POPULATION AND SAMPLE BASED STATISTICAL ANALYSIS FOR NHANES PCB-153 AND 180 DATA

William Mills[1], Colleen Simpson[2], Courtney D. Sandau[3]

[1]Mills Consulting Inc.
[2]Jake's Whitford limited
[3]Jacques Whitford limited

Introduction

The Second National Exposure Report (NER) released in January 2003[1], presented the results from biomonitoring for 116 chemicals in the "noninstitutionalized, civilian US population" derived from a subset of the National Health and Nutrition Evaluation Survey (NHANES) [2] 1999-00 population [3]. There were 27 polychlorinated biphenyls (PCB) congeners and 15 polychlorinated dibenzo-p-dioxin/polychlorinated dibenzofuran (PCDD/DF) congeners analyzed for in this data set [3, 4]. While the NHANES study looks at parameters on the US population on the basis of a number of population subgroups such as geographical location, age, sex, and race/ethnicity , to name a few, the NER is more limited (age, gender, race/ethnicity) and provides no information on other potentially important factors such as geographical location. The NER presents "reference ranges," as levels of chemicals in blood and urine found in the general population. These ranges are also sometimes referred to as background levels. A level of a chemical similar to those found in the NER would indicate exposure that was no different from exposure found in the general population, while a level much higher "would indicate unusual exposure that might require further investigation"[4]. It is important to note that a level higher than the NER "reference range" does not indicate a health effect. Nor does it necessarily indicate that a level is elevated relative to background for a specific geographic area such as city or state.

Recently the NER database has been used in a manner for which it was never intended. A person is determined to be “higher” than background based on an analysis of the NHANES database and the cause is then postulated to be due to exposure from a specific site or group of sites. This is usually done without determining the normal background for the vicinity (on a state and/or local area). Additionally, statistics are being produced which are not taking sampling survey design and weightings and are incorrectly represented as being population based statistics.

Because of its predominance, PCB 153 has been used as a marker of exposure for the other PCBs.[1] The most frequently detected PCB congeners in general populations are 138, 153, and 180[1]. We present here a comparison of the statistical analysis approaches and different software packages for PCB-153, and to more limited extent PCB 180 and PCB-180: 153 ratio in the NHANES 1999-00 database.

Methods and Materials

Data from the NHANES 1999-2000 cycle in SAS transport files (.xpt) via the National Center for Health Statistics website [8] were used. Demographic information collected for all individuals in the survey (filename=demo.xpt, n=9965) were merged with PCB measurements taken from a sub-sample of survey participants (filename=lab28poc.xpt, n=1992). Observations were weighted by the sub-sample weight variable (WTSPO2YR) from lab28poc.xpt. Variance estimation for NHANES 1999-2000 data should be based on a nested PSU within stratum sampling design [9]. Statistical analyses were conducted using a standard statistical package, JMP version 5.1 [10], and two programs designed to analyze complex survey data; WesVar version 4.2 [11] and SUDAAN version 9.0 [12]. Different statistical programs estimate variance using different techniques. . Brogan [13] has recommended that “sample survey software be used to analyze sample survey data, especially for estimation of population parameters, descriptive analyses and analytical analyses”. Only WesVar and SUDAAN take into account the complex sampling design (nested PSU within stratum) of NHANES 1999-2000 when estimating population statistics.

Results and Discussion

Table 1 shows results from the SUDAAN analysis of PCB-153 data. The results were compared to those published in the NER[1] and show good agreement. One issue of note is that the published NER does not have a value for percentiles below 75, even though values can be calculated the data set. This is because of the variability in the limits of detection and do not necessarily represent actual numbers. The issue of how to handle nondetected (or censored) data is beyond the scope of this paper.

In Table 2 comparisons are made for stratified NER data. SUDAAN and WESVAR provide values that are quite close to each other, while the JMP values are higher by as much as 25%. Unfortunately one cannot tell ahead of time that such conservative values would be calculated from a standard statistical software package. In general it is probably easier, less time-consuming and more importantly, statistically correct, to use a sample survey software package for analyzing the NER data.

The ratio of PCB 180 to 153 is greater than 1 in Aroclors 1260, 1262 [5]and 1268 [6] and a ratio of greater than 1 was observed in an occupationally exposed cohort and comparison group [7]. In addition certain food groups have been reported to have PCB-180:153 greater than 1. Interestingly, we found that the PCB-180:153 generally increased with age (unpublished data) and more than 100 values have a ratio greater than 1.

The data results presented show potential pitfalls for the unwary or inexperienced data analyst.

Acknowledgement

Partial funding for this work was provided by Latham and Watkins. Helpful discussions with Dr. Sam Caudill and Don Patterson of CDC are acknowledged. This paper has not been reviewed and the opinions expressed do not represent endorsement by either of these parties

References

1. ***Second National Report on Human Exposure to Environmental Chemicals***. 2003, **Department of Health and Human Services, Centers for Disease Control and Prevention**: Atlanta.

2. ***National Health and Nutrition Examination Survey***, National Center for health statistics, Editor. 2005, Centers for Disease Control, Department of Health and Human Services: Atlanta, Georgia.

3. Botman, S., et al., *Design and Estimation for the National Health Interview Survey, 1995-2004.* Vital and Health Stat, 2000. **2**: p. 130.

4. *Second National Report on Human Exposure to Environmental Chemicals-Summary*. 2003.

5. Frame, G., J. Cochran, and S. Bowadt, *Complete PCB Congener Distributions for 17 Aroclor Mixtures Determined by 3 HRGC Systems Optimized for Comprehensive, Quantitative, Congener-Specific Analysis.* J. High Resol. Chromatogr., 1996. **19**: p. 657-668.

6. Schwope, D.M., ***IMPROVEMENTS IN SAMPLING AND ANALYTICAL METHODS FOR POLYCHLORINATED BIPHENYLS IN ENVIRONMENTAL SAMPLES***, in *Forensic Sciences*. 2005, University of Illinois at Chicago: Chicago.

7. Fait, A., et al., *Polychlorinated biphenyl congeners in adipose tissue lipid and serum of past and present exposed transformer repair workers and a comparison group.* Fundamental and Applied Toxicology, 1989. **12**: p. 42-55.

8. CDC, *NHANES 1999-2000 Subsample Notes and Data*. 2003, Centers for Disease Control, Department of Health and Human Services: Atlanta, GA.

9. NCHS, *NHANES Analytic Guidelines*, in *NHANES*. 2004, National Center for Health Statistics, Centers for Disease Control, The Department of Health And Human Services: Atlanta, Georgia.

10. Institute, S., *JMP Statistics and Graphics Guide*. 2002, SAS Institute Inc.: Cary, NC.

11. *WESVAR 4.0 users guide*. 2000, Westat Inc: Rockville, MD.

12. RTI, *SUDAAN Language Manual, Release 9.0.* 2004, Research Triangle Institute: Research Triangle Park, NC.

13. Brogan, D.U., *PITFALLS OF USING STANDARD STATISTICAL SOFTWARE PACKAGES FOR SAMPLE SURVEY DATA*, in *Encyclopedia of Biostatistics*, P. Armitage and T. Colton, Editors. 1998, John Wiley: New York.

Table 1: SUDAAN Analysis (including influence of weights and sampling design) for PCB 153

					Percentiles					
Subpopulation	**N**	**Arith-metic Mean**	**Lower 95%**	**Upper 95%**	**10**	**25**	**50**	**75**	**90**	**95**
NER Reported Values	1926	-	-	-	-	-	-	-	77.8	112
Total, age 12 and older	1926	41.0	*37.5*	*44.4*	*16.6*	*19.5*	*26.7*	*48.5*	77.6	114
Age Group										
12-19 yrs	668	25.4	*23.5*	*27.2*	*16.9*	*19.9*	*23.2*	*27.0*	32.4	35.6

20 yrs +	1258	43.5	*40.0*	*47.1*	*16.6*	*19.4*	*28.4*	*52.0*	82.7	121
Gender										
Males	917	40.6	*37.2*	*44.1*	*17.0*	*19.9*	*26.9*	*47.0*	74.9	110
Females	1009	41.3	*37.1*	*45.5*	*16.1*	*19.2*	*26.5*	*50.1*	78.9	119
Ethnicity										
Non-hispanic whites	725	40.7	*37.4*	*44.0*	*17.0*	*19.7*	*27.6*	*50.1*	76.4	101
Non-hispanic blacks	412	57.9	*48.4*	*67.5*	*17.8*	*21.0*	*30.7*	*59.4*	120	175
Mexican Americans	634	28.9	*27.4*	*30.3*	*15.1*	*18.1*	*22.5*	*28.9*	50.4	67.2

Table 2: Comparison of statistical software analyses for PCB 153, 180 and PCB 180 to 153 ratio

PCB 153 for African American Females, 20+ yrs old							**Percentiles**					
Stats Program	**N**	**Mean**	**Lower 95% CL**	**Upper 95% CL**	**Geometric Mean**	**Std Error Geo Mean**	**10**	**25**	**50**	**75**	**90**	**95**
SUDAAN	130	65.8	48.6	83.0	43.4	3.08	17.3	20.3	39.8	72.4	128	189
SUDAAN (*n.d. excluded*)	85	92.0	65.5	119	67.8	6.81	30.2	41.1	57.6	108	154	234
WesVar	130	65.8	48.5	83.1	43.4	3.13	17.3	20.3	39.8	72.4	128	189
WesVar (*n.d. excluded*)	85	92.0	65.4	119	67.8	6.85	30.2	41.1	57.6	108	154	234
JMP	130	65.8	51.3	80.3	43.4	n/a	17.2	20.5	43.4	100	153	231
JMP (*n.d. excluded*)	85	92.0	71.3	113	67.8	n/a	30.1	45.9	71.8	118	190	393
JMP no weights	130	73.6	58.4	88.8	47.2	n/a	17.2	20.5	43.4	100	153	231
JMP no weights, (*n.d. exclude*	85	102	80.8	123	74.9	n/a	30.1	45.9	71.8	118	190	393
PCB 153 for African Americans, 20+ yrs old												
SUDAAN	228	64.6	52.3	76.9	43.0	2.9	17.8	21.8	36.7	72.8	127	212
WesVar	228	64.6	52.2	77.0	43.0	2.9	17.8	21.8	36.7	72.8	127	212
JMP	228	64.6	54.3	74.9	43.0	n/a	17.6	22.3	44.1	99.3	154	274
JMP no weights	228	74.8	63.8	85.9	48.7	n/a	17.6	22.3	44.1	99.3	154	274
PCB 180 for African Americans, 20+ yrs old												
SUDAAN	228	39.3	31.9	46.7	25.3	1.66	9.18	11.3	23.7	44.6	86.4	120
WesVar	228	39.3	31.8	46.7	25.3	1.68	9.18	11.3	23.7	44.6	86	120
JMP	228	39.3	33.3	45.2	25.3	n/a	8.89	11.6	29.7	63.0	104	143
JMP no weights	228	45.6	39.2	51.9	28.9	n/a	8.89	11.6	29.7	63.0	104	143
Ratio: PCB180:PCB153 for African Americans, 20+ yrs old												
SUDAAN	227	0.634	0.586	0.682	0.592	0.009	0.451	0.507	0.510	0.684	0.855	0.991
WesVar	227	0.634	0.586	0.683	0.592	0.009	0.451	0.507	0.510	0.684	0.855	0.991
JMP	227	0.634	0.584	0.684	0.592	n/a	0.463	0.507	0.520	0.728	0.865	1.044