Bull. Environ. Contam. Toxicol. (1988) 40:641–646

Bulletin of Environmental Contamination and Toxicology

Multivariate Statistical Approach to a Data Set of Dioxin and Furan Contaminations in Human Milk

Gunilla U. M. Lindström,[1] Michael Sjöström,[1] Stephen E. Swanson,[1] Peter Fürst,[2] Christiane Krüger,[2] Hans-Albert Meemken,[2] and Wilhelm Groebel[2]

[1]Department of Organic Chemistry, University of Umeå, S-901 87 Umeå, Sweden and [2]Chemisches Landes-Untersuchungsamt, NordrheinWestfalen, Sperlichstr. 19, 4400 Münster, BRD

The levels of chlorinated dibenzodioxins, PCDDs, and dibenzofurans, PCDFs, in human milk have been of great concern after the discovery of the toxic 2,3,7,8-substituted isomers in milk of European origin (WHO 1985). The toxicity of the different 2,3,7,8-isomers varies. In estimating the total toxicity of PCDDs and PCDFs in milk all isomers may be transformed to TCDD-equivalents representing the toxicity of 2,3,7,8-TCDD. The amounts of dioxins and furans in human milk have been found to exceed the tolerable daily intake values, 1-5pg TCDD-equivalents per kg of body weight and day for an infant beeing nursed (Nygren et al. 1986).

As knowledge of environmental contamination of human breast milk increases, questions will continue to be asked about possible risks from breast feeding. Before any recommendations can be made, there must be knowledge of contaminant levels in mothers' breast milk. Since testing of all mothers is not feasible it is urgent to find potential risk groups of mothers.

Fürst et al. (1986a) have earlier measured PCB and 17 different dioxins and furans in human breast milk samples. Personalia of the mothers are known. To date the data has only been analyzed by univariate and bivariate statistical methods (Fürst et al. 1986b). However to extract as much information as possible from this data set, multivariate statistical methods (see e.g. Mardia et al. 1979) must be used. Here we present a multivariate analysis where the relationships between the polychlorinated compounds and the personalia of the mothers have been studied.

For the data analysis partial least squares (PLS) modelling (Wold et al. 1984, 1985) has been used.

Send reprint requests to G. Lindström at the above address.

MATERIALS AND METHODS

The isomer specific analyses of PCDDs and PCDFs in the milk samples were performed by GC-MS (NCI) after clean-up procedure including extraction of the milk fat and further purification of this fraction by GPC on a Florisil column. For a detailed description of the methodology (Fürst et al. 1986a and Fürst et al. unpublished work 1987).

All calculations were performed with the SIMCA 3-B program on an IBM PC microcomputer (SIMCA 3-B is available from SEPANOVA AB, S-122 43 Enskede, SWEDEN or Principal Data Components, Columbia, 65201 Mo. USA). PLS and SIMCA 3-B system have been described in detail elsewhere (Wold 1984). Hence only a short presentation will be given here.

Chemical investigations frequently produce multivariate data, i.e. data arranged in tables or matrices, as the present data set, see figure 1. With PLS the causal relationships between two data matrices X and Y with the same samples can be investigated. In PLS each of the matrices is decomposed in two low dimensional matrices T and P for the X block and U and Q for the Y block. The causal relationship between the X block and Y block is expressed by mutual relationships between columns in T and U. The number of significant dimensions A, i.e. the number of columns in the T and U matrices and rows in the P and Q matrices, extracted from the X and Y block, is determined with cross-validation (Wold et al. 1984). Usually the number of significant dimensions are low. If just one component is significant as in the present investigation (see Results) T and U are just column vectors (t(1) and u(1)) and P and Q are row vectors (p(1) and q(1)), see figure 1. The size of the absolute value of an element in the p and q loading vectors describes the contribution of the variable to the t and u score vectors and the sign shows whether the variable is negatively or positively correlated to t or u. A plot of the u vector against the t vector reveals the causal relationship between the Y and X block.

The analyzed data set consists of 39 milk samples from lactating women. The samples were characterized by 15 chemical measurements and 7 additional variables which described the personalia of the mothers. Thus the data could be represented by a 39X15 matrix, Y, and a 39X7 matrix, X, see figure 1. In the original data set reported by Fürst et al., 17

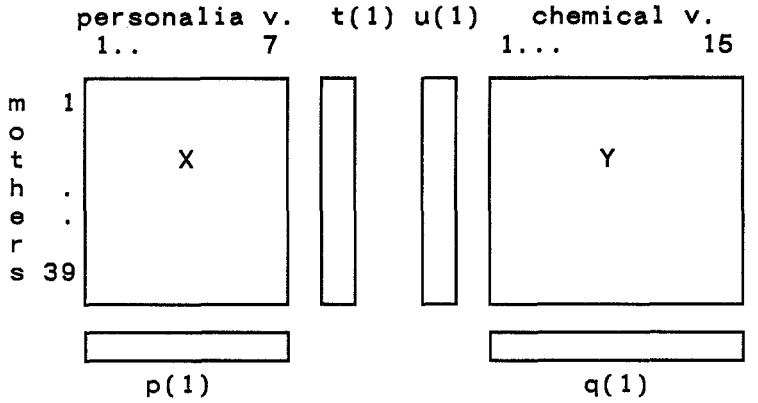

Figure 1. Schematic representation of the present multivariate data set. The score (t and u) and loading (p and q) vectors of the first PLS dimension are also shown.

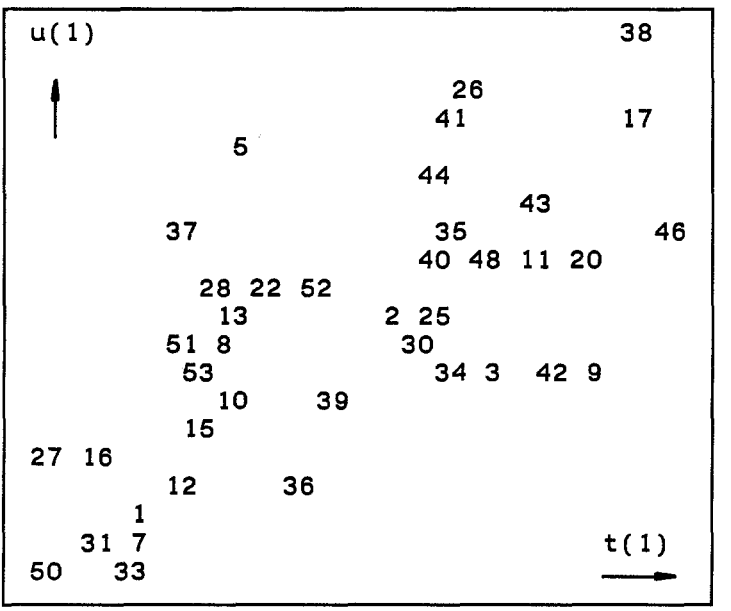

Figure 2. The latent variable t(1) from the personalia data block plotted against the latent variable u(1) from the chemical data. The plot shows the correlation between the X and Y blocks. Each number (see Fürst et al. (1986a)) represents a mother.

chemical variables for 53 mothers were reported. However, personalia was only recorded for the 39 samples here used. Two chemical variables were not used in this study due to the fact that their levels were below the detection limit in most samples analyzed.

Prior to the PLS analysis, established data preprocessing of the data was performed. Thus due to skewed distributions of the chemical data, logarithms of the original measurements were used. Further, all variables were scaled and the variable means were subtracted from the data.

Table I. The variables and loadings p and q for the Y and X blocks used in the PLS analysis.

	variables	loadings	m. power*
1	OCDD	.23	.11
2	1234678-HpCDD	.26	.14
3	123789-HxCDD	.32	.24
4	123678-HxCDD	.31	.22
5	123478-HxCDD	.33	.26
6	sum of HxCDDs	.33	.27
7	12378-PeCDD	.29	.18
8	TCDD equivalents	.23	.10
9	OCDF	.10	.01
10	1234678-HpCDF	.21	.09
11	123478-HxCDF	.24	.12
12	123678-HxCDF	.21	.08
13	234678-HxCDF	.23	.11
14	23478-PeCDF	.20	.08
15	PCB	.28	.17
16	fat % in milk	.32	.08
17	mothers fat status	-.13	.00
18	place of residence	.08	.00
19	number of children	-.54	.28
20	lactation time	-.41	.14
21	mothers age	.21	.00
22	weight of child	-.60	.34

*A modelling power equal to zero means that no variation in the variable is described by the model and a modelling power equal to one means that all variation in the variable is described by the model.

RESULTS AND DISCUSSION

In the PLS-analysis the chemical variables 1-15 were used as Y and the personalia variables (16-22) as X. Thus an optimal model was calculated with respect to prediction of the chemical variables from the personalia of the mothers. Cross-validation showed that just one latent variable from the personalia data block (X) contained predictive information on the chemical variables. A plot of the latent variable from the X block, t(1), against the latent variable, u(1), from the Y block is shown in figure 2.

The PLS model describes about 10% and 27% of the variance in the X and Y blocks, showing that there is little overall relationship between the personalia variables and chemical variables. With a separate principal components analysis (Mardia 1979) of the chemical data block Y, we have confirmed that a two components model describes about 75% of the variance in this block. The number of significant components were determined with cross-validation. This shows the high precision in the chemical analysis and that the personalia variables can describe only one third of the systematic variation of the dioxins and furans in the samples.

In the PLS analysis variables 19, 20 and 22 contribute much (50%, 24% and 57% of variance explained) to the t vector and the dioxins (variables 2-7) and PCB (variable 15) contribute most (30%-50% of variance explained) to the u vector. From the sign of the loadings it is seen that the major contributing personalia variables are negatively correlated to the major chemical variables. For example the concentration of the dioxins decreases with increasing number of lactation periods. We also note that whether the mothers lived in an urbanized area or a rural environment does not contribute to the model.

From the analysis, it was detected that for the studied lactation period the two-children mothers at the time of the collection of the milk samples had about twice as long lactation times as the one-child mothers. This shows that the sampling can be improved and that we can not directly from this data set draw the conclusion that the concentration of dioxins decreases with the number of children.

The data can also be analyzed with separate models for one and two child mothers. However, these analyses have not been dealt with in this report.

REFERENCES

Fürst P, Meemken H-A, Groebel W (1986a) Determination of polychlorinated dibenzodioxins and dibenzofurans in human milk. Chemsphere 15:1977-1980

Fürst P, Meemken H-A, Krüger C, Groebel W (1986b) Polychlorinated dibenzodioxines in milk samples from Western Germany, 6th international dioxin-symposium, Fukuoka

Mardia KW, Kent JT, Bibby JN (1979) Multivariate analysis. Academic Press, New York

Nygren M, Rappe C, Lindström G, Hansson M, Bergqvist P-A, Marklund S, Domellöf L, Hardell L, Olsson M (1986) Identification of 2,3,7,8-substituted Polychlorinated Dioxins and Dibenzofuranes in Environmental and Human Samples. In: Rappe C, Choudhary G, Keith LH (eds) Chlorinated Dioxins and Dibenzofurans in Perspective. Lewis Publisher, Chelsea, Michigan. p 17

WHO (1985) Organohalogen compounds in human milk and related hazards. Report on a WHO Consultation. WHO Regional Office for Europe, ICP/CEH 501/m 05

Wold S, Albano C, Dunn III WJ, Edlund U, Esbendsen K, Geladi P, Hellberg S, Johansson E, Lindberg W, Sjöström M (1984) Multivariate Data Analysis. In: Kowalski BR (ed) Chemometrics: Mathematics and Statistics in Chemistry, NATO ASI Series, Series C: Mathematical and Physical Sciences Vol 138. Reidel Publishing Company, Dordrecht, Holland. p 17

Wold S, Dunn III WJ, Hellberg S. (1985) Toxicity modelling and prediction with pattern recognition. Environ Health Pers 61:257-268

Recieved October 15,1987; accepted January 18, 1988.