DUPONT
ESTABLISHED 1802
E. I. DU PONT DE NEMOURS & COMPANY
INCORPORATED
WILMINGTON, DELAWARE 19898

ENGINEERING DEPARTMENT
LOUVIERS BUILDING

April 14, 1986

Dr. Morton Lippman
Chairman, Clean Air Scientific
Advisory Committee
Department of Environmental Medicine
New York University Medical Center
Tuxedo, NY 10987

Dear Dr. Lippman:

Attached for your information is Du Pont's response to an April 4, 1986 inquiry (attached) by Dr. Lester Grant, EPA - ECAO, requesting explanations on reasons for discrepencies in the magnitude of the blood pressure - log blood lead regression coefficient estimated from NHANES II data by Du Pont and Dr. Schwartz. This inquiry was a result of Dr. Schwartz' March 3, 1986 memo (attached) to Dr. Grant.

I thought you and Drs. Ware and Royall would be interested in this correspondence due to the extensive discussions on NHANES II blood pressure results that have taken place during the 1985 and 1986 CASAC meetings on the Air Quality Criteria for Lead.

If you have any questions or concerns, please feel free to contact me collect at (302) 366-3540.

Very truly yours,

ENGINEERING SERVICE DIVISION
Applied Statistics Group

Charles G. Pfeifer

Charles G. Pfeifer
Consultant Supervisor

CGP:jcl
Atchs
5.20

cc: Dr. Richard M. Royall
Johns Hopkins University

Dr. James H. Ware
Harvard School of Public Health

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
Environmental Criteria and Assessment Office (MD-52)
Research Triangle Park, North Carolina 27711

April 4, 1986

Dr. Charles G. Pfeifer
Consultant Supervisor
Applied Statistics Group
Engineering Services Division
E. I. Du Pont de Nemours & Co., Inc.
Wilmington, Delaware 19898

Dear Dr. Pfeifer:

Thank you for your letter of March 25 to Dr. David Weil of my staff, in which you responded to his telephone request for additional tables illustrating the predicted drop in blood pressure associated with a given drop in blood lead, assuming a causal relationship. These tables make very clear the magnitude of the impact small changes in blood lead can have on this important physiological parameter.

I remain concerned about the discrepancy between the results of the regression analyses presented by Du Pont in its public comments on the Addendum to the Air Quality Criteria for Lead and the work of Schwartz et al. As you are well aware, there are noticeable differences between the regression coefficients found by your group and by Schwartz et al., especially after adjusting for site. These differences are highlighted in the attached Table 1, which has been constructed merely for comparative purposes. Note that the coefficient for systolic blood pressure vs. blood lead (5.91) for white males aged 40-59 has been inserted based on further analyses adjusting for site done by Schwartz at my request.

Because we would like to understand the reasons for and, if possible, resolve these differences, I would like to draw your attention to the attached commentary provided to me by Joel Schwartz, in which he discusses some of the differences in the way your analyses were performed vs. the way in which his were done. These differences may be summarized as follows:

1. A number of significant variables other than those listed in Table 2 of the Pirkle et al. (1985) paper were considered in the analyses of white males aged 40-59 done by Schwartz et al. (including family history of hypertension, taking hypertension medication, SES, height, and coffee), whereas Du Pont did not use them.

2. Schwartz and colleagues, in their published paper, used the log transform of dietary variables, whereas Du Pont did not.

3. Schwartz et al. used weights in their analyses taking site into account, so that the results could be more directly compared to "non-site" analyses using SURREGR, whereas Du Pont did not.

4. Schwartz argues that one must pick a reference PSU as an intercept when adjusting for site, rather than having no intercept, as in the Du Pont analyses.

5. It is difficult to understand the Du Pont variable Age-yy, as it yields a root mean square age of 5.8.

6. Analyses performed by Schwartz et al., subsequent to the Pirkle et al. (1985) paper, have used the age group of all men ≧ 20 and variables not included in the Harlan et al. analyses, whereas Du Pont used all men aged 12-74 for its analyses and did not include all variables found to be significant by Schwartz. This further complicates comparison of results.

As you can see, there are a number of questions concerning the different results obtained by Du Pont and Schwartz which we would like to resolve. It would be of considerable help to this office if you and your colleagues at Du Pont would consider re-analyzing the subject data by rerunning regression analyses which include the same variables, age groups, and weights used in the Schwartz analyses. Alternatively, I would very much like to see an explanation of why Du Pont (1) did not include the additional variables such as SES, height, etc., which are not found in Table 2 of the Pirkle et al. paper, but which were included in the Schwartz analyses; (2) did not run weighted regression analyses where PSU's were included; (3) did not use the log transform of the dietary variables or an intercept in adjustment for site; and so on in order to better match the actual analyses done by Schwartz and colleagues.

I would be happy to discuss this matter by telephone with you next week. I would also be grateful for your prompt attention to this matter, as I'm sure you can appreciate the time constraints under which we are working.

Thank you for your cooperation and help in this matter.

Sincerely yours,

Lester D. Grant

Lester D. Grant, Ph.D.
Director, Environmental Criteria
and Assessment Office (MD-52)

Enclosure

TABLE 1

COEFFICIENTS FOR THE NATURAL LOG OF BLOOD LEAD CONCENTRATION (lnPbB) VS. BLOOD PRESSURE (BP) IN MEN WITH AND WITHOUT ADJUSTMENT FOR SITE VARIABLES.

ANALYSIS PERFORMED BY	STUDY GROUP	COEFFICIENT OF lnPbB vs. BP: UNADJUSTED FOR SITE	COEFFICIENT OF lnPbB vs. BP: ADJUSTED FOR SITE
Schwartz (U.S. EPA)	NHANES II Males aged 20-74		
	Systolic (n=2254)	5.23 ***	3.23 **
	Diastolic (n=2248)	2.96 ***	1.39 *
Pfiefer et al. (Du Pont)	NHANES II Males aged 12-74		
	Systolic (n=2794)	3.43 ***	0.36
	Diastolic (n=2789)	2.02 ***	1.95 *
Schwartz (U.S. EPA)	NHANES II White males aged 40-59		
	Systolic (n=543)	8.44 **	5.91 **
	Diastolic (n=565)	3.95 **	3.12 *
Pfiefer et al. (Du Pont)	NHANES II White males aged 40-59		
	Systolic (n=553)	6.27 **	3.46
	Diastolic (n=575)	4.01 **	1.93
Pocock et al. (Univ. of London)	British Regional Heart Study White males aged 40-59		
	Systolic (n=7371)	1.68 **	2.09 **
	Diastolic (n=7371)	0.30	1.81 ***

* $p < 0.05$
** $p < 0.01$
*** $p < 0.001$

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

MAR 3 1986

SUBJECT: NHANES II BLOOD PRESSURE ANALYSIS

FROM: Joel Schwartz
OPA

Joel

To: Les Grant
ECAO

As I indicated last week, this memo encloses and discusses the additional results that I performed for you. It also amplifies on the question of what analysis Dupont actually performed. Since we spoke I went back and looked at the Dupont submission to CASAC last month. I had always assumed that they had duplicated our analysis, except for using variables for PSU, and some minor differences (we excluded fingerprick blood samples, they may not have, we defined Body Mass Index using the MKS system of measurements, they obviously did not, etc).

On closer examination it is not surprising that they got different results than us since they did a very different analysis. First, as I mentioned over the phone, they did not include a number of variables that we did. These variables were discussed in the A.J. Epid paper, and in the Lead RIA which were both published in Feb. 1985, and were in the analyses that I presented to CASAC by memo in April 1985, and in my oral presentation. They include family history of hypertension, (which can hardly be considered a marginal variable), use of hypertensive medication, and recreational exercise, which has been the subject of considerable discussion with respect to cardiovascular disease and blood pressure. In addition, when Jim Pirkle and I made our presentations to the peer review panel you convened in RTP, one of the reviewers suggested that we consider height as a variable in addition to body mass index. I did, and it was significant, as I mentioned in the second meeting of the reviewers the next week. This variable was therefore included in the regression tables in the Lead RIA which were published in Feb. 1985, 3 months before Duponts submission. They were also discussed and are included in my paper given at the Heavy Metals Conference in Athens, and published 6 months before Dupont made its presentation to CASAC this spring. In addition, another lead industry analysis, submitted in the fall of 1984 as a comment to the lead phasedown docket, claimed that we erred in our analysis by not including coffee. To answer this criticism we included it in the analysis I submitted to CASAC, since it was significant. Let me mention that these changes were not made by randomly testing variables for significance. Rather, the entire stepwise formalism discussed in our paper and the Lead RIA was repeated each time, with the new candidate variable included in the list of variables available for selection. Maximum R-square

regression again selected the model that maximized R-square subject to all the variables being significant, and this model was used. The only difference in the procedure was the inclusion of additional variables in the candidate list. This is discussed in detail in the Lead RIA with respect to the new variables.

The exclusion of significant variables increases the error sum of squares, which can decrease the significance of other variables. In addition, I see no reason to exclude variables that are selected by objective stepwise criteria such as explanatory power, particularly since biological theory suggests a priori that they should be included.

The second major difference between our analysis and the Dupont analysis was that they did not use the log transform for the dietary variables, even though our published analysis did, and even though the transformed variables are more significant than the untransformed ones. Indeed, in the base models that Dupont uses, the untransformed dietary variables are not significant. On p 42 of their "reanalysis of Pirkle et al" they report a diastolic regression, before including PSU variables with dietary vitamin C, serum vitamin C, Riboflavin, and Oleic fatty acid intake all insignificant. This is because they have not used the transformed form, although there may be other differences in how they coded the variable. In any case, our coefficients and t-statistics for these variables were confirmed by independent analysis by epidemiologists at the University of North Carolina. Therefore, we do not need to determine how Dupont misdefined the variables, since we have confirmation of both our definitions and the results of those definitions.

There is also something very bizarre about their variable Age_yy, which presumably is age in years. Their descriptive statistics table for variables(the first page in their appendix on the Pirkle et al paper) reports a sum of squares of 18764 for 552 observations, or a mean square (reported as zero in their table presumably because of some coding error) of 33.9, which gives a root mean square of 5.8. Obviously, the mean age was not 5.8 for the white males aged 40-59. While it plausible that they would divide the variable by 10, the mean age could not have been 58 either. Our results give a mean age, and a root mean square age of about 49.5. It appears that they have subtracted the means from all of their independent variables (and 5.8 is the root mean square deviation for age), but not their dependent variable. What exactly they did is not clear, but their coefficients for age, for the model without psu's in it are different from ours, and different from the UNC results.

However, one factor that tells alot about the cumulative impacts of these miscodings and omitted variables is that they report an R-square of 0.17 for systolic (p11 of appendix) and 0.148 for diastolic (p 42 appendix) for their model without psu terms. Yet our R-square for our models with all the terms except psu in were 0.242 systolic and 0.249 diastolic, and were 0.22 and 0.23 even for the models just including the correctly coded age,

BMI, lead, and dietary terms. Since UNC has already replicated our results these large differences appear to be the result of some substantial errors on the part of Dupont. All other things being equal a reduction of 10% in R-square would be expected to reduce the t-statistics of all the independent variables by about 10% as well, which will make a substantial difference in the significance level of some variables. Duponts failure to replicate both our age coefficients and UNC's may also have a specific impact on their lead coefficient since age and lead are collinear

The same pattern of miscoded dietary variables, and omitting important covariates applies to Duponts analysis of the full age range. Given all of this it is not surprising that their results are different, especially since their sum of square errors is much larger.

As I mentioned on the phone, one other difference is the use of the sample weights. The NHANES II sample deliberately oversampled the poor, the elderly, etc. To obtain regression results that can be generalized to the population, such as the expected change in mean blood pressure from a change in mean blood lead, which is a calculation that dupont themselves performed, one needs to use the weights, which return the results to those that would have obtained in a random sample. In addition, since SURREGR must use weights, and since all of our published analyses use SURREGR, to determine the change in the coefficient that occurs when you include PSU dummies, one needs to use the weights, so that the two coefficients will be comparable. Otherwise, you don't know what change came from the PSU's and what from the lack of weighting.