# Clusters of Kin Analytes: Detection Thresholds of Individual Components and Representativeness of Cumulative Results

Alessandro di Domenico, Carlo Lupi, Elena De Felip, Fabiola Ferri, Nicola Iacovella, Roberto Miniero, Emerenziana Scotto di Tella, Paola Tafani, Luigi Turrio Baldassarri, and Fabrizio Volpi

Laboratory of Comparative Toxicology and Ecotoxicology, Istituto Superiore di Sanità, Viale Regina Elena 299, 00161 Rome, Italy

## BACKGROUND

In recent years we have repeatedly investigated whether a correlation existed between PCB (polychlorobiphenyl) levels detected in environmental soil samples and the parallel PCDD and PCDF (polychlorinated dibenzo-*p*-dioxin and dibenzofuran, respectively) TE values found in the same matrices.[1-6] Cumulative PCB amounts were obtained from the assessment of tri- through heptachlorosubstituted homologues, whereas cumulative PCDD and PCDF figures resulted from quantitation of individual tetra- through octachlorosubstituted congeners followed by conversion into international TCDD equivalents—TE units—through the current I-TEF system.[7] In all cases, analytical preparative and GC-MS(SIM) procedures were practiced in compliance with common GLP and QC/QA protocols.[1,8,9]

Given the analytical findings from a set of congruous soil matrices, regression analyses were carried out by associating two "chemical" coordinates to each sample: X = ***ln*** [PCB, in ng/kg] and Y = ***ln*** [PCDD+PCDF, in ngTE/kg]. Since each cumulative figure came from a cluster of closely related analytes—that is, five PCB homologues and 17 PCDD and PCDF 2,3,7,8-congeners—it included the contribution of below-detection-threshold data, if any: in fact, the latter were entered in calculations as half the nominal value of the pertinent threshold. Therefore, before performing regression analysis or other forms of data elaboration, the question was raised how much a given cumulative figure of an analytical species—namely, PCBs or PCDDs and PCDFs—was representative of the amount estimated to be present in that given sample, so that only data with an adequate level of representativeness could be selected. A few approaches developed to meet the above question are briefly described hereafter.

## EXPERIMENTAL

***Polychlorobiphenyls*** The term representativeness was first introduced for PCBs determined in the contaminated soil of an industrial site.[1,2,6,8] In fact, to speed up analytical operations, a homologue-specific detection bias was set at 500 ng/kg (of dry matrix), regardless of the real detection sensitivity attainable. Since the latter was over one order of magnitude below the aforecited value, the final outcome of the assessment was substantially independent of matrix size (≈0.1–2 kg). To estimate a cumulative concentration, all homologue data expressed as <500 ng/kg were entered as 250 ng/kg, as said. Therefore, the question was posed of how many homologues—of the five to be measured—should be quantified (≥500 ng/kg) for the final cumulative analytical figure to be representative of PCB level estimated to be present in the sample. The outcome of the statistical approach utilized is shown in Figure 1.

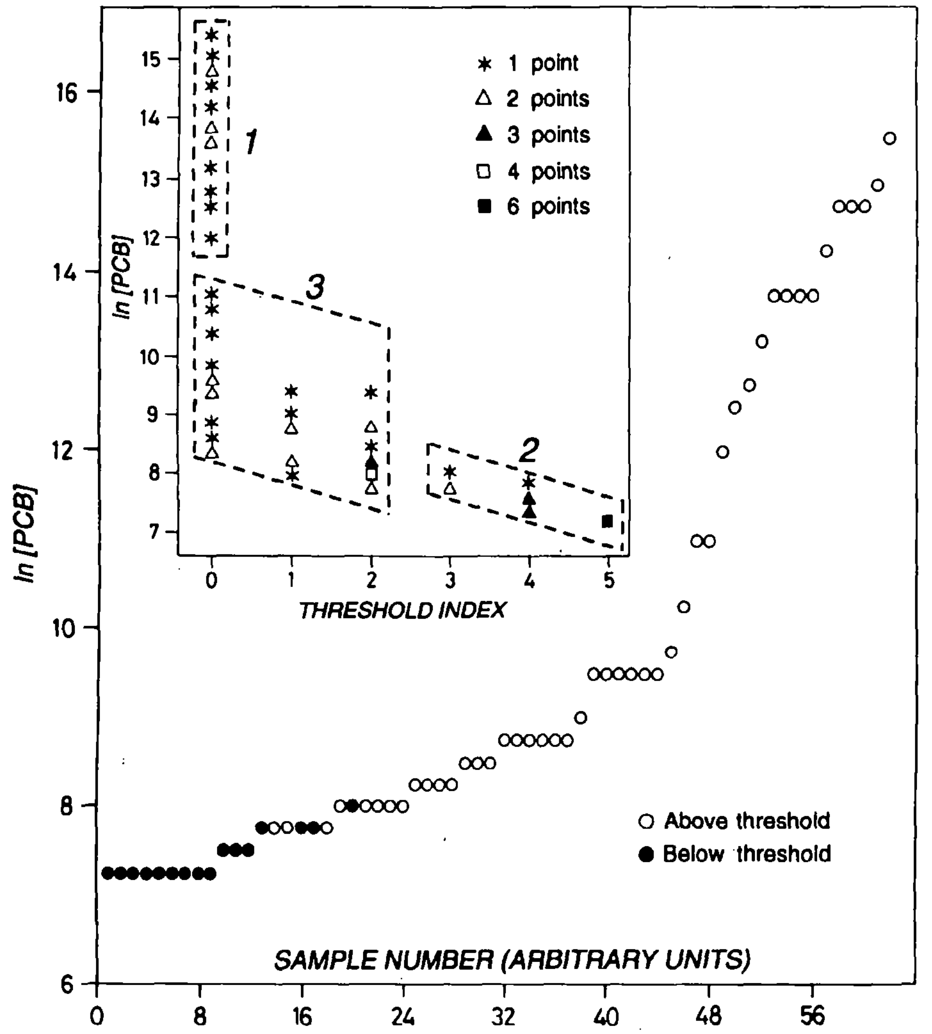


*Figure 1* Cluster analysis (inset) of 62 PCB data (range, 1200–5,000,000 ng/kg) from the assay of 43 environmental[1,6] and 19 blank[1] soil matrices. The latter were obtained by reprocessing the environmental soil already analyzed; this matrix retained some (<3 %) of original PCBs. Each cumulative figure was log-transformed and associated with a "threshold index" indicating how many homologues were not determined (<500 ng/kg); the index assumed integer values between 0 and 5 (all homologues not determined). Of the three clusters composed, Cluster 2 was deemed to be characterized by data with insufficient representativeness (only two or less homologues assessed). In the larger frame, the data available are ordered according to an increasing trend: below-representativeness-threshold values are clearly visible. By combining the cluster analysis outcome and the shown trend, the threshold was estimated between 7.7 and 8.1 (2200 and 3300 ng/kg, respectively) and conservatively set at ≈3000 ng/kg. In the inset, marks indicate that more data may come in at the same point due to lack of graphic resolution.

***Polychlorinated dibenzo-p-dioxins and dibenzofurans*** Contrary to PCBs, PCDD and PCDF congeners frequently happened to go unquantified for lack of detection sensitivity.[1] Since matrix weight had a significant bearing on that, analytical results associated with sampling outliers (dry sample matrix, <0.300 kg) were subjected to an evaluation of acceptability prior to being used.[1,4,5] The selected cumulative data were then grouped into Sets A (N = 15; range, 2.1–130 ngTE/kg) and B (N = 26; range, 3.7–3400 ngTE/kg), respectively associated with soil matrices from the general environment neighboring the aforecited industrial site and from the industrial site itself;[1,6] TCDD detection threshold appeared to vary mostly between 2.2 and 6.0 ng/kg (S/N ≈ 3). *Because PCDDs and PCDFs were never quantified in blank soil matrices,*[1] *these have no* bearing in this text.

Representativeness thresholds were determined by various statistical approaches, as summarized in Figures 2–4 (for all estimates, a $P \leq 0.05$ was selected). In all cases, each cumulative TE datum was associated with a threshold index obtained by summing up the pertinent I-TEF values of below-detection-threshold congeners—the contributions of quantified PCDDs and PCDFs being set equal to 0—so that the index ranged from 0 to 2.882 (only threshold data available).

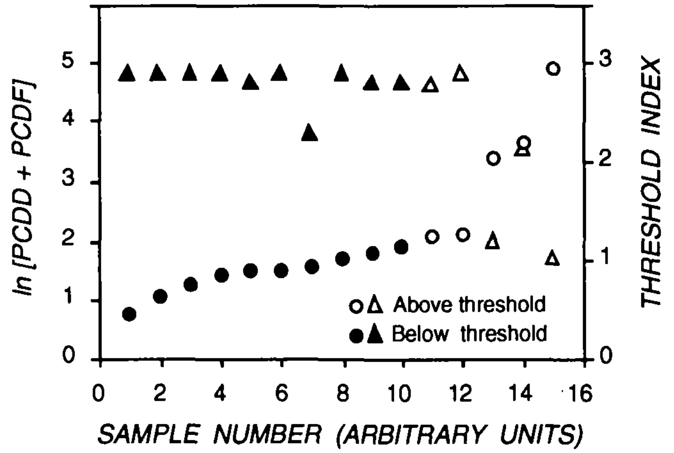


***Figure 2*** Set A layout. *The original data* (ngTE/kg) *were log-transformed and arrayed* according to their increasing trend (O); the paired threshold index figures are also exhibited (Δ). Up to approximately Sample 12, the TE trend appears to be rather flat: this feature may be characteristic of an analytical procedure which has a loss of response sensitivity and, in fact, in this case is associated with a large number of below-detection-threshold contributions (as clearly indicated by the high index values). Moving farther to the right, the TE trend shows an evident upslope, as the index dips to lower values due to an increase of the number of PCDDs and PCDFs quantified. To estimate the representativeness threshold, a line parallel to the X-axis—and set below the top TE datum—was moved down stepwise: at each step, an additional TE datum was included in the upper group (Subset A1), whereas the lower group (Subset A2) lost one. At each step, the index means associated with Subsets A1 and A2 were re-estimated and their difference evaluated by means of Student's *t* test, until the chosen value of *P* was reached. It may be noted that *P* started out at <<0.05, this indicating that the threshold was—as expected—below the top TE datum (however, that line setting *did* eliminate too many TE data!). The threshold was thereby estimated at approximately 1.9 (6.7 ngTE/kg). The statistical approach described failed to yield a threshold estimate for Set B.

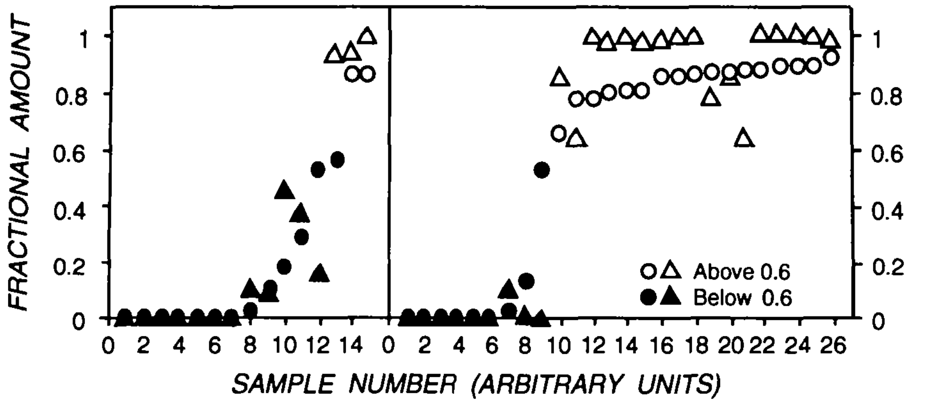


***Figure 3a*** In this display, the representativeness associated to each sample (Set A, left; Set B, right) is indicated by the relative contribution—$f$, fractional amount—of the PCDD and PCDF levels above detection threshold.[5,10] Therefore, by definition $f$ varies between 100 and 0 %, the latter value being associated with virtual contributions alone (only threshold data available). Due to the way we carry out readings of near-detection-threshold signals, the virtual contribution of data below detection threshold determines a final cumulative figure to be overestimated rather than underestimated. Generally, the overestimate increases as the physical meaning of the measure ($f$) decreases: for $f < \approx 60$ %, cumulative figures should be regarded as upper estimates. The figure shows the paired distributions of $f_A$ (O) and $f_{TE}$ ($\Delta$), the fractional amounts estimated from cumulative data expressed in ng/kg (sum of 10 homologue findings) and ngTE/kg, respectively. The distributions are ordered according to the increasing values of $f_A$.

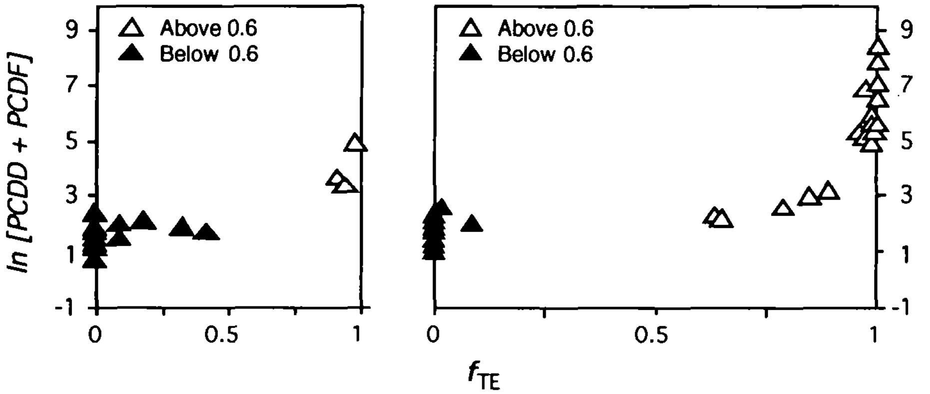


***Figure 3b*** Scattergraphs of Set A (left) and B (right) cumulative PCDD and PCDF TE data. The position of each point is determined by the pertinent $f_{TE}$ value (X-axis) and its paired log-transformed figure expressed in ngTE/kg (Y-axis). From both displays (3a and 3b) and with reference to the limit previously indicated ($f \approx 60$ %), representativeness thresholds may be readily estimated in a more or less conservative way. The subsidiary application of the Wald-Wolfowitz nonparametric test, as described in Figure 4, allowed to associate a value of $P$ and, possibly, redirect threshold selection if required.[5] According to the approach outlined, for both sets thresholds were set at 2.1 (8.2 ngTE/kg). The Wald-Wolfowitz test was also successfully applied to each pair of $f_{TE}$ subsets.

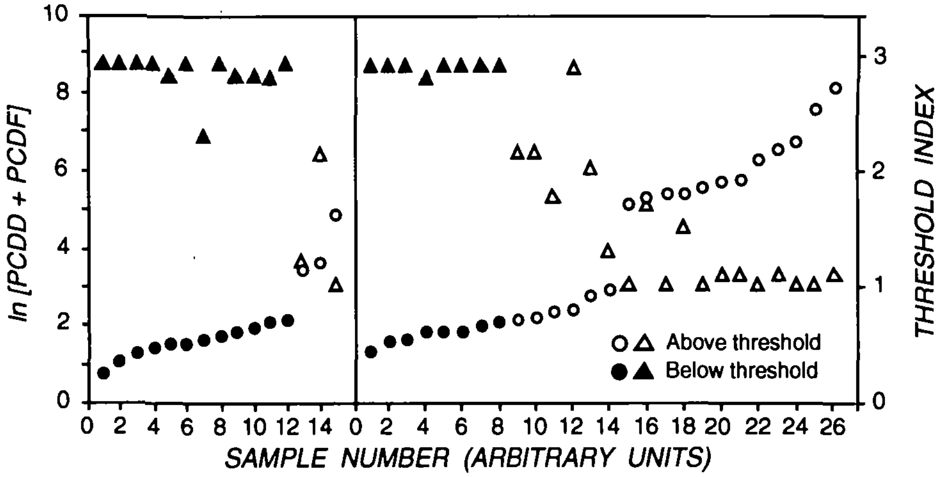


***Figure 4*** Set A (left) and B (right) TE data are shown—log-transformed and laid out according to their pertinent increasing trends (O)—together with the paired threshold indexes (△). Reference should be made to Figure 2 for general information and hints on the statistical approach. As previously seen, a moving horizontal line was used to obtain Subsets A1 and B1 (and the respective complementary Subsets A2 and B2). Again, as described a stepwise procedure was carried out: at each step, the difference $z$ between the index subsets of the pair being investigated (A1-A2 or B1-B2) was estimated by the Wald-Wolfowitz nonparametric test and associated to a $P$ level. For each set, the process was repeated until the selected value of $P$ was reached. For both sets, the representativeness thresholds were estimated at approximately 2.1 (8.2 ngTE/kg). For Set A, this estimate is in good agreement with the value reported in Figure 2, which however appears to be not as conservative. The application of nonparametric statistics has been successfully reported by di Domenico *et al.*,[4,5] although applied to different sets of data.

## CONCLUSIONS

The use of statistical means to estimate PCB representativeness threshold was required due to the specific analytical approach we adopted to determine the chemicals in a group of industrial soil samples. However, PCB levels in environmental matrices are normally high enough to yield representative cumulative results *almost* regardless of sample size. The particular case of TCDD-like PCBs was not investigated.

PCDDs and PCDFs are known to be present in environmental matrices at levels generally much lower than those assessed for PCBs. Therefore, sample size may have an important bearing on cumulative (analytical or TE) finding representativeness, a fact calling for a threshold evaluation strategy. The statistical approaches described in this paper were applied pragmatically to specific cases. However, a study to gain more insight and clarify some shortcomings is presently underway (Lupi *et al.*, in preparation). For example, the various approaches correctly yielded comparable threshold estimates; however, testing the difference of two means by Student's $t$ distribution failed for Set B (Figure 2), for which instead a more flexible nonparametric testing procedure did succeed (Figure 4). Further, the TE level (11.2 ngTE/kg) of Sample 8 of Set B in Figure 3a was qualified representative (>8.2 ngTE/kg; Figures 3b and 4) in spite of the very low $f_A$ and $f_{TE}$ values (<<60 %) and of the high threshold index (2.880; Sample 8 becomes Sample 12 in

Figure 4). Threshold evaluation was also successfully carried out on sets of TE data obtained by using the 1987 EPA-TEF conversion system.[7]

REFERENCES

*1* di Domenico A., E. De Felip, F. Ferri, A. Fulgenzi, N. Iacovella, C. La Rocca, R. Miniero, F. Rodriguez, E. Scotto di Tella, S. Silvestri, P. Tafani, L. Turrio Baldassarri, and F. Volpi (1991): Accertamenti Analitici per il Rilevamento di Policlorodibenzodiossine e Policlorodibenzofurani in Campioni Ambientali Prelevati presso l'ACNA di Cengio e nell'Ambiente Circostante - Parte I. Report for the Ministry of the Environment, Rome (Italy).

*2* De Felip E., A. di Domenico, F. Ferri, N. Iacovella, R. Miniero, E. Scotto di Tella, S. Silvestri, P. Tafani, L. Turrio Baldassarri, and F. Volpi (1992): An Outline of Approaches and Criteria to Investigate Soil Contamination of an Industrial Site and Some Preliminary Results. *Fresenius Environmental Bulletin* **1**, 565-570.

*3* di Domenico A., F. Ferri, A.R. Fulgenzi, N. Iacovella, C. La Rocca, R. Miniero, F. Rodriguez, E. Scotto di Tella, S. Silvestri, P. Tafani, L. Turrio Baldassarri, and F. Volpi (1993): Polychlorinated Biphenyl, Dibenzodioxin, and Dibenzofuran Occurrence in the General Environment in Italy. *Chemosphere* **27**, 83-90.

*4* di Domenico A., F. Ferri, N. Iacovella, C. La Rocca, C. Lupi, R. Miniero, F. Rodriguez, E. Scotto di Tella, S. Silvestri, L. Turrio Baldassarri, and F. Volpi (1993): Accertamenti Analitici per il Rilevamento di Policlorodibenzodiossine e Policlorodibenzofurani in Campioni Ambientali Prelevati presso l'ACNA di Cengio e nell'Ambiente Circostante - Area con Priorità 01 - Parte II. Report for the Ministry of the Environment, Rome (Italy).

*5* di Domenico A., E. De Felip, F. Ferri, N. Iacovella, C. La Rocca, C. Lupi, R. Miniero, F. Rodriguez, E. Scotto di Tella, S. Silvestri, L. Turrio Baldassarri, and F. Volpi (1994): Accertamenti Analitici per il Rilevamento di Policlorodibenzodiossine e Policlorodibenzofurani in Campioni Ambientali Prelevati presso l'ACNA di Cengio e nell'Ambiente Circostante - Aree con Priorità 02 e 03 - Parte III. Report for the Ministry of the Environment, Rome (Italy).

*6* Miniero R., E. De Felip, A. di Domenico, F. Ferri, A. Fulgenzi, N. Iacovella, C. La Rocca, F. Rodriguez, E. Scotto di Tella, S. Silvestri, P. Tafani, L. Turrio Baldassarri, and F. Volpi (1994): PCDDs, PCDFs, and PCBs in the Soil of an Industrial Site. II. Preliminary Assessment. *Fresenius Environmental Bulletin* **3**, 339-344.

*7* US EPA (1989): *Interim Procedures for Estimating Risks Associated with Exposures to Mixtures of Chlorinated Dibenzo-p-dioxins and Dibenzofurans (CDDs and CDFs) and 1989 Update*. EPA/625/3-89/016, Risk Assessment Forum, Environmental Protection Agency, Washington (DC).

*8* di Domenico A., E. De Felip, F. Ferri, N. Iacovella, R. Miniero, E. Scotto di Tella, P. Tafani, and L. Turrio Baldassarri (1992): Determination of the Composition of Complex Chemical Mixtures in the Soil of an Industrial Site. *Microchemical Journal* **46**, 48-81.

*9* Turrio Baldassarri L., E. De Felip, A. di Domenico, F. Ferri, N. Iacovella, O.A. Longo, R. Miniero, E. Scotto di Tella, and F. Volpi (1994): PCDDs, PCDFs, and PCBs in the Soil of an Industrial Site. I. Outline and Evaluation of the Analytical Method. *Fresenius Environmental Bulletin* **3**, 281-286.

*10* De Felip E., A. di Domenico, M. Falleni, F. Ferri, N. Iacovella, G. Menale, P. Tafani, G. Tommasino, and L. Turrio Baldassarri (1994): Polychlorodibenzodioxin and Polychlorodibenzofuran Levels in Dielectric Fluids Containing Polychlorobiphenyls. *Toxicological and Environmental Chemistry* **46**, 239-260.