ECAO-...
WS-5-25

DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service

Centers for Disease Control
Atlanta GA 30333
FTS 236-4111

December 7, 1982

David E. Weil, Ph.D.
Project Manager, Environmental Criteria
and Assessment Office (MD-52)
U. S. Environmental Protection Agency
Research Triangle Park, North Carolina 27711

Dear David:

Enclosed is the comments on the DuPont Analysis of the NHANES II blood lead data which was done by Dr. James Pirkle of my staff. I understand that Dr. Annest is separately and specifically addressing the Lucas paper on overascertainment. We will send separately a more detailed analysis as it relates to Hammond and Bornchein concerns about the reference laboratories. That concern has no bearing on the blind quality control that was in place during the analysis of the majority of the NHANES II data, which is discussed on page 11 of the Pirkle comments.

Would you please acknowledge receipt of this material. See you on the 18th.

Sincerely yours,

V

Vernon N. Houk, M.D.
Acting Director
Center for Environmental Health

Enclosure

N33897

TEH 0533288

Subject: Comments on the Dupont Analysis of the NHANES II Blood Lead Data

From: James L. Pirkle M.D. Ph.D.
Clinical Chemistry Division
Center for Environmental Health
Centers for Disease Control

Date: December 1, 1982

N33897.01

TEH 0533289

DUP050034556

INTRODUCTION

Recent comments from Dupont (1) on the NHANES II data warrant careful analysis. Dupont asserts that conclusions drawn on the NHANES II data are "suspect" because (1) demographic characteristics of subject sample groups changed during the study and (2) analytical error leads to overestimation of the percentage of the population with blood lead levels of 30 or more. In addition, DuPont claimed there were "several potential difficulties" in the blind quality control data that "could impact on the reported interpretations with NHANES II trends."

The points concerning accounting for demographic variables and the "potential difficulties" in the blind quality control are summarized and discussed below. The question of analytical error leading to an overestimation of children with lead toxicity is addressed here and by Annest (2).

SUMMARY

(1) After accounting for the variability of demographic variables and incorporating the complex survey design, a statistically significant downward trend of 31.1% (blacks and whites together) is still present in the NHANES II lead values over the period of the survey. This drop is noted in blacks (27.1%), whites (31.8%), and in different age groupings of whites: .5-5 yrs (32.1%), 6-17 yrs. (31.3%) and 18-74 yrs. (35.4%). (Blacks were not broken into age groupings due to small sample sizes.)

(2) Changes in leaded paint consumption is an unlikely explanation for the downward trend in blood lead levels since the trend is also present in adults.

(3) The best available data on dietary lead intake indicate that lead in the diet certainly was not declining over the survey period and may have been rising. This makes changes in dietary lead intake also an unlikely explanation of the downward trend.

(4) After accounting for the variability of demographic variables and incorporating the complex survey design, lead in gasoline is highly statistically significant ($p < .0001$) and <u>alone</u> can account for the magnitude of the downward trend in lead values seen with time. Over the survey period, changes in gasoline lead account for the following decreases in blood lead levels: blacks and whites - 34.4%, blacks - 25.6%, whites - 35.3%, whites .5-5 yrs old - 34.7%, whites 6-17 yrs old - 34.9% and whites 18-74 yrs old - 34.0%.

(5) Multiple regression analysis of the NHANES II lead data must incorporate the complex survey design. Otherwise, all of the F statistics in the regression will be incorrect. The DuPont analysis did <u>not</u> incorporate the complex survey design.

(6) The contention that the downward trend can be explained by variability in the sampling of urban areas (cities with greater than 1 million population) has previously been shown to be incorrect (3). In fact, in subpopulations with no urban dwellers (i.e. small cities with less than 1 million population and rural areas), the trend is still present. Furthermore, after accounting for urbanization, the trend is still present in the total population (see item (1) above).

(7) The NHANES II data either correctly estimate or underestimate the number of persons with lead toxicity. Inasmuch as the original "toxicity" criteria of 30 ug/dl was based on blood lead data which included measurement error, the only pertinent question is whether the measurement error in the NHANES II data is more or less than the data on which the 30 ug/dl cutoff was established. If the NHANES II measurement error were higher, then there would be potential for overestimating the number of persons with lead toxicity. Actually, the measurement error is equal to or less than the measurement error from standard or reference laboratories and thus the data either correctly estimate or underestimate the number of persons with lead toxicity.

(8) The blind quality control data show no significant downward trend in lead values due to laboratory measurement.

DISCUSSION

ACCOUNTING FOR THE DOWNWARD TREND IN THE NHANES II BLOOD LEAD DATA

Demographic Variables

Age, sex, race and urbanization are variables known to effect blood lead levels. Ideally, standard stepwise regression techniques could be used on the 9937 lead values to control for the influence of these variables. However, the NHANES II data was collected utilizing a stratified sampling design which has to be incorporated into the regression analysis for proper estimation of errors. The regression program SURREGR (4) permits multiple regression analysis incorporating the effects of the survey design. In the current analysis, the dependent variable was the natural log of blood lead.

Originally, the lead data was analyzed with SURREGR in subgroups broken down by age, sex, race, urbanization, income level, season, and geographic location. The downward trend was consistently greater than 30% and statistically significant in all subgroups except urban areas with more than one million population. Normally, there would be plenty of degrees of freedom to include all these variables plus their interactions in a stepwise linear regression approach. However, incorporating the complex survey design reduces the degrees of freedom to 32. Thus some selection of variables is necessary. The following method was used to select the most significant variables to enter into SURREGR:

(1) indicator variables were constructed representing:
 -- sex
 -- age (.5-5 yrs, 6-17 yrs, 18-74 yrs)
 -- urbanization (1 million or more, less than 1 million, rural)
 -- SMSA (whether the person lived in or out of an SMSA)
 -- all the first order interactions of the above variables

(2) A model using these variables was fit using stepwise linear regression (5) with an F-to-enter of 2.7 and F-to-remove of 2.7 (6). In this manner, variables with even marginal significance would not be excluded but clearly insignificant variables could be dropped. This process did not incorporate the complex survey design but consistently included more variables than found to be significant following subsequent analysis with SURREGR.

(3) The remaining variables were then used with SURREGR. Depending on the regression, the variable "time" or the variable "gas" was entered in both a linear and quadratic term.

A manual backward elimination was performed with SURREGR, deleting the least significant variable in each regression run until all variables were significant at approximately the .05 level. This elimination procedure had very little effect on the coefficients of "time" or "gas".

Regression on TIME

When time and time-squared were entered as variables along with the demographic variables mentioned above, time-squared was consistently significant at the $p < .0001$ level. The regression results by race and selected age groups are given in Appendix A. After accounting for the demographic variables, the downward trend over time was still present. The amount of the downward trend is given by race and selected age groups in Table 1.

Table 1
Regression Results for the "Time" Variable After Controlling Age, Sex, Urbanization and SMSA

	Blood Lead at Start of Survey	Blood Lead at End of Survey	Difference	Percent Decrease
Blacks & Whites	14.45 ug/dl	9.96	4.50	31.1
Blacks	16.23	11.84	4.39	27.1
Whites	14.24	9.71	4.53	31.8
Whites .5-5 yrs	14.26	9.69	4.57	32.1
Whites 6-17 yrs	14.79	10.17	4.62	31.3
Whites 18-74 yrs	13.38	8.64	4.74	35.4

Lead in Paint

In attempting to account for this downward trend, the different major sources of lead exposure must be examined. Lead in paint is a well known source of lead exposure for children. However, the downward trend also occurred in teens and adults which makes decreased leaded paint consumption an unlikely explanation of the trend.

Lead in the Diet

The Food and Drug Administration (FDA) has measured lead in its Total Diet ("Market Basket") Studies since 1973 (7). The survey includes 120 individual food items representative of the 2-week diet of a teenage male for the regions of the country in which it is collected. The results from 1976 through 1980 are given in Table 2.

Table 2
Lead in the Diet (micrograms/day)
Teenage Male

1976	71.1
1977	79.3
1978	95.1
1979	81.7
1980	82.9

No downward trend is present, suggesting that diet is also an unlikely explanation of the decrease in blood lead levels seen over the course of the NHANES II survey.

Lead in gasoline

The only other widespread source of lead exposure (other than occupational) is from leaded gasoline. The amount of lead used in gasoline production across the nation is available from the quarterly refiner's reports to EPA. During a 3-month period, the NHANES II caravans on-the-average sampled 4 different sites in the USA and over a 6-month period they sampled on-the-average 8 sites. Three gas variables were tested to cover both the 3-month and 6-month breakdown of gasoline lead data:

(1) Gas1 = amount of lead used in gasoline for each of the quarters of the year during the NHANES II survey
(2) Gas2 = amount of lead used in gasoline for the January-June and July-December 6-month periods during the NHANES II survey
(3) Gas3 = amount of lead used in gasoline for the April-September and October-March 6-month periods during the NHANES II survey

The regression procedure included the demographic variables itemized in Appendix A and was performed separately for each of these gas variables. A "gas-squared" term was also permitted to enter each model to account for a curvilinear relation. The regression results by race and selected age groups are given in Appendix A. The amount of downward trend attributable to gasoline lead is tabulated by race and selected age groups in Table 3 and the gas variable with the generally most conservative results (Gas1) is shown in Figure 1. AFTER ACCOUNTING FOR DEMOGRAPHIC VARIABLES, THE GASOLINE LEAD VARIABLE (GAS1, GAS2 OR GAS3) WAS ALWAYS HIGHLY STATISTICALLY SIGNIFICANT (p .0001) AND ALONE COULD ACCOUNT FOR THE DOWNWARD TREND IN LEAD VALUES OVER TIME SEEN IN THE NHANES II SURVEY.

PERCENT DECREASE IN BLOOD LEAD VALUES EXPLAINED BY LEAD

NHANES II BLOOD LEAD DATA

AFTER ACCOUNTING FOR AGE, SEX, URBANIZATION, SMSA

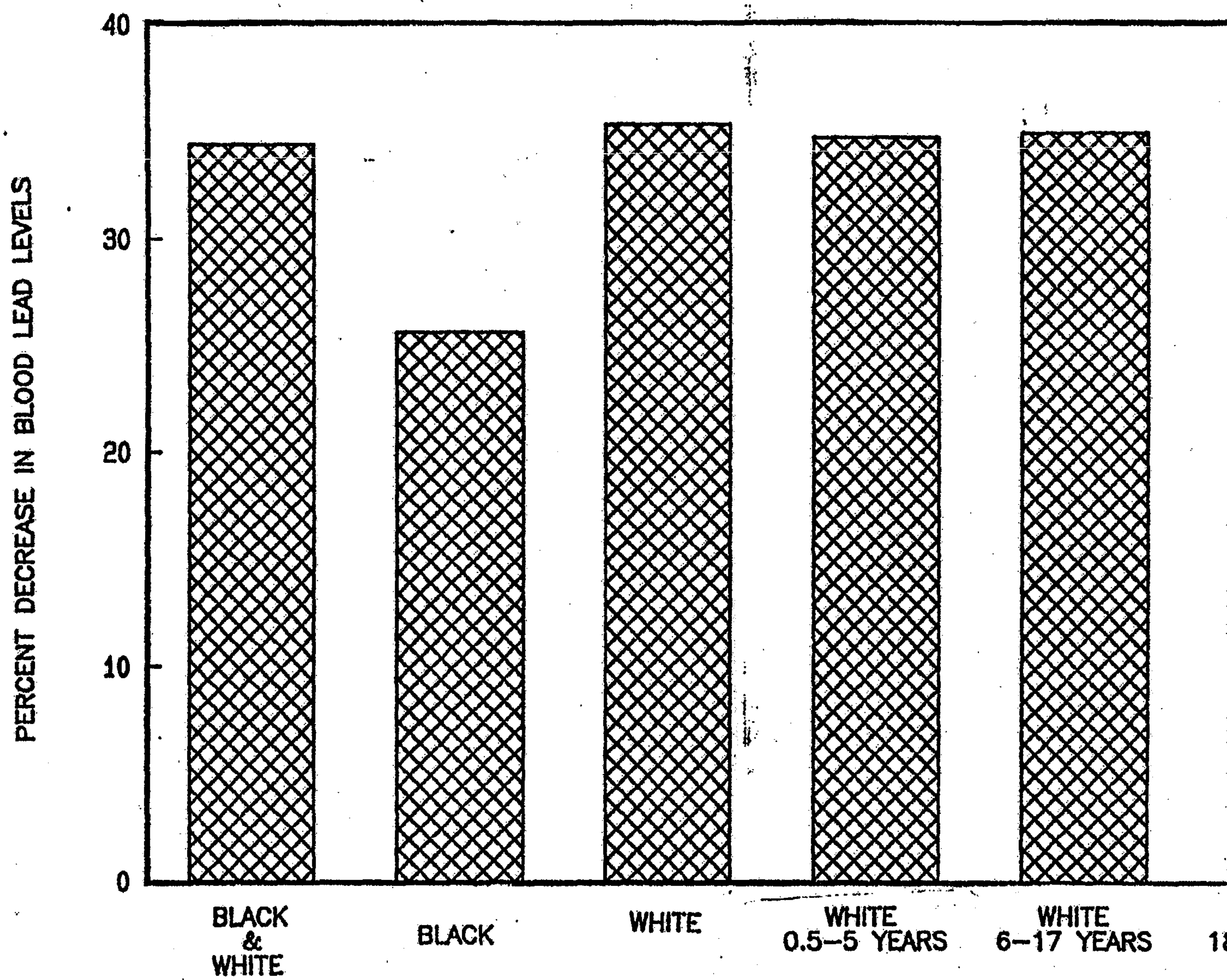

FIGURE 1

Table 3
Regression Results for the Three Gas Variables After Accounting for Age, Sex, Urbanization and SMSA

	Blood lead during initial period of NHANES survey*	Blood lead during final period of NHANES survey*	Difference	Percent decrease
Blacks & Whites				
Gas1	14.72	9.67	5.06	34.4
Gas2	14.94	9.43	5.50	36.9
Gas3	14.80	9.96	4.97	33.6
Blacks				
Gas1	16.18	12.05	4.13	25.6
Gas2	16.72	11.66	5.07	30.3
Gas3	16.31	11.84	4.39	25.5
Whites				
Gas1	14.53	9.40	5.13	35.3
Gas2	14.76	9.19	5.58	37.8
Gas3	14.68	9.47	5.21	35.5
Whites .5-5 yrs				
Gas1	14.44	9.43	5.01	34.7
Gas2	14.65	9.27	5.38	36.7
Gas3	14.60	9.49	5.10	35.0
Whites 6-17 yrs				
Gas1	15.03	9.79	5.25	34.9
Gas2	15.33	9.63	5.70	37.2
Gas3	15.22	9.93	5.29	34.8
Whites 18-74 yrs				
Gas1	13.36	8.82	4.54	34.0
Gas2	13.75	8.34	5.41	39.3
Gas3	13.65	8.69	4.97	36.4

*see text for definition of 3-month initial period of Gas1 and 6-month initial period of Gas2 and Gas3.

Importance of Incorporating the Complex Sample Design into the Regression

The complex sample design must be considered in the regression analysis or the F statistics will be incorrect. The DuPont analysis did not incorporate the complex sample design.

Explaining the Downward Trend by Variability in Sampling of the Urban Areas (Cities with more than one million population)

This contention has previously been shown to be incorrect. In fact, in subpopulations with no "urban dwellers" (i.e. small cities with less than one million population or rural areas), the trend is still present. Furthermore, after accounting for urbanization as a demographic variable the trend is still present in the total population (see analysis above).

IN SUMMARY, after accounting for all these demographic variables and considering alternative etiologies, lead in gasoline remains very significantly related to blood lead levels and the most reasonable explanation of the downward trend in the NHANES II lead data.

THE EFFECT OF ANALYTICAL ERROR ON ESTIMATING THE PERCENTAGE OF THE POPULATION WITH LEAD TOXICITY FROM THE NHANES II DATA

"Lead toxicity" is a medical rather than a numeric determination. For example, suppose some unusual systematic bias pervaded all lead measurements causing them to be two times their true value. The same people who medically had "lead toxicity" would still have it, only the lead value with which they were associated would be different.

This concept is important when considering the 30 ug/dl cutoff for defining "lead toxicity". Physicians examined and followed individuals and determined that those with lead values of 30 ug/dl or more medically had "lead toxicity". These blood lead values, on which the 30 ug/dl cutoff was based, came from several laboratories and all contained measurement error. If the measurement error was removed from this data, the cutoff of 30 ug/dl would decrease to a lower lead value so, for example, 23 or 27 ug/dl would be the cutoff value for "lead toxicity". This has not happened. No one has removed measurement error from the data on which the 30 ug/dl cutoff was based. Consequently, the laboratory determination of "lead toxicity" by blood lead levels is based on data which includes measurement error.

The essential point in determining if the NHANES II lead data falsely classify persons "lead toxic" due to laboratory measurement error, is the comparison of the amount of measurement error in NHANES II data relative to the amount of measurement error in the studies on which the 30 ug/dl cutoff was based. If the NHANES II measurement error is greater than the measurement error of these other studies, then some persons would be falsely classified "lead toxic" due to this relatively greater measurement error (i.e. false positives). If the NHANES II measurement error is less than the measurement error of these studies on which the 30 ug/dl cutoff was based then the NHANES II data underestimates the number of persons who should be classified "lead toxic".

To make this comparison, the measurement error of the studies on which the 30 ug/dl cutoff was based needs to be known. This cannot be determined exactly since multiple studies using several laboratories influenced the setting of the 30 ug/dl cutoff. However, the NHANES II measurement error can be compared to the error from standard and reference laboratories which provide good estimates of what the error was in the studies of interest.

The Centers for Disease Control (CDC) conducts a Blood Lead Proficiency Testing Program which has estimates of the coefficients of variation for standard and reference laboratories. The 1980 report (8) has data which are near the mean levels for the two NHANES II blind quality control pools (13.5 ug/dl and 25.5 ug/dl). This data was compiled from about 27 reference laboratories and about 150 standard laboratories. The coefficient of variation of lead measurements is known to increase as the magnitude of the lead value decreases. A comparison of these standard and reference labs with the NHANES II data is given below.

Table 4

Comparison of Measurement Error in NHANES II Data with that of Standard and Reference Laboratories

	Lower Lead Level		Higher Lead Level	
	Mean (ug/dl)	C.V. (%)	Mean (ug/dl)	C.V. (%)
Standard Laboratory	not available		25.1	27.9
Reference Laboratory	15.9	24.7	24.8	14.4
NHANES II Lead Data	13.5	16.0	25.5	12.5

The important conclusion from Table 4 is that the NHANES II measurement error is at least as good as and probably better than standard or reference laboratories. Without any additional calculations, it can be concluded that RELATIVE TO THE STUDIES ON WHICH THE 30 UG/DL CUTOFF FOR "LEAD TOXICITY" WAS ESTABLISHED, THE NHANES II DATA EITHER CORRECTLY ESTIMATE OR UNDERESTIMATE THE NUMBER OF PERSONS WITH "LEAD TOXICITY".

It is altogether a separate issue as to whether measurement error correction procedures should be implemented on lead data from all studies. If so, it would certainly be incorrect to adjust the NHANES II data for measurement error and not adjust the data on which the 30 ug/dl cutoff was based. This would result in falsely assigning 30 ug/dl as the cutoff for "lead toxicity" in the NHANES II data when the actual measurement-error-free cutoff would be less than 30 ug/dl.

In summary, after considering the effects of measurement error, the NHANES II data either correctly estimate or underestimate the number of persons with lead toxicity.

QUALITY CONTROL DATA

The NHANES II lead samples were covered by two independent quality control systems: a bench and a blind quality control. The bench quality control started at the beginning of the survey and finished at the end. The blind quality control began approximately 12 months after the survey began and continued uninterrupted through the end of the survey. Approximately 90% of the drop in lead levels occurred while both bench and blind quality control were in effect. If the average of replicate values of either bench or blind quality control specimens fell outside of their esablished 95% confidence limits, the run was repeated. Also, if replicate absorbance values for any given specimen differed by more than 0.25 absorbance units or the difference between calculated concentrations for duplicates was greater than 7 ug/dl, the analysis was repeated.

The bench quality control utilized a number of bovine pools over the course of the survey so a long term trend in lead values could not be statistically tested on this data. The blind quality control used two pools simultaneously for the entire period of blind quality control (a high pool with mean of 25.5 ug/dl and a low pool with mean of 13.5 ug/dl). In order to statistically test for a trend in the entire blind quality control, the high and the low pools were combined. First, the lead data from the low and high blind pools were checked to insure they were approximately normally distributed. Then the values in each pool were transformed using a z transformation; that is, (lead value - lead pool mean)/lead pool standard deviation. Finally, the transformed data from each pool, having mean of zero and variance of one, were combined and linear regression of z on "time" was performed. The "time" variable was defined to correspond with the chronology of the collection of the NHANES II specimens. No significant <u>linear</u> trend was found from the regression ($p=.69$, $n=544$).

DuPont separately analyzed the low pool and high pool and found that the high pool rose very slightly over time and the low pool declined very slightly over time and that there were statistically significant curvilinear terms in each pool when analyzed separately.

We tested the entire blind quality control, and though there was no statistically significant linear trend with time, a statistically significant ($p=.0001$) quadratic term was found. This quadratic term accounts for a slight dome curvature in the blind quality control data but does <u>not</u> account for <u>any</u> decline in the blood lead levels with time. On the contrary, including this quadratic term with the linear term predicts a very small net increase in lead levels of .09 ug/dl from the beginning to the end of the blind quality control data. If a first-order correction is applied to the data to compensate for the slight dome curvature, the average absolute amount of the correction is only about 1.8% of the lead value. The adjustment changes the coefficients of the "time" or "gas" variable by about 3%. For the sake of detail, all the lead results in this report include this slight adjustment.

In summary, a small statistically significant curvature is present in the blind quality control data. However, it does <u>not</u> explain <u>any</u> of the downward trend seen in the NHANES blood lead data nor diminish the significance of the relationship between blood lead levels and gasoline lead.

CONCLUSIONS

After considering the laboratory quality control and the effect of demographic variables, a statistically significant downward trend of about 31% remains in the NHANES II blood lead levels over the course of the survey. Lead in paint is an unlikely explanation of this trend since the trend is present in adults. The best available data indicate lead in the diet did not decrease over the survey period making it also an unlikely explanation of the trend. The only other widespread source of lead (other than occupational) is leaded gasoline. Multiple regression analysis accounting for demographic variables and incorporating the complex survey design indicates that lead in gasoline is significantly related to NHANES II blood lead levels and it alone can explain the entire magnitude of the downward trend seen in the NHANES II blood lead data.

Appendix A

Regression Variables

The variables for the regression model were constructed as follows:

Characteristic	Variable value
male	male=1
female	male=0
age .5-5 yrs	child=1 teen=0
age 6-17 yrs	child=0 teen=1
age 18-74 yrs	child=0 teen=0
live in SMSA	outsmsa=0
live outside SMSA	outsmsa=1
live in city with one million or more pop.	small=0 rural=0
live in city with less than million pop.	small=1 rural=0
live in rural area	small=0 rural=1

Interaction terms:
- maleout = male X outsmsa
- smallout = small X outsmsa
- malesmal = male X small
- malerurl = male X rural
- malechild = male X child
- maleteen = male X teen
- childsmal = child X small
- teensmal = teen X small
- childrural = child X rural
- teenrurl = teen X rural
- childout = child X outsmsa
- teenout = teen X outsmsa

For the regression on blacks, whites or blacks + whites, all of the above terms were initially in the model. For regression on whites .5-5 yrs, whites 6-17 yrs or whites 18-74 yrs, the population was already divided into an age group so no age variables were used in the model.

Regression Results on "Time" Variable

Time and time-squared were entered as variables together with the demographic variables listed above. Time was defined as the number of days from 2/20/76 to the sample collection date divided by 2800. This reduced the magnitude of the time variable to approximately that of the 0-1 indicator variables. Some of the final models contain interaction terms without the corresponding main effects. All of the main effects were reintroduced into each of these final models and the regression rerun to insure that the omission of these main effects did not significantly alter the "time" coefficients or the percent decrease in lead calculated over time (Table 1). The "time" coefficients changed less than 2% and the percent decrease in lead changed less than 0.5% (e.g. 35.5% to 35.0%).

The final models resulted from the manual backward elimination procedure described on page 3. In each regression the denominator degrees of freedom (df) is 32. Results are given below for each race and selected age groups of whites. The time-squared variable is consistently significant.

Group: Black & White

Variable	F Stat.	DF	Prob.	Coeff.	Variance
MALE	901.9	1	0.0000	0.3420	0.0001300
MALECHILD	355.0	1	0.0000	-0.3361	0.0003200
CHILD	210.1	1	0.0000	0.3002	0.0004300
MALETEEN	87.6	1	0.0000	-0.1921	0.0004200
TIMESQ	67.6	1	0.0000	-1.3576	0.0272500
RURAL	23.6	1	0.0000	-0.1207	0.0006200
CHILDRURAL	13.2	1	0.0010	-0.0833	0.0005300
MALEOUT	11.7	1	0.0017	0.0631	0.0003400
SMALLOUT	10.6	1	0.0027	-0.0965	0.0008800
TEEN	4.9	1	0.0335	-0.0387	0.0003000
INTERCEPT				2.5692	0.0004700
Overall Model	256.2	10	0.0000		

Group: Black

Variable	F Stat.	DF	Prob.	Coeff.	Variance
MALE	211.0	1	0.0000	0.4035	0.0007700
CHILD	120.8	1	0.0000	0.4690	0.0018215
MALECHILD	52.9	1	0.0000	-0.4163	0.0032780
TIMESQ	51.7	1	0.0000	-1.1492	0.0255330
MALETEEN	23.2	1	0.0000	-0.2073	0.0018554
RURAL	10.5	1	0.0028	-0.2170	0.0044866
SMALLOUT	6.7	1	0.0142	-0.2205	0.0072169
OUTSMSA	4.3	1	0.0456	0.1616	0.0060313
INTERCEPT				2.6232	0.0011671
Overall Model	72.6	8	0.0000		

Group: White

Variable	F Stat.	DF	Prob.	Coeff.	Variance
MALE	1334.1	1	0.0000	0.3619	0.0000980
MALECHILD	245.3	1	0.0000	-0.3229	0.0004300
CHILD	109.3	1	0.0000	0.2425	0.0005400
MALETEEN	59.6	1	0.0000	-0.1880	0.0005900
TIMESQ	54.8	1	0.0000	-1.3938	0.0354706
RURAL	11.4	1	0.0020	-0.1011	0.0009000
TEEN	7.4	1	0.0105	-0.0540	0.0004000
SMALL	6.4	1	0.0165	-0.0581	0.0005300
CHILDRURAL	4.5	1	0.0422	-0.0456	0.0004600
INTERCEPT				2.5664	0.0005300
Overall Model	146.7	9	0.0000		

Group: White .5-5 yrs.

Variable	F Stat.	DF	Prob.	Coeff.	Variance
MALE	1212.7	1	0.0000	0.3169	0.0000830
TIMESQ	55.6	1	0.0000	-1.4079	0.0356367
RURAL	8.9	1	0.0055	-0.0683	0.0005300
INTERCEPT				2.5218	0.0004100
Overall Model	392.7	3	0.0000		

Group: White 6-17 yrs.

Variable	F Stat.	DF	Prob.	Coeff.	Variance
MALE	1358.2	1	0.0000	0.3426	0.0000860
TIMESQ	49.1	1	0.0000	-1.3654	0.0379896
RURAL	12.6	1	0.0012	-0.1043	0.0008600
SMALL	7.7	1	0.0092	-0.0614	0.0004900
INTERCEPT				2.5793	0.0004300
Overall Model	330.6	4	0.0000		

Group: White 18-74 yrs.

Variable	F Stat.	DF	Prob.	Coeff.	Variance
MALE	58.3	1	0.0000	0.1490	0.0003800
TIMESQ	47.0	1	0.0000	-1.5944	0.0540572
RURAL	5.5	1	0.0258	-0.0670	0.0008200
INTERCEPT				2.5452	0.0009100
Overall Model	35.3	3	0.0000		

Regression on the Lead in Gasoline Variables

Each of the gas variables is in units of 1000 tons of lead divided by 100 to reduce the magnitude of the variable to approximately that of the 0-1 indicator variables. As with the time regressions, some of the final models contain interaction terms without the corresponding main effects. All of the main effects were reintroduced to these final models and the regression rerun to insure that their omission did not significantly alter the "gas" coefficients or the percent decrease in blood lead levels accounted for by lead in gasoline. The "gas" coefficients changes less than 3.5% and the percent decrease in blood lead levels accounted for by gasoline lead changed less than 1.1% (e.g. 34.4% to 33.3%).

The final models resulted from a manual backward elimination procedure (decribed on page 3) performed separately for each of the three gas variables. In each regression the denominator degrees of freedom (df) is 32. Results are given below for each race and selected age groups of whites. The "gas" variable (Gas1, Gas2, or Gas3) is consistently significant.

Lead in Gasoline Variable: Gas1

Group: Black & White

Variable	F Stat.	DF	Prob.	Coeff.	Variance
MALE	928.2	1	0.0000	0.37584	0.0001500
MALECHILD	327.6	1	0.0000	-0.33918	0.0003500
CHILD	200.4	1	0.0000	0.27565	0.0003800
MALETEEN	128.5	1	0.0000	-0.19898	0.0003100
GAS1	89.6	1	0.0000	1.42693	0.0227163
MALERURL	13.8	1	0.0008	-0.08638	0.0005400
SMALLOUT	5.6	1	0.0247	-0.05091	0.0004700
TEENOUT	4.4	1	0.0434	-0.07768	0.0013639
MALEOUT	4.0	1	0.0541	0.05167	0.0006700
INTERCEPT				1.78655	0.0044702
Overall Model	256.2	9	0.0000		

Group: Black

Variable	F Stat.	DF	Prob.	Coeff.	Variance
MALE	196.3	1	0.0000	0.39318	0.0007900
CHILD	145.3	1	0.0000	0.48829	0.0016415
GAS1	91.2	1	0.0000	0.99997	0.0109703
MALECHILD	68.2	1	0.0000	-0.44610	0.0029167
MALETEEN	18.5	1	0.0002	-0.23137	0.0028991
TEEN	6.4	1	0.0168	0.07328	0.0008400
TEENOUT	3.0	1	0.0919	-0.17989	0.0107213
MALEOUT	2.5	1	0.1288	0.08737	0.0030057
INTERCEPT				2.06451	0.0008300
Overall Model	110.2	8	0.0000		

TEH 0533304

Group: White

Variable	F Stat.	DF	Prob.	Coeff.	Variance
MALE	674.9	1	0.0000	0.34352	0.0001700
MALECHILD	250.3	1	0.0000	-0.32582	0.0004200
CHILD	126.9	1	0.0000	0.23235	0.0004300
GAS1	104.2	1	0.0000	1.47503	0.0208812
MALETEEN	59.1	1	0.0000	-0.18393	0.0005700
RURAL	22.3	1	0.0000	-0.13970	0.0008800
SMALL	13.3	1	0.0009	-0.10110	0.0007700
TEEN	7.7	1	0.0090	-0.05611	0.0004100
MALEOUT	6.1	1	0.0193	0.04796	0.0003800
INTERCEPT				1.83529	0.0038429
Overall Model	160.6	9	0.0000		

Group: White .5-5 yrs

Variable	F Stat.	DF	Prob.	Coeff.	Variance
MALE	570.9	1	0.0000	0.29737	0.0001500
GAS1	92.3	1	0.0000	1.44410	0.0225929
RURAL	23.8	1	0.0000	-0.14286	0.0008600
SMALL	14.1	1	0.0007	-0.10304	0.0007500
MALEOUT	5.9	1	0.0209	0.05095	0.0004400
INTERCEPT				1.83879	0.0041133
Overall Model	257.0	5	0.0000		

Group: White 6-17 yrs.

Variable	F Stat.	DF	Prob.	Coeff.	Variance
MALE	1174.3	1	0.0000	0.34271	0.0001000
GAS1	114.6	1	0.0000	1.45531	0.0187422
RURAL	20.6	1	0.0001	-0.12087	0.0007100
SMALL	14.9	1	0.0005	-0.09452	0.0006000
INTERCEPT				1.84415	0.0037919
Overall Model	300.8	4	0.0000		

Group: White 18-74 yrs.

Variable	F Stat.	DF	Prob.	Coeff.	Variance
MALE	66.6	1	0.0000	0.15381	0.0003600
GAS1	31.8	1	0.0000	1.40769	0.0622089
RURAL	5.1	1	0.0312	-0.07161	0.0010101
INTERCEPT				1.79792	0.0111279
Overall Model	35.5	3	0.0000		

Lead in Gasoline Variable: GAS2

Group: Blacks & Whites

Variable	F Stat.	DF	Prob.	Coeff.	Variance
MALE	909.4	1	0.0000	0.34076	0.0001300
MALECHILD	377.9	1	0.0000	-0.33764	0.0003000
CHILD	234.0	1	0.0000	0.30608	0.0004000
GAS2	100.6	1	0.0000	0.77418	0.0059587
MALETEEN	95.8	1	0.0000	-0.19386	0.0003900
RURAL	23.8	1	0.0000	-0.14514	0.0008900
MALEOUT	15.1	1	0.0005	0.06695	0.0003000
CHILDRURAL	14.9	1	0.0005	-0.09014	0.0005500
TEEN	5.4	1	0.0273	-0.03967	0.0002900
SMALL	3.8	1	0.0598	-0.05578	0.0008200
SMALLOUT	3.6	1	0.0657	-0.06046	0.0010061
INTERCEPT				1.80786	0.0048407
Overall Model	200.0	11	0.0000		

Group: Black

Variable	F Stat.	DF	Prob.	Coeff.	Variance
MALE	166.0	1	0.0000	0.38893	0.0009100
CHILD	136.7	1	0.0000	0.47271	0.0016347
MALECHILD	56.8	1	0.0000	-0.42610	0.0031954
GAS2SQ	36.2	1	0.0000	0.39912	0.0044039
MALETEEN	21.9	1	0.0001	-0.20361	0.0018978
MALERURL	6.8	1	0.0134	-0.17387	0.0044112
MALEOUT	5.2	1	0.0291	0.15268	0.0044635
SMALLOUT	4.2	1	0.0479	-0.10112	0.0024159
INTERCEPT				2.20442	0.0026534
Overall Model	46.7	8	0.0000		

Group: White

Variable	F Stat.	DF	Prob.	Coeff.	Variance
MALE	748.7	1	0.0000	0.34426	0.0001600
MALECHILD	262.3	1	0.0000	-0.32447	0.0004000
CHILD	129.9	1	0.0000	0.22909	0.0004000
GAS2	83.6	1	0.0000	0.79861	0.0076249
MALETEEN	64.0	1	0.0000	-0.18931	0.0005600

RURAL	19.5	1	0.0001	-0.12810	0.0008400
SMALL	8.7	1	0.0059	-0.07655	0.0006700
TEEN	8.2	1	0.0073	-0.05520	0.0003700
MALEOUT	6.6	1	0.0148	0.04425	0.0002900
INTERCEPT				1.77408	0.0063072
Overall Model	165.4	9	0.0000		

Group: White .5-5 yrs.

Variable	F Stat.	DF	Prob.	Coeff.	Variance
MALE	621.6	1	0.0000	0.29721	0.0001400
GAS2	71.9	1	0.0000	0.77009	0.0082465
RURAL	20.3	1	0.0001	-0.13155	0.0008500
SMALL	9.0	1	0.0052	-0.07902	0.0006900
MALEOUT	5.7	1	0.0230	0.04678	0.0003800
INTERCEPT				1.78925	0.0068279
Overall Model	247.8	5	0.0000		

Group: White 6-17 yrs

Variable	F Stat.	DF	Prob.	Coeff.	Variance
MALE	1343.7	1	0.0000	0.34215	0.0000870
GAS2	77.0	1	0.0000	0.78245	0.0079501
RURAL	16.2	1	0.0003	-0.11078	0.0007600
SMALL	8.6	1	0.0061	-0.07120	0.0005900
INTERCEPT				1.79346	0.0067350
Overall Model	315.5	4	0.0000		

Group: White 18-74 yrs

Variable	F Stat.	DF	Prob.	Coeff.	Variance
GAS2	66.5	1	0.0000	0.84111	0.0106363
MALE	62.3	1	0.0000	0.14837	0.0003500
RURAL	5.7	1	0.0235	-0.07527	0.0010012
INTERCEPT				1.68533	0.0074795
Overall Model	45.5	3	0.0000		

Lead in Gasoline variable: GAS3

Black & White

Variable	F Stat.	DF	Prob.	Coeff.	Variance
MALE	880.3	1	0.0000	0.34937	0.0001400
MALECHILD	336.3	1	0.0000	-0.33742	0.0003400
CHILD	212.8	1	0.0000	0.30832	0.0004500
GAS3	102.8	1	0.0000	0.68839	0.0046114
MALETEEN	85.3	1	0.0000	-0.18984	0.0004200
RURAL	23.2	1	0.0000	-0.14733	0.0009300
CHILDRURAL	12.8	1	0.0011	-0.08462	0.0005600
SMALL	12.7	1	0.0012	-0.09803	0.0007600
MALEOUT	5.3	1	0.0279	0.04429	0.0003700
TEEN	4.3	1	0.0462	-0.03654	0.0003100
INTERCEPT				1.88479	0.0038981
Overall Model	185.9	10	0.0000		

Group: Black

Variable	F Stat.	DF	Prob.	Coeff.	Variance
MALE	204.2	1	0.0000	0.39472	0.0007600
CHILD	144.3	1	0.0000	0.47893	0.0015895
GAS3	103.5	1	0.0000	0.49571	0.0023736
MALECHILD	64.6	1	0.0000	-0.47215	0.0028230
MALETEEN	16.9	1	0.0003	-0.19477	0.0022499
INTERCEPT				2.08256	0.0013784
Overall Model	167.6	5	0.0000		

Group: White

Variable	F Stat.	DF	Prob.	Coeff.	Variance
MALE	641.2	1	0.0000	0.34266	0.0001800
MALECHILD	238.6	1	0.0000	-0.32357	0.0004400
CHILD	121.2	1	0.0000	0.23069	0.0004400
GAS3	101.0	1	0.0000	0.73855	0.0054024
MALETEEN	58.1	1	0.0000	-0.18459	0.0005900
RURAL	20.9	1	0.0001	-0.13380	0.0008600
SMALL	14.8	1	0.0005	-0.10004	0.0006800
TEEN	6.8	1	0.0136	-0.05338	0.0004200
MALEOUT	5.4	1	0.0263	0.04778	0.0004200
INTERCEPT				1.82802	0.0042439
Overall Model	160.7	9	0.0000		

Group: White .5-5 yrs.

Variable	F Stat.	DF	Prob.	Coeff.	Variance
MALE	554.3	1	0.0000	0.29635	0.0001600
GAS3	91.9	1	0.0000	0.72462	0.0057151
RURAL	22.4	1	0.0000	-0.13704	0.0008400
SMALL	15.8	1	0.0004	-0.10193	0.0006600
MALEOUT	6.0	1	0.0200	0.05124	0.0004400
INTERCEPT				1.83083	0.0044237
Overall Model	250.3	5	0.0000		

Group: White 6-17 yrs.

Variable	F Stat.	DF	Prob.	Coeff.	Variance
MALE	1151.6	1	0.0000	0.34204	0.0001000
GAS3	105.1	1	0.0000	0.71918	0.0049208
RURAL	16.6	1	0.0003	-0.11581	0.0008100
SMALL	14.1	1	0.0007	-0.09344	0.0006200
INTERCEPT				1.84978	0.0042993
Overall Model	280.2	4	0.0000		

Group: White 18-74 yrs.

Variable	F Stat.	DF	Prob.	Coeff.	Variance
MALE	65.4	1	0.0000	0.15290	0.0003600
GAS3	32.8	1	0.0000	0.76154	0.0177065
RURAL	3.9	1	0.0578	-0.06269	0.0010148
INTERCEPT				1.74183	0.0124679
Overall Model	36.9	3	0.0000		

REFERENCES

1) Comments by E.I. du Pont de Nemours and Co. in response to EPA's proposed rule regarding lead in gasoline. October 7, 1982.

2) Annest J.L., et al., manuscript in preparation

3) Pirkle J.L., Sampling of the Urban Population and the Downward Trend in the NHANES II Blood Lead Levels, August 24, 1982 (EPA docket)

4) Holt M.M., Revised by Shah B.V., SURREGR: Standard Errors of Regression Coefficients from Sample Survey Data. Research Triangle Institute, Research Triangle Park, North Carolina April 1982.

5) BMDP Statistical Software, Stepwise Regression Program P2R, 1981.

6) Neter J, Wasserman W, Applied Linear Statistical Models, p. 382, 1974.

7) Jelinek C.V., "Levels of Lead in the U.S. Food Supply" presented at the Analytical Methodology for Lead in Foods Symposium, Assoc. of Official Analytical Chemists Meeting, Oct. 19, 1981.

8) Dudley M.T., Boone D.J., Blood Lead Analysis 1980, Laboratory Improvement Program Office, Centers for Disease Control.