PATTERN RECOGNITION ANALYSIS OF VA/EPA PCDD AND PCDF DATA

K. M. Bauer[a], J. S. Stanley[a], J. Remmers[b], J. J. Breen[b], J. Schwemberger[b], B. Schultz[b], and H. K. Kang[c]

[a]Midwest Research Institute, 425 Volker Boulevard, Kansas City, Missouri 64110; [b]U. S. Environmental Protection Agency, 401 M. Street, S.W., Washington, DC 20460; and [c]Department of Veterans Affairs, Washington, DC

ABSTRACT

The objective of this study was to evaluate, by pattern recognition techniques, relationships of PCDDs and PCDFs in human adipose tissue among the analytes themselves and with respect to demographic variables (military/civilian status, age, collection year, race, and geographic region). The conclusions of this study provide a multivariate context for the data and strengthen the conclusions obtained in the univariate study (on 2378-TCDD) previously conducted by the VA and EPA.

From a principal component analysis, the result of most interest, based on a set of matched NHATS specimens (36 Vietnam veterans, 79 non-Vietnam veterans, and 80 civilians), is that Vietnam veterans in this study do not exhibit higher levels of 2378-TCDD, PCDDs, and PCDFs than the other two groups. Also, cluster analysis of the PCDDs and PCDFs demonstrates that the 2378-TCDD is not strongly correlated with any of the other compounds, indicating possible differences in exposure routes to these chemicals.

KEY WORDS

Pattern recognition, multivariate statistical analysis, cluster analysis, principal component analysis, adipose tissue, dioxins, furans.

INTRODUCTION AND BACKGROUND

This study presents the results and interpretation of a series of multivariate statistical analyses performed on a data base of levels of polychlorinated dibenzo-p-dioxins (PCDDs) and dibenzofurans (PCDFs) in human adipose tissue. The objective of this study is to provide a means of evaluating the relationships between, and significance of, chemical and demographic variables and military/civilian status of the individuals represented in this data base. This work is to provide a complementary examination of the previous univariate analyses performed by the VA and USEPA (1989).

Under a joint VA/USEPA agreement, approximately 200 human adipose tissue specimens were analyzed for PCDDs and PCDFs. These tissue samples were obtained from the National Human Adipose Tissue Survey (NHATS). The availability of military service status on males born between 1936 and 1954 was the basis for the selection of specimens in the three groups--Vietnam veterans, non-Vietnam veterans, and civilians. To the extent that the NHATS population is not representative of Vietnam veterans, the results are not representative of TCDD levels in the population of Vietnam veterans. The relatively small sample size (36) of Vietnam veterans would limit the significance of results relating to TCDD levels in the population of Vietnam veterans in any case. However, within the restricted NHATS population, the comparative results are valid.

Since the data had been reported for all 2378-substituted compounds, some effort was needed for the assessment of relationships between compounds and with respect to demographic data. Pattern recognition was proposed as an exploratory analysis tool. This approach--investigation of potential multidimensional relationships among the analyte results--was followed by classification of the subjects into groups, if groups could be identified. The VA provided demographic information on these subjects. After completion of all statistical analyses, the affiliation of each donor to one of the three groups was made available by the VA so that possible groupings of the results could be identified that would parallel the three groups of donors.

APPROACH

Multivariate techniques, or pattern recognition techniques, have been developed and provide rules for combining variables in an optimal way (the criterion of optimality varies from one technique to another) so as to reduce the dimensionality of the set of variables with as little loss of information as possible. In this context, two statistical techniques have been used to complement each other: cluster analysis and principal component analysis.

There are three major objectives in pattern recognition: data reduction, feature extraction, and classification. In cluster analysis, in which little a priori knowledge of possible grouping of variables is required, natural groupings and structures among variables are studied without prejudice. The same technique can be applied to the observations (specimens) in the study. In principal component analysis, the goal is to define linear combinations of the original variables and thus to reduce the dimension of the data space.

The Biomedical (BMDP) and Statistical Analysis System (SAS) software packages have been used exclusively in this study. Also, all concentration data were log-transformed prior to analysis.

RESULTS

Figure 1 and the following points highlight the cluster analysis results:

- 2378-TCDD does not group with any of the other analytes--its behavior is unlike the behavior of the others.
- Apart from 2378-TCDD and those analytes with relatively high frequencies of nondetects, the levels of the higher chlorinated PCDDs and PCDFs are strongly interrelated. All correlations are high and positive, that is, the levels of these analytes are proportional to each other.

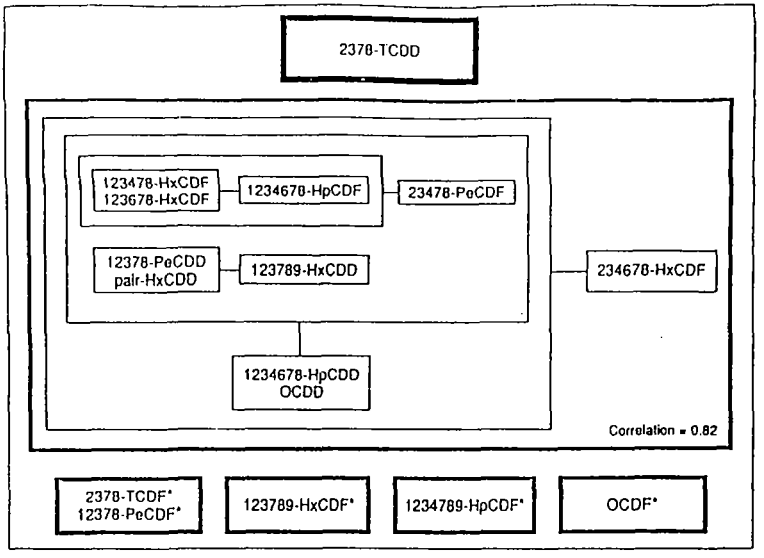
The following conclusions can be drawn from the principal component analysis:

- None of the analytes, demographic variables, or military service/civilian status provided any means of classifying the NHATS specimens into more than one large group. That is, no pattern-distinguishing--for example, between Vietnam veterans, non-Vietnam veterans, and civilians--was visible on a plot of specimen scores along the most important principal components.
- A large proportion (77%) of the variability in the total data set, as determined by the values of the NHATS specimens on all the variables, is explained by the combination of two principal components. The first and predominant principal component (explaining 64% of the total variability) is a combination of the higher chlorinated PCDDs and PCDFs, each analyte having about equal contribution to this principal component. This is consistent with the results of the cluster analysis. The second, and by far less important principal component (13% of total variability explained), consists mainly of age of specimen at time of death and specimen collection year (see Table 1). A third principal component (6.4% of total variance) consists of lower chlorinated PCDDs and PCDFs, including 2378-TCDD.

CONCLUSIONS

The most important outcome from the analyses performed in this study is that there is no visible stratification of the NHATS specimens in groups as defined by their military service or civilian status with respect to either the PCDD and PCDF levels or the demographic variables. In other words, based on the set of matched NHATS specimens, Vietnam veterans in this study do not exhibit higher levels of PCDDs and PCDFs, including 2378-TCDD, than either the non-Vietnam veterans or the control group of civilians.

The results of these multivariate statistical analysis efforts provide evidence of interrelationships of PCDDs and PCDFs that could be possible indicators of sources of exposure. That is, the source of 2378-TCDD may be completely different from those of the other PCDDs and PCDFs. Further statistical evaluations of body burden data may require more comprehensive data bases that identify body burden half-life of the higher chlorinated PCDDs and PCDFs vs. 2378-TCDD, as well as the potential contributions arising from commercial products and environmental matrices (e.g., incinerator emissions, ambient air, water, and fish).



* Analytes with high occurrences of non-detects (over 25%)

Note: The relative position of the analytes in this figure is for illustrative purposes only.

Each rectangle outlines a cluster. One cluster may consist of sub-clusters and individual analytes.

The 6 clusters outlined by heavy dark lines are those formed when a 0.80 cutoff is selected for the similarity measure between clusters.

The 10 analytes within the large rectangle are all highly inter-correlated.

The 4 rectangles at the bottom contain those analytes detected in less than 75% of the specimens.

The cluster on top shows 2378-TCDD, which does not significantly correlate with any other analyte or group of analytes.

Figure 1. Cluster analysis results of all 16 analytes based on the correlations of the log-transformed concentrations.

Table 1. Principal Component Analysis Results

Principal Component 1 (64.4% of total variance explained)
"Higher Chlorinated Furans and Dioxins"

Variable (Analyte No.)*1 (ordered by % variance)		Variable Loading	Variance Explained (%)	Cumulative Variance (%)
123478-HxCDF	(6)	0.94	10.6	10.6
123789-HxCDD	(11)	0.92	10.0	20.6
123678-HxCDF	(7)	0.91	10.0	30.6
123478/123678-HxCDD	(10)	0.90	9.7	40.3
OCDD	(16)	0.88	9.2	49.5
1234678-HpCDF	(12)	0.87	9.0	58.5
1234678-HpCDD	(14)	0.86	8.8	67.3
12378-PeCDD	(5)	0.85	8.7	76.0
23478-PeCDF	(4)	0.83	8.3	84.3

Principal Component 2 (12.7% of total variance explained)
"Age and Collection Year"

Variable (Analyte No.)*1 (ordered by % variance)	Variable Loading	Variance Explained (%)	Cumulative Variance (%)
Age at Sampling	0.92	51.7	51.7
Collection Year	0.84	42.5	94.2

Principal Component 3 (6.39% of total variance explained)
"Lower Chlorinated Furans and Dioxins"

Variable (Analyte No.)*1 (ordered by % variance)		Variable Loading	Variance Explained (%)	Cumulative Variance (%)
234678-HxCDF	(8)	0.43	22.1	22.1
1234678-HpCDD	(14)	0.37	16.9	38.9
12378-PeCDD	(5)	-0.37	16.7	55.7
2378-TCDD	(2)	-0.37	16.6	72.3

Note 1*: All concentrations were log-transformed
Note 2: Only variables with a variance contribution above average (1/13=7.7%) are shown within each principal component

REFERENCES

USEPA. 1986. U.S. Environmental Protection Agency, Office of Toxic Substances. Analysis for polychlorinated dibenzo-p-dioxins and dibenzofurans in human adipose tissue: method evaluation study. EPA 560/5-85-022.

VA and USEPA. 1989. The Department of Veterans Affairs, Office of Environmental Epidemiology, and the U.S. Environmental Protection Agency, Office of Toxic Substances. Dioxins and dibenzofurans in adipose tissue of U.S. Vietnam Veterans and controls. EPA 560/5-89-002.

Stanley, J. S., R. E. Ayling, P. H. Cramer, K. R. Thornburg, J. C. Remmers, J. J. Breen, J. Schwemberger, H. K. Kang, K. Watanabe, Chemosphere, Proceedings of the 9th International Dioxin Conference, Toronto, Canada, 1989 (in press.)

VA and USEPA. 1990. The Department of Veterans Affairs, Office of Environmental Epidemiology, and the U.S. Environmental Protection Agency, Office of Toxic Substances. Pattern Recognition Analysis of VA/EPA PCDD and PCDF Data. EPA 560/5-90-006.