Du Pont Response to Points of Discrepancy
Between Analyses of Dr. Schwartz and Du Pont
on Relating Blood Pressure to Blood Lead
Using NHANES II Data

April 11, 1986
Charles G. Pfeifer
E. I. du Pont de Nemours & Co., Inc.

As you requested in your letter of April 4, 1986, I am providing explanations to the six points you raised concerning possible reasons for differences in the estimated BP-ln(BPb) coefficient between analyses performed by Du Pont and Dr. Schwartz. My reply is organized in the order of the points you raised.

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.**

   o Du Pont's analysis first reported in our May 9, 1985 comments to EPA and CASAC on Corrigenda and later reported in our March 7, 1986 comments to EPA and CASAC on Addendum focussed on a methodological flow in the papers of Harlan et al (1985) and Pirkle et al (1985). The identified flaw was the nonadjustment for site effects of the BP-ln(BPb) regression coefficient. Our analysis was intended to address the effect on both the magnitude and statistical significance of the coefficient when attempting to replicate the published age-race-sex groups and regression models. Recognizing that some differences in our analysis and those of Harlan et al and Pirkle et al would inevitably remain, we performed our analyses without site adjustment as well as with site adjustment to enable a fair comparison on the effect of site adjustment on the BP-ln(BPb) coefficient. The other terms noted in the heading were not incorporated not because we believe they are unimportant but rather because they were not considered in the two published papers. Attention was restricted to the published papers because of EPA rules for citing in Criteria Document.

   o The analyses of Dr. Schwartz which included additional predictors such as family history of hypertension, coffee, etc, were a part of his memos of June 26, 1985 (Response to Du Pont Critique of Blood Pressure) and of October 2, 1985 (Resonse to Richard Royall's questions on the blood lead-blood pressure relationship in NHANES II). These memos were in response to Du Pont's critique. Dr. Schwartz has complicated the fair comparison between Du Pont's results, the results in Harlan et al/Pirkle et al and his results. Remember it is the credibility of the published results that Du Pont has questioned as a result of the identified methodological flaw. We support continued improvements in analysis to best assess the magnitude and significance of the blood pressure-blood lead relationship (see Objective 2 in cover letter). However, the first step was to get consensus agreement on the invalidity of the results in the published papers.

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

   o This observation is correct and is due to an unintentional oversight on my part. The footnote in Table 2 of Pirkle et al was missed. Dietary variables in the Harlan et al models apparently were not log transformed.

   o One would hardly expect this difference to account in any substantive way for the differences in the magnitude of the BP-ln(BPb) coefficients between the Du Pont and Schwartz analyses. If it did then very serious questions regarding both the data and appropriateness of the models used by both Du Pont and Schwartz would be raised. At best, use of nonlog transformed variables might result in some minor loss of explanatory power. Contrary to Dr. Schwartz' comment in his March 3 memo to you, all but Oleic acid maintained significance ($p \leq .05$) in our Systolic BP analysis when not transforming whereas none did for the Diastolic BP analysis.

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.**

   o While the details of the Du Pont analyses reported in our May 9, 1985 comments and March 7, 1986 comments are unweighted regressions, we noted (Section III.C.3, item 1 of May 9, 1985 comments) that we ran corresponding weighted regressions which provided very comparable results. Summaries from these analyses are provided in Tables 1 and 2 attached, along with the unweighted results previously reported. These results demonstrate the effect of weighting has little effect on either the magnitude or significance of the results. In fact, the one adjusted coefficient that was significant in the unweighted analysis (Males, 12-74) was not significant in the weighted analysis.

   o The complex survey design was not used in our analyses. The coefficient estimates would be the <u>same</u> as those obtained by the weighted analysis, shown in Tables 1 and 2. An effect of using the survey design is that it tends to produce a larger error term, thereby decreasing likelihood that a term is flagged as statistically significant. The NHANES II Statistical Review Panel, CDC and Du Pont have previously all acknowledged that incorporation of the survey design structure has little impact on the qualitative results based on their experiences with NHANES II analyses.

- In fact, use of the complex survey design in analyses affected by within study temporal trends is an area of disagreement among statisticians. It may be argued that only when the design accounts for time-related representativeness (which NHANES II did not) is it valid to use the design structure in such analyses.

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.**

   - This point is without basis. There will be <u>no</u> effect on the blood lead coefficient whether a reference PSU is picked or an intercept suppressed as Du Pont did. We suppressed the intercept so that the site effect coefficients have the convenient interpretation of denoting the site mean blood pressure. Suppressing the intercept imposes a linear constraint on the model parameters. Using an intercept term with a "restrict" statement (as Dr. Schwartz claimed he did at the March 11-12 1986 CASAC meeting) also imposes a linear constraint but a different one than we imposed.

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

   - Dr. Schwartz' comments here are perplexing and apparently reflect an inability to understand the "Descriptive Statistics" table (Appendix IV, first page, Du Pont's March 7, 1986 Comments). There is <u>no</u> coding error and these results are accurate. Age_YY variable does denote age in years centered about the overall age mean (see comment below). Dr. Schwartz equates root mean square of 5.8 with mean age of white males aged 40-59. This is incorrect. Root mean square is a standard deviation of the age data (whether age is centered about its mean or not). An age standard deviation of 5.8 is quite consistent with the expected value of 5.5 assuming ages are uniformly distributed over the white male age range, 40-59.

   - All of our analyses centered the predictor variables (except site terms) about their means which is clearly shown on page 18 of our March 7, 1986 comments. One may interpret Dr. Schwartz' comments on page 2 of his March 3, 1986 memo as implying centering may account for differences in BP-ln(BPb) coefficient values. Centering

TEH 0412639

covariates as Du Pont did has no effect on the estimated coefficients such as BP-ln(BPb). It is accepted and widespread statistical practice to center predictor variables in order to remove nonessential correlation among predictor variables. This results in smaller standard errors and increased sensitivity to detect significant effects. Numerous statistical references can be provided to support use of this procedure. (As an example, if one were to plot age and age-squared data against each other an obvious increasing trend relationship would be seen. By entering these two variables as such in the regression equation the effects of age and age-squared would be highly confounded due to nonessential correlation. By centering the age data about its mean and then squaring the centered age data to form age-squared the nonessential correlation is removed. A plot of the centered age data against age-squared formed in this way will demonstrate no linear trend.)

- Centering the predictor variables, in fact, is very important in the analyses including site effects because of the temporal trends inherent in NHANES II. If this were not done it would be expected that the sensitivity of the analysis to detect a BP-ln(BPb) relationship would be diminished. It is apparent from the results shown by Harlan et al in Table 2 that they standardized their predictor variables. Standardizing signifies that the predictor variables are centered about their mean and divided by their standard deviation. Division by a standard deviation enables a fair comparison among all coefficients in a given model. Standardization is proper and widely used and also has the benefits of centering but expresses the coefficients in terms of standard deviation units rather than the physical measurements.

6. **Analyses performed by Schwartz et al, subsequent to the Pirkle et al (1985) paper, have used 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.**

- Our response here is the same as in item 1. Du Pont's analyses attempted to mimic the published analyses and not deliberately introduce other considerations that might confuse the effect on the BP-ln(BPb) coefficient by adjusting for site. It was Dr. Schwartz in response to Du Pont's analyses who introduced these other considerations. The age group men 20-74 and the other variables alluded to in (1.) above were not considered in the published papers of Harlan et al and Pirkle et al.

TEH 0412640

DUP050453694

Summary of Items 1-6

The concern you expressed in your letter to me of April 4, 1986 about the inconsistency in the magnitude of the BP-ln(BPb) coefficients estimated by Dr. Schwartz and Du Pont is certainly warranted. Based on the foregoing remarks it may be due in part to differences in

- o explanatory variables considered
- o male age groups (but not for 40-59 group)
- o dietary variable metric
- o sample size for male 40-59 group (Du Pont excluded fingerprick BPb's also)
- o other potential reasons not apparent

It is not due to

- o data centering
- o age_YY data coding errors
- o suppression of the intercept
- o weighting or lack thereof (see Tables 1 and 2)
- o use or nonuse of SURREGR

Recommendations

The question of the consistency of the coefficient values could be addressed very directly. Since Dr. Schwartz has not yet mimicked the published analyses while adjusting properly for site, I suggest he do this. Thus, I would like to see his results for all the age-race-sex groups, both adjusted and unadjusted for sites, considered in Tables 1 and 2 using the model terms reported in the published papers. Then, redo those analyses but with the additional covariates he added to his recent analyses that were not considered in the published papers. The totality of these analyses will provide some useful comparisons to help better assess the total picture.

TABLE 1

Comparison of BP-ln(BPb) Coefficients and Regression Summaries
Diastolic Blood Pressure

	Unadjusted for Sites						Adjusted for Sites					
	n	R^2	Coeff	Std Error	t	P	n	R^2	Coeff	Std Error	t	P
UNWEIGHTED REGRESSIONS												
Harlan et al (1985) Groups												
Males (12-74)	2789	.24	2.017	.580	3.477	.0005	2789	-	0.360	.606	.593	.553
Males (21-55)	1306	.23	2.213	.843	2.623	.0088	1306	-	-.0242	.877	-.028	.978
White Males (21-55)	1175	.24	2.582	.875	2.950	.0032	1175	-	.289	.909	.318	.750
Males (56-74)	913	.07	2.228	1.104	2.018	.0439	913	-	.802	1.185	.677	.499
White Males (56-74)	821	.08	1.665	1.165	1.429	.1534	821	-	.0351	1.259	.028	.978
Females (12-74)	2918	.27	1.222	.549	2.224	.0262	2918	-	-.00141	.605	-.002	.998
Females (21-55)	1382	.26	1.245	.785	1.586	.1130	1382	-	-.170	.855	-.199	.842
Females (56-74)	1008	.15	0.709	1.024	.693	.4890	1008	-	-.242	1.157	-.209	.835
Pirkle et al (1985) Group												
White Males (40-59)	574	.18	4.012	1.246	3.219	.0014	575	-	1.925	1.385	1.389	.165
WEIGHTED REGRESSIONS												
Harlan et al (1985) Groups												
Males (12-74)	2789	.27	1.788	.557	3.209	.0013	2789	-	.047	.578	.081	.936
Males (21-55)	1306	.23	1.910	.821	2.325	.0202	1306	-	.044	.849	.052	.959
White Males (21-55)	1175	.24	2.165	.849	2.549	.0109	1175	-	.150	.879	.171	.865
Males (56-74)	913	.10	2.578	1.087	2.372	.0179	913	-	.802	1.184	.677	.499
White Males (56-74)	821	.11	2.228	1.330	1.966	.0496	821	-	.559	1.231	.455	.650
Females (12-74)	2918	.27	.870	.536	1.623	.1048	2918	-	-.423	.589	-.719	.472
Females (21-55)	1382	.25	1.179	.775	1.521	.1286	1382	-	-.231	.847	-.273	.785
Females (56-74)	1008	.18	-1.416	1.008	-1.406	.1601	1008	-	-1.877	.076	-.040	.968
Pirkle et al (1985) Group												
White Males (40-59)	574	.20	3.682	1.218	3.02	.0026	575	-	1.750	1.371	1.276	.203

NOTE: R^2 summaries in SAS statistical computing package are misleading when intercept is suppressed in site-adjusted analyses.

PFIEFE:13

TEH 0412642

DUP050453696

N 31008.01

TABLE 2

Comparison of BP-ln(BPb) Coefficients and Regression Summaries
Systolic Blood Pressure

	Unadjusted for Sites						Adjusted for Sites					
	n	R^2	Coeff	Std Error	t	P	n	R^2	Coeff	Std Error	t	P
UNWEIGHTED REGRESSIONS												
Harlan et al (1985) Groups												
Males (12-74)	2794	.23	3.434	.933	3.680	.0002	2794	-	1.947	.975	1.996	.046
Males (21-55)	1307	.17	3.560	1.255	2.840	.0046	1307	-	1.081	1.311	.825	.410
White Males (21-55)	1175	.18	3.821	1.297	2.945	.0033	1175	-	1.205	1.352	.891	.373
Males (56-74)	915	.06	5.219	1.960	2.663	.0079	915	-	3.908	2.104	1.858	.064
White Males (56-74)	823	.06	5.165	2.125	2.431	.0153	823	-	3.980	2.300	1.730	.084
Females (12-74)	2921	.39	.199	.932	.213	.8311	2921	-	-.505	1.049	-.480	.631
Females (21-55)	1385	.28	-.694	1.140	-.609	.5428	1385	-	-1.542	1.272	-1.213	.226
Females (56-74)	1007	.14	1.897	2.053	.924	.3555	1007	-	1.702	2.342	.727	.468
Pirkle et al (1985) Group												
White Males (40-59)	553	.18	6.271	1.999	3.14	.0018	553	-	3.461	2.231	1.551	.122
WEIGHTED REGRESSIONS												
Harlan et al (1985) Groups												
Males (12-74)	2794	.24	2.650	.859	3.080	.0021	2794	-	1.103	.896	1.230	.219
Males (21-55)	1307	.17	3.257	1.204	2.710	.0069	1307	-	1.136	1.260	.902	.368
White Males (21-55)	1175	.18	3.448	1.245	2.769	.0057	1175	-	1.256	1.303	.964	.336
Males (56-74)	915	.07	5.529	1.941	2.849	.0045	915	-	4.189	2.069	2.025	.043
White Males (56-74)	823	.07	5.705	2.082	2.740	.0063	823	-	3.980	2.300	1.730	.084
Females (12-74)	2921	.36	-.646	.862	-.750	.4533	2921	-	-1.377	.967	-1.424	.155
Females (21-55)	1385	.26	-.883	1.132	-.780	.4354	1385	-	-1.249	1.262	-.990	.323
Females (56-74)	1007	.14	1.897	2.053	.924	.3555	1007	-	.086	2.306	.037	.970
Pirkle et al (1985) Group												
White Males (40-59)	553	.20	7.312	1.933	3.78	.0002	553	-	3.985	2.196	1.815	.070

NOTE: R^2 summaries in SAS statistical computing package are misleading when intercept is suppressed in site-adjusted analyses.

PFIEFE:13