Corporate Health Physics
Corporate Occupational Medicine
Corporate Product Responsibility
Corporate Toxicology
3M Medical Department

3M Center, 220-2E-02
PO Box 33220
St. Paul, MN 55133-3220
651 733 1110

3M

Working Memorandum on Data Quality Assessment (of the Mean and Range of PFOS (ppb) Levels in Current and Historical Human Populations)

This report documents Battelle's evaluation of the statistical quality and defensibility of 12 data sources of serum perfluorooctanesulfonate (PFOS) levels found in non-occupational human populations. These data were summarized in the 3M white paper, "Perfluorooctane Sulfonate: Current Summary of Human Sera, Health and Toxicology Data" provided to the EPA in January, 1999. The purpose of this initial collection of human non-occupational serum samples by the 3M Medical Department was to determine the presence or absence of PFOS in human blood from current and historical samples. 3M requested a review of these data by Battelle to determine what additional inferences could be obtained by analyses of these data. Battelle statisticians felt that only the pooled serum from the 18 U.S. blood banks was judged defensible for a statistical analysis. Of these data, the geometric mean PFOS blood level was 28.2 ppb (95% confidence interval 23.9-33.3). Regional differences could not be explained without additional data and analyses.

WORKING MEMORANDUM
on
DATA QUALITY ASSESSMENT
by

Bruce Buxton, Warren Strauss, Owen Chang
BATTELLE
September 22, 1998

This report documents Battelle's evaluation to date of the statistical quality and defensibility of 12 data sources presented by 3M staff who suggested that these data sources might be used to estimate blood levels of perfluorooctanesulfonate (PFOS) in non-occupational human populations. According to 3M, these data were collected to help address three discovery questions posed to the 3M Medical Department: (1) Is PFOS detectable in individual human blood from a corporate-based 3M employee population known not to have worked in 3M fluorochemical manufacturing plants (but knowledgeable about fluorochemicals)? (2) Is PFOS detectable in general population pooled blood samples from different geographical locations across the United States? and (3) Is PFOS detectable in human blood which would have been collected prior to the introduction of PFOS-based chemicals into the marketplace of the general population? Subsequently, Battelle was asked whether any further statistically defensible inferences could be derived from these data; this report addresses that question. In this report we (1) list important characteristics which data sets must have in order to support statistically defensible inferences, (2) judge each of 3M's 12 data sources against these criteria, and (3) statistically analyze the available data, where appropriate.

Our findings from this initial assessment of statistical data quality can be summarized in the following points:

- We suggest 12 criteria which define statistically defensible data sources. Of the 12 data sources presented by 3M staff, no source was judged sufficient on all 12 criteria, and only one source, the 1998 pooled serum from U.S. blood banks (data source A3), was judged sufficient for subsequent statistical analysis.
- The 1998 pooled serum blood bank data indicated a national geometric mean PFOS blood level of 28.2 ppb with a 95% confidence interval of 23.9 to 33.3 ppb. Significant regional differences are apparent but can not be explained without additional data and analysis.

Data Characteristics for Statistical Defensibility

In order to make statistically defensible conclusions from a set of data, several requirements must be met regarding design of the study and implementation of the data collection effort. Each of these requirements can be associated with important quantitative assumptions underlying the numerical models which are used to make statistical inferences. In our assessment here, we considered three different types of estimation problems regarding PFOS blood levels which one might want to consider: (1) temporal changes in PFOS blood levels over the past several decades, (2) spatial trends in PFOS blood levels for various regions across the U.S., and (3) demographic trends in PFOS blood levels among various subpopulations in the U.S., or worldwide. Table 1 summarizes 12 statistical requirements for data defensibility, and associates each with the three types of estimation problems listed above. Note in this table that the first five

Table 1. Statistical Requirements for Defensible Quantitative Conclusions

Statistical Requirements	Types of Estimation Problems		
	Temporal Changes	Spatial Trends	Demographic Trends
1. Detailed documentation of the study objectives and approach, including sampling design (number and types of samples), and sampling and chemical analysis protocols	✔	✔	✔
2. Specification of a well defined target population about which inference is to be made (e.g., the entire U.S. population, or some specified subpopulation)	✔	✔	✔
3. Unbiased data collection protocols which ensure that all samples are demographically representative of the target population; generally this implies random sampling from the target population	✔	✔	✔
4. Along with PFOS blood levels, accurate collection of important predictor or explanatory information, such as demographics like age, gender, race, etc.	✔	✔	✔
5. Compositing of samples, when done, is performed only for samples within the same subpopulation so that important potential differences among subpopulations are preserved in the resulting data	✔	✔	✔
6. Sampling is performed at several (i.e., at least three) points in time which are spaced at regular time intervals (e.g., every five years)	✔		
7. At each point in time, sampling is performed cross-sectionally from the entire target population or subpopulations	✔		
8. Sampling is performed to provide uniform coverage of the entire spatial domain of interest (e.g., the entire U.S.)		✔	
9. Data is collected both for multiple locations (e.g., regions or cities) across the spatial domain and for multiple individuals within each location		✔	
10. All data are collected at approximately the same point in time to eliminate the potential for confounding due to temporal trends		✔	
11. Specification of target subpopulations which are well defined in terms of important demographic parameters (e.g., age, gender, race, etc.) and from which sampling is randomly performed			✔
12. Either fix sampling at one time and spatial location (e.g., city) or ensure that sampling for all demographic subpopulations is performed in a balanced design across the same times and locations (i.e., to eliminate potential confounding due to time or location)			✔

requirements are more general in nature and apply equally to all three types of estimation problems, while the last seven requirements apply more specifically to a different type of estimation problem.

Assessment of 12 Data Sources

On May 12, 1998, 3M staff faxed to Battelle a table listing 12 data sources they believe might be used to estimate PFOS blood levels in non-occupational human populations. This table has been retyped as shown in Figure 1. Subsequently, 3M staff sent to Battelle electronic files with the PFOS blood data they had available, as well as associated written documentation on the data sources which they had. Based on this information, Battelle attempted, and was partially successful, to obtain through literature searches more extensive documentation on these 12 studies. However, in many cases no additional written documentation could be found.

Given the information sent to Battelle by 3M and other documentation Battelle obtained from the literature search, Table 2 summarizes our assessment of how well each of the 12 data sources meets each of the 12 defensibility requirements listed in Table 1. In Table 2, a "✓" indicates that the given data source seems to meet the associated defensibility criterion, while a "X" indicates that the data source does not meet the criterion. This assessment by Battelle is based solely on the written documentation we were able to obtain, either from 3M or the literature review. In a large majority of cases, no written documentation was found and so we can not judge whether the defensibility criteria are met by a data source, or not. In these cases, a "?" is listed in Table 2. In some cases, noted as "✓?", documentation may have been found about the initial study which collected the blood samples, but no documentation was available describing how particular subsamples were selected to be sent to 3M. Also, in some cases we judged that a given data source would be appropriate for addressing just one type of estimation problem (i.e., temporal changes, spatial trends, or demographic trends), and so defensibility criteria for other problems would not be applicable. These cases are denoted by a "–" in Table 2. Table 3 lists more detailed notes about each of the data sources for those criteria where sufficient information was found in the written documentation to warrant comment.

Figure 1
Summary of Mean and Range of PFOS (ppb) Levels in Current and Historical Human Populations.
All Data Analyzed in February – April 1998

Populations	Description of Sample	Mean* (ppb)	Range* (ppb)
A. Current Populations (blood collected in 1998)			
1. Non-occupationally exposed 3M employees	31 individuals	47	28-96
2. Commercial pooled serum samples			
Intergen Laboratory	3 samples each with ≥100 donors	44	43-44
Sigman Laboratory	3 samples each with ≥100 donors	33	26-45
3. Pooled serum from 21 separate U.S. blood banks (see Figure 1 for more detail)	3 to 6 pooled samples per location with 5 to 10 donors per sample	29	9-56
B. Historical Samples (chronological order)			
1. Korean War era U.S. military recruits, 1948 to 1951	10 pooled samples with 10 donors per sample	N.D.**	N.D.
2. Swedish samples, 1957	10 individual samples	2	N.D. – 2
3. Michigan Breast Cancer Study, 1969-1971	5 individual samples	33	N.D. – 59
4. Swedish samples, 1971	10 individual samples	1	N.D. – 1
5. MRFIT pooled calibration samples, 1976	6 pooled samples with unknown number of donors per sample	31	14-56
6. MRFIT pooled calibration samples, 1980	3 pooled samples with unknown number of donors per sample	23	14-41
7. China samples (Linxian, rural province), 1984	6 individual samples	N.D.	N.D.
8. MRFIT individual samples, 1985	3 individual samples	31	N.D. – 44
9. China samples (Shandong, rural province), 1994	6 individual samples	N.D.	N.D.

* Rounded to nearest ppb

** Not detected

Note: This figure was retyped from a fax sent to Battelle on 5/12/98 without changing possible mistakes.

Table 2. Overall Assessment of Defensibility of 12 Data Sources

Statistical Requirement	Data Source											
	A1	A2	A3	B1	B2	B3	B4	B5	B6	B7	B8	B9
1. Detailed documentation	✓?	✓?	✓?	✓?	✓?	✓?	✓?	✓?	✓?	✓?	✓?	✓?
2. Defined target population	✓	?	?	✓	?	?	?	✓	✓	✓	✓	?
3. Unbiased sampling protocols	?	?	X	?	?	?	?	X	X	✓	X	?
4. Accurate explanatory information	✓	?	?	?	?	?	?	✓?	✓?	✓?	✓?	?
5. Appropriate compositing	–	?	?	?	–	?	–	?	?	–	–	–
6. Regular sampling in time	–	?	–	?	?	?	?	?	?	–	?	?
7. Complete sampling at each time	–	?	–	?	?	?	?	?	?	–	?	?
8. Uniform sampling across space	–	?	✓	?	?	–	?	–	–	–	–	?
9. Multiple samples locally	–	?	✓	?	?	–	?	–	–	–	–	?
10. Sampling at same time(s)	–	?	?	?	?	–	?	–	–	–	–	?
11. Defined demographic subpopulations	?	?	–	?	?	?	?	?	?	?	?	?
12. Sampling at same time(s) and location(s)	✓	?	–	?	?	?	?	?	?	?	?	?

✓ = Adequately designed and implemented
X = Inadequately designed or implemented
? = Unknown due to insufficient documentation
– = Not applicable to this data source

Table 3. Notes Regarding Defensibility of 12 Data Sources

A1. Non-occupationally exposed 3M employees
1. Detailed documentation – The only documentation is in Geary Olsen's 8/15/98 letter 2. Defined target population -- The target population was 3M employees with minimal occupational exposure 4. Accurate explanatory information -- Some demographic information was collected and included, such as age and gender 12. Sampling at same time(s) and locations(s) -- The data appears to be only from employees at the corporate center
A2. Commercial pooled serum samples
1. Detailed documentation – The only documentation is in Geary Olsen's 8/15/98 letter
A3. Pooled serum from 21 separate U.S. blood banks
1. Detailed documentation – The only documentation is in Geary Olsen's 8/15/98 letter 3. Unbiased sampling protocols – There is potential bias as a result of the 50% refusal rate 8. Uniform sampling across space -- The samples are geographically dispersed 9. Multiple samples locally—There are multiple regions and multiple individuals within each region
B1. Korean War era U.S. military recruits, 1948 to 1951
1. Detailed documentation – The only documentation is in Geary Olsen's 8/15/98 letter 2. Defined target population -- The target population was U.S. military recruits from the Korean War era
B2. Swedish samples, 1957
1. Detailed documentation – The only documentation is in Geary Olsen's 8/15/98 letter
B3. Michigan Breast Cancer Study, 1969-1971
1. Detailed documentation – The only documentation is in Geary Olsen's 8/15/98 letter
B4. Swedish samples, 1971
1. Detailed documentation – The only documentation is in Geary Olsen's 8/15/98 letter

Table 3. Notes Regarding Defensibility of 12 Data Sources (Continued)

B5. MRFIT pooled calibration samples, 1976
1. Detailed documentation – MRFIT is documented in Geary Olsen's 8/15/98 letter and in multiple journals including JAMA, Vol. 235, No. 6, Feb. 23, 1976, but subsampling for 3M is not documented 2. Defined target population -- The target population was men aged 35-57 with multiple risks for heart disease 3. Unbiased sampling protocols – The sampling protocol selected certain characteristics, with multiple screenings 4. Accurate explanatory information – Information about the original sample population was recorded, but no information about the pooled calibration sample is known
B6. MRFIT pooled calibration samples, 1980
1. Detailed documentation – MRFIT is documented in Geary Olsen's 8/15/98 letter and in multiple journals including JAMA, Vol. 235, No. 6, Feb. 23, 1976, but subsampling for 3M is not documented 2. Defined target population -- The target population was men aged 35-57 with multiple risks for heart disease 3. Unbiased sampling protocols – The sampling protocol selected certain characteristics, with multiple screenings 4. Accurate explanatory information – Information about the original sample population was recorded, but no information about the pooled calibration sample is known
B7. China samples (Linxian, rural province), 1984
1. Detailed documentation – Documented in Geary Olsen's 8/15/98 letter and in Journal of the National Cancer Inst., Vol. 85, No. 18, Sept 15, 1993, but subsampling for 3M is not documented 2. Defined target population -- The target population was all the residents of Linxian aged 40-69 without debilitating diseases or prior esophageal or stomach cancer 3. Unbiased sampling protocols – The sample appears to be representative of the target population 4. Accurate explanatory information -- Information about the original sample population was recorded, but no information about the individual sample is known
B8. MRFIT individual samples, 1985
1. Detailed documentation -- MRFIT is documented in Geary Olsen's 8/15/98 letter and in multiple journals including JAMA, Vol. 235, No. 6, Feb. 23, 1976, but subsampling for 3M is not documented 2. Defined target population -- The target population was men aged 35-57 with multiple risks for heart disease 3. Unbiased sampling protocols – The sampling protocol selected certain characteristics, with multiple screenings 4. Accurate explanatory information – Information about the original sample population was recorded, but no information about the pooled calibration sample is known
B9. China samples (Shandong, rural province), 1994
1. Detailed documentation -- The only documentation is in Geary Olsen's 8/15/98 letter

The main conclusion from our overall assessment of the 12 data sources is that in most cases there is simply not enough written documentation of the study design and sampling protocols to judge whether the resulting data are statistically defensible, or not. In some cases (e.g., MRFIT) documentation is available on the original study under which an entire set of blood samples was collected, but no documentation is then available on how the subsampling was performed to send samples to 3M for chemical analysis. If either the original sampling or the subsampling was biased, then the data obtained by 3M will also be biased (i.e., systematically too high or too low). Given this lack of documentation, there is currently little, if any, justification for statistically analyzing 3M's data and presenting statistical conclusions. Pooling the data across studies (e.g., to look for blood temporal, spatial, or demographic trends) is not recommended because of the widely different study objectives and target populations among the studies.

Of all the data sources provided by 3M, probably the most defensible for statistical analysis purposes is the data set of 1998 pooled serum from U.S. blood banks (data source A3). Although

detailed written documentation is lacking, it does appear that the samples may have been selected in an unbiased manner, at least within a particular blood bank. We are concerned with possible non-response bias potentially introduced by the large number of blood banks (approximately half of those contacted) which declined to send 3M the samples they requested. There are also potential biases resulting from the probable lack of control by each blood bank concerning the demographics of the donors of the samples which were sent to 3M. However, recognizing these potential serious data deficiencies which limit our ability to make valid inferences, in the next section we have subjected the resulting PFOS blood level data to a limited statistical analysis.

Statistical Analysis of Blood Bank Data

The recently collected and analyzed pooled blood samples from an independent Minnesota and 17 geographically dispersed Red Cross blood banks have been combined to estimate a nationally representative mean PFOS (ppb) level. However, a variety of assumptions are necessary, including:

1) Each pooled sample is representative of the regional population (i.e., each member of the population had an equal chance of being represented in the sample).

2) Each pooled sample has a known number of donors, and each donor is represented equally in the sample.

3) The sampling and analytical methods for obtaining serum PFOS levels is the same across all samples.

It is unlikely that the samples obtained by 3M completely satisfy all of the above assumptions. Also, the data have the following additional limitations:

1) The blood banks represent only those that donated blood – i.e., cities where the blood bank(s) refused to participate had no chance of being represented in 3M's sample. Furthermore, it is unlikely that the samples obtained from participant blood banks are representative of the general population in each geographic region. Rather, these samples are representative of people who donate blood to the Red Cross.

2) It is unknown how many donors are in each pooled sample, nor is it known how each donor was selected.

3) Most of these samples came from Red Cross blood banks, so it is likely those sampling methods are the same. The Minnesota samples probably do not have significantly different methods. As all of the analysis was performed by 3M, the analytical methods should be the same.

Although these data may not satisfy all of the assumptions detailed above, the following random effects model was used to estimate a national geometric mean PFOS level:

$$\ln(PFOS_{ij}) = \mu + B_i + \varepsilon_{ij}, \text{ where}$$

$\ln(PFOS_{ij})$ represents the natural-log transformed PFOS level associated with the jth pooled sample from the ith Blood Bank, $\exp(\mu)$ is the national geometric mean PFOS level, B_i is the difference (on the natural log scale) between the mean PFOS level at the ith blood-bank and the national mean, and ε_{ij} is the residual error left unexplained by the model (the difference between the individual sample PFOS measurements and the mean PFOS level at its corresponding blood bank). It is assumed that B_i follows a normal distribution with mean zero and variance $\sigma^2_{\text{Between Blood Banks}}$, and that ε_{ij} follows a normal distribution with mean zero and variance $\sigma^2_{\text{Within Blood Banks}}$.

Table 4 provides the estimates obtained from fitting this model to the data. Table 5 lists the individual region geometric means which were found to fall outside the 95% and 99% confidence intervals for the overall geometric mean.

Table 4. Parameter Estimates for Blood Bank Data

Parameter	Estimate	95% Confidence Interval	99% Confidence Interval
National Geometric Mean PFOS Level (e^{μ}) (ppb)	28.2	(23.9, 33.3)	(22.5, 35.4)
σ^2 Between Blood Banks $[\ln(ppb)]^2$	.095	-	-
σ^2 Within Blood Banks $[\ln(ppb)]$	.052	-	-

Table 5. Regions With Geometric Means Falling Outside Confidence Intervals for Overall Geometric Mean

Region	n	Geometric Mean (ppb)	Outside 95% CI	Outside 99% CI
Santa Barbara, CA	3	15.9	✓	✓
Omaha, NE	5	16.2	✓	✓
Billings, MT	3	22.4	✓	✓
Grand Rapids, MI	3	23.1	✓	
Cheyenne, WY	3	23.3	✓	
East Orange, NJ	3	23.5	✓	
Newark, DE	5	23.9	✓	
Anchorage, AK	3	25.3		
Las Vegas, NV	3	26.6		
Santa Rosa, CA	3	27.3		
Corpus Christi, TX	3	28.9		
Kansas City, MO	3	29.6		
Davenport, IA	3	31.9		
Scottsdale, AZ	3	36.5	✓	✓
Meridian, MS	3	37.5	✓	✓
Minneapolis, MN	6	45.6	✓	✓
Lafayette, LA	3	45.7	✓	✓
Greenville, SC	3	51.5	✓	✓

The above results demonstrate that 12 of 18 blood banks had geometric mean PFOS levels that fell outside the 95% confidence interval constructed for the estimated national geometric mean PFOS concentration. This high number of regional means falling significantly above and below the overall mean is a result of the fact that there is significant variation between the geometric

mean PFOS concentrations observed at each blood bank. It is possible that there are meaningful factors that can explain a portion of this variability. For example, the blood banks in Minneapolis, MN, Lafayette, LA and Greenville, SC may be located in close proximity to industries that apply 3M products (such as carpet manufacturers), while Santa Barbara, CA, Omaha, NE and Billings, MT may be isolated from such exposure sources. However, differences in the donor population at each blood bank may be another possible explanation for the observed variability between blood banks. To investigate the reasons for these regional differences, additional data analysis with more explanatory information would be required.

Resampling the 18 Blood Banks

3M asked Battelle their opinion of the statistical merit of resampling these 18 blood banks to help evaluate the uncertainty in PFOS blood levels within geographical locations. In the current data, there are variations of 20 ppb or more between the minimum and maximum PFOS levels measured in different pooled samples from the same geographical location.

If additional information can be collected from the blood banks, Battelle recommends that the most beneficial data would be demographic information on the pooled samples which were previously sent to 3M. We suggest that it is most important at this point in time to better understand the factors influencing the PFOS blood level variability. Collecting new blood samples without demographic information would help quantify the variability one should expect from the 18 local populations, but it would not help explain the source(s) of that variability. For designing efficient future sampling programs, we need to know how much variability to expect, but more importantly the major environmental, behavioral, nutritional, demographic, socioeconomic, etc. factors that affect that variability so that we can try to control these factors (e.g., stratify by these factors) to obtain more accurate sampling results. A second concern with resampling these 18 locations is that they may constitute a biased sampling of the general U.S. population. By taking more samples from these cities, we might be simply generating more precise but highly biased estimates.