PRINCIPAL COMPONENTS ANALYSIS OF SERUM PCDDs, PCDFs, AND PCBs FROM A COMMUNITY IN MICHIGAN, USA

Hedgeman E[1], Chang S-C[2], Towey T[2], Demond A[2], Adriaens P[2], Chen Q[3], Franzblau A[1], Gillespie BW[3], Sima C[3], Garabrant D[1]

[1]Department of Environmental Health Sciences, University of Michigan School of Public Health, 109 S Observatory, Ann Arbor, MI 48109; [2]Department of Civil and Environmental Engineering, University of Michigan College of Engineering, 1351 Beal, Ann Arbor, MI 48109; [3]Department of Biostatistics, University of Michigan School of Public Health, 109 Observatory, Ann Arbor, MI 48109

Introduction and Study Goals

The University of Michigan Dioxin Exposure Study (UMDES) was undertaken to determine whether environmental contamination downstream of the Dow Chemical Plant in Midland, MI is contributing to the body-burden of dioxins, furans and PCBs in the surrounding population. To investigate, blood samples taken from a randomized sample of the population living within three counties surrounding the Dow Chemical plant and the Tittabawassee River were compared to a referent population of individuals living in two removed counties. To complement the simple distributional analysis of serum data, an analysis of congener patterns in serum samples was performed using multivariate chemometric methods. Principal component analysis (PCA) and hierarchical cluster analysis (HCA) were performed using the UMDES serum data set. The PCA and HCA output were used to generate a concentration heatmap and cluster-centroid pattern profiles. Initial results suggest that the patterns of congener exposure, uptake, degradation and excretion do not vary greatly in the general population.

Materials and Methods

The final dataset used for the principal components analysis included the results of 946 serum samples. Methods for selecting study participants and obtaining blood samples can be found elsewhere. [1,2,3]

Multiple PCA runs were completed using all 29 congeners as well as subsets including only those congeners with the highest limits of detection.[4] A second set of runs was completed omitting one observation with a unique pattern.

Since congener data exhibited log-normal distributions, a natural logarithm transformation of ln(x+1) was undertaken. A constant-row-sum transformation was used to convert the sum of each row to unity and the natural log transformed concentration of each congener in each sample was converted to a fraction of unity. Finally, a range transformation was applied to each congener column to ensure the variation among observations would be similar. This final step kept the PCA from being driven by congeners with extreme variation.[5]

PCA was performed using Minitab software (Minitab, Inc., State College, Pennsylvania, USA). A Scree plot, a cumulative variance plot, and principal component loading graphs were generated. The principal components that accounted for 95% of the cumulative variance were selected for further use in the HCA.[6,7]

Using the selected principal components, HCA was performed based on a correlation matrix and average linkage of Euclidean distance between samples.[6] Each serum sample in the dataset was assigned a cluster membership. A dendrogram, indicating the similarity between clusters, was generated. Also, two and three-dimensional principal component score plots - grouped by cluster membership, were produced.

The constant-row-sum transformed data (not range transformed) was sorted according to cluster membership. Using the sorted data, a heatmap was generated to represent the congener patterns of all the serum samples in a single graph. In addition, supplemental information items, such as untransformed serum samples exceeding a particular threshold concentration, were indicated in columns adjacent to the congener pattern for each sample.

Minitab software allows for the creation of a distance to cluster-centroid matrix. The sample with the smallest value for the distance to each centroid was selected to represent that cluster. The congener pattern of each cluster centroid was produced using the untransformed serum data to create 100% stacked bar graphs. Both original concentration and TEF-weighted patterns were produced.

Results and Discussion

Results and discussion will be provided after August 15th, 2006.

Acknowledgements

The authors acknowledge the Dow Chemical Company for funding the study and Ms. Sharyn Vantine for her continued assistance.

References

1. Franzblau A, Garabrant D, Adriaens P, Gillespie BW, Demond A, Olson K, Ward B, Hedgeman E, Knutson K, Zwica L, Towey T, Chen Q, Ladronka K, Sinibaldi J, Chang S-C, Lee S-Y, Gwinn D, Sima C, Swan S, Lepkowski J. *Organohalogen Comp* (forthcoming).
2. Lepkowski J, Olson K, Ward B, Ladronka K, Sinibaldi J, Franzblau A, Adriaens P, Gillespie BW, Chang S-C, Chen Q, Demond A, Gwinn D, Hedgeman E, Knutson K, Lee S-Y, Sima C, Swan S, Towey T, Zwica L, Garabrant D. *Organohalogen Comp* (forthcoming).
3. Hedgeman E, Luksemburg W, Patterson D, Knutson K, Franzblau A, Garabrant D. *Organohalogen Comp* (forthcoming).
4. Orloff KG, Hewitt D, Metcalf S, Kathman S, Lewin M, Turner W. *J of Expo Anal & Environ Epid* 2001; 11:352.
5. Johnson GW, Ehrlich R. *Environmental Forensics* 2002; 3: 59.
6. Jolliffe IT. *Principal Component Analysis*. Springer-Verlag, New York. 1986; 8.
7. Hair Jr JF, Black WC. Cluster analysis. In: *Multivariate Data Analysis*, Hair Jr JF, Anderson RE, Tatham RL, Black WC. (ed.), Prentice-Hall Inc, New Jersey, 1998: 147.