-of-fit, and cross-validation analysis to separate random from meaningful contributions to sample variance
(iv) Analysis using self-training receptor models (polytopic vector analysis, PVA) to develop chemical fingerprints, source contributions, and mixing proportions.
(v) Sensitivity analysis and jack-knifing to eliminate random contributions and retain statistically robust fingerprints.
(vi) Evaluation of the contribution of alteration processes (e.g. degradation, volatilization) on mixing model outcomes using goodness-of-fit diagnostics[10,11]

This paper presents the results of the PCA and cluster analysis only; the PVA results will be available at the time of the conference.

Mixing models such as PVA resolve 3 parameters of concern; 1) the number of sources (called end-members) in the system; 2) the composition of the end-members; and 3) the relative proportions of each end-member in each sample. Mixing of various source emissions and bio/chemical alteration processes leads to the final mixture of contaminants detected in a sample. In phenomenological terms, PVA un-mixes complex mixtures into the contributing source fingerprints (end-members) and their corresponding source contributions (sample loadings) in each sample. Mathematically, PVA is based on traditional principal component analysis (PCA) and subsequent iterative rotation and shifting of principal component axes until all PC compositions (now called end-members (EMs)) and their sample loadings are positive. Although the link between the mathematical representation (EMs and their loadings) and real world phenomena (fingerprints and their contribution to each sample) is not rigorous, it has been shown repeatedly that the PVA procedure can decompose constructed data sets such that the mathematical representation is in agreement with the phenomenological representation.

Results and Discussion

<u>*Soil samples*</u> were analyzed using principal component analysis (PCA) and hierarchical cluster analysis[12]. All analysis was deterministic rather than probabilistic in nature, and did not take into account region-specific exposures. The most common pattern, characterized by elevated HpCDD, OCDD and PCBs, believed to be due to combustion, was found in soils throughout Jackson/Calhoun (JC) and Midland/Saginaw (MS). The next most

common pattern, characterized by elevated TCDF and 2,3,4,7,8-PeCDF was found mainly in the Floodplain and Near Floodplain. We believe this pattern is the result of Dow's historic discharges into the river. The Midland Plume region, downwind of the Dow plant in Midland, is related to the combustion pattern. The magnitude of the combustion pattern was higher, and 2,3,7,8-TCDD and 1,2,3,7,8-PeCDD were elevated. We believe this pattern is a result of Dow's historic airborne discharges. The Jackson-Calhoun region is similar to other Midland-Saginaw. The highest TEQ values in Jackson/Calhoun were due to contributions of PCBs. The source of this/these pattern(s) is undetermined at this time. There was substantial overlap in the congener contributions to the patterns, in particular for the JC/MS cluster, and the cluster including samples from the Midland Plume region, suggesting that the pattern analysis may require regionalization of the samples prior to analysis. The transitioning between incinerator outfall patterns and background patterns from domestic combustion, and long range atmospheric transport has been observed elsewhere[1].

Principal components analysis and hierarchical cluster analysis were performed on the UMDES *household dust* dataset (n = 764)[13]. Four congeners were dropped from the PCA analysis because 50% or more of their values were below the limit of detection. These congeners were TCDD, 12378-PeCDD, 23789-HxCDF, and 1234789-HpCDF. The largest group (n = 621) was scattered across the entire UMDES population and had a pattern with an undistinguishable source. The second largest group (n = 62) was found mainly in and near the Tittabawassee River floodplain. Its pattern was consistent with the pattern found in the river sediment. The third group (n = 49) was scattered across the entire UMDES population and had a pattern consistent with a combustion source. The apparent similarity between the main dust and soil clusters translated in profiles with overlapping congener contributions, and underscores the need for regionalization of the samples prior to analysis to better correlate soil as a source of dust vs. other (e.g. indoor) sources.

Blood samples taken from a randomized sample of the population surrounding the Dow Chemical plant and the Tittabawassee River were compared to a referent population of individuals living in two removed counties[14]. All 946 lipid-adjusted serum samples collected were considered for the principal components analysis; one outlying observation was dropped leaving a final N = 945 samples. Congeners where 50% or more of the observations fell below the limit of detection were dropped (2378-TCDF, 12378-PeCDF, 123789-HxCDF, 234678-HxCDF, 1234789-HpCDF, OCDF, PCB 81) leaving 22 congeners in the dataset. The results showed that 93% of all the serum observations in this study fall within a single cluster indicating that the pattern of PCDDs, PCDFs and PCBs does not vary substantially from person to person. The second largest cluster (n=48) appears to be driven by age suggesting either a change in exposure for younger individuals or a difference in metabolism related to age. The only clusters that may contain unique personal, regional or occupational activities are minor in size. Interestingly, each cluster exhibited a different median age and age range, suggesting that further resolution of fingerprints may be achieved when samples are age-stratified prior to analysis. DiCaprio et al.[3] demonstrated age-dependence in a multivariate analysis using polytopic vector analysis of PCB patterns in blood serum of an adult Native American population living in an industrialized area, each of which was characterized by a unique mix of congeners. Orloff et al.[2] showed that, using principal component analysis, profiles of dioxin congeners were different in people with elevated dioxin concentrations compared to those with background concentrations. The differences were largely explained by 12378-PeCDD, TCDD, and 1234678-HxCDD, in particular in older individuals. Wingfors et al.[6] indicated country-specific (Spain, Sweden) congener profiles (70 compounds measured) in blood using Soft Independent Modeling of Class Analogy (SIMCA) and Cooman's plot analysis, largely explained by differences in OCDD, 1234678-HpCDD, OCDF, TCDD/F and PCB 169.

Discussion

At the time of this analysis, it was apparent that, whereas PCA revealed pattern similarities between soil and dust, there was no apparent relationship between blood patterns and the three main contributing soil/dust source patterns (floodplain, incineration, background combustion). However, the blood analysis revealed the possibility of patterns being masked by age-dependent elimination processes. Pharmacokinetic processes result in the equilibration of patterns in blood, but the elimination rates and sources of exposure are age-dependent. We anticipate that future

work using age- (and potentially region-) corrected subgroups of the UMDES population will allow a better differentiation between variably exposed populations. Second, the use of a self-training receptor model such as PVA will allow for the extraction of identifiable chemical fingerprints in soil and dust that may help inform serum profiles.

Acknowledgements
Financial support for this study comes from the Dow Chemical Company through an unrestricted grant to the University of Michigan. The authors acknowledge Ms. Sharyn Vantine for her continued assistance and Drs. Linda Birnbaum, Ron Hites, Paolo Boffetta and Marie Haring Sweeney for their guidance as members of our Scientific Advisory Board.

References

1. Leem, J.H., D.S. Lee, and J. Kim. 2006. *Arch. Environ. Contam. Toxicol*. 51: 478-484.
2. Orloff, K.G., D. Hewitt, S. Metcalf, S. Kathman, M. Lewin, and W. Turner. 2001. *J. Expos. Anal. Environ. Epid.* 11: 352-358.
3. DeCaprioa,A.P., G. W. Johnson, A. M. Tarbellc, D. O. Carpentera, J. R. Chiarenzellid, G. S. Morsee, A. L. Santiago-Riverae, M. J. Schymuraf, Akwesasne Task Force on the Environment. 2005. *Environ. Res*. 98: 284–302
4. Cole, D.C., J. Kearney, J.J. Ryan, and A.P. Gilman. 1997. *Chemosphere* 34: 1401-1409.
5. Anderson, H.A., C. Falk, L. Hanrahan, J. Olson, V.W. Burse, L. Needham, D. Paschal, D. Patterson Jr., R.H. Hill, Jr. and the Great Lakes Consortium. 1998. *Environ.Health Perspect*. 106: 279-289.
6. Wingfors, H., G. Lindstrom, B. Van Baval, M. Shuhmacher, and L. Hardell. 2000. *Chemosphere* 40: 1083-1088.
7. Suzuki, G., M. Nakano, and S. Nakano. 2005. *Biosci. Biotechnol. Biochem*. 69: 1836-1847.
8. CDC. 2005. Third National Report on Human Exposure to Environmental Chemicals. NCEH Pub# 05-0725.
9. Johnson, G.W., Ehrlich, R., Full, W., 2002. *In: Morrison, R.D., Murphy, B.L. (Eds.), Introduction to Environmental Forensics*. Academic Press, New York, pp. 461–515.
10. Barabas, N., P. Goovaerts, and P. Adriaens. 2004. *Environ. Sci. Technol*., 38: 1813-1820.
11. Barabas, N., P. Goovaerts, and P. Adriaens. 2004. *Environ. Sci. Technol*., 38: 1821-1827.
12. Chang SC, Adriaens P, Towey T, Wright D, Demond A, Gillespie B, Franzblau A, Garabrant D. Proceed. Dioxin 2006, Oslo, Norway.
13. Zwica L, Chang SC, Towey T, Knutson K, Adriaens P, Demond A, Chen Q, Gillespie BW, Franzblau A, Garabrant D. Proceed. Dioxin 2006, Oslo, Norway.
14. Hedgeman E, Chang SC, Towey T, Demond A, Adriaens P, Chen Q, Franzblau A, Gillespie BW, Sima C, Garabrant D. Proceed. Dioxin 2006, Oslo, Norway.