Journal of Official Statistics
Vol. 5. No. 4, 1989, pp. 421–431

The Relationship Between Gasoline Lead and Blood Lead in the United States

Joel Schwartz and Hugh Pitcher[1]

Abstract: The relationship between gasoline lead and blood lead was examined using a large sample chosen to be representative of the U.S. population. A strong, highly significant relationship was found, indicating that gasoline lead may have accounted for over half of the lead in blood in the late 1970s. The relationship held controlling for socioeconomic status, food intake, alcohol consumption, smoking, and occupational status. Separate regressions for males, females, whites, blacks, urban, and less urban gave comparable results. A separate analysis of blood lead levels in inner-city black children in Chicago, Illinois gave an essentially identical regression coefficient. The relationship appears to be causal.

Key words: Gasoline; blood lead; air pollution.

1. Introduction

Estimating the health benefits of an environmental regulation requires predicting the regulation's effect on human exposure levels. In some cases, only indirect assessments are possible, based on estimates of contributions from different pathways and modeling emissions. Unless information is available on the variance as well as the mean contribution for each pathway, such exercises often predict the distribution of exposure more poorly than the mean.

In the case of lead in gasoline, however, the United States Environmental Protection Agency used several large data sets that allowed direct estimation of the effect of changes in gasoline lead usage on the blood lead distribution in the U.S., including changes in the variance. Analyses of these data have shown a strong and consistent relationship between the amount of lead in gasoline and the amount of lead in blood.

Several articles in the last decade have shown a strong correlation between gasoline lead and blood lead in the U.S. Billick, Curran, and Shier (1979) showed a very strong correlation between gasoline lead and blood lead in the several hundred thousand children in the New York City lead screening program. Annest et al. (1983) analyzed the blood lead levels from 1976 to 1980 in the second National Health and Nutrition Examination Survey (NHANES II), a representative sample of the U.S.

[1] U.S. Environmental Protection Agency, PM 221, 401 M. St. SW, Washington, D.C. 20460, U.S.A.

Acknowledgements: Lee Annest provided valuable support in dealing with the complexities of the NHANES II survey, as well as the NIOSH occupational codes for lead exposure. Asa Janney provided considerable help in earlier analyses of this data. We would also like to thank Robert Murphy of the National Center for Health Statistics for providing the NHANES II data tapes. Some of this work was done while Joel Schwartz was a visiting scientist at the Department of Biostatistics, Harvard School of Public Health.

population. They found a downward trend in blood lead levels after controlling for socioeconomic and regional factors, which persisted in each separate demographic subgroup. The only factor correlated with this trend was gasoline lead. It showed a highly significant correlation after controlling for the other factors and explained the trend in blood lead. No regression analyses were presented, however.

This analysis examines the relationship between gasoline lead and blood lead in the NHANES II survey in greater detail, and uses the results of the screening program in Chicago, Illinois to confirm those results. Changes in both the mean and the variance of the blood lead distribution are examined.

2. Materials and Methods

2.1. Data

The NHANES II sample was selected to represent the civilian, non-institutionalized U.S. population aged 6 months to 74 years. The survey was conducted between February 1976 and February 1980. Medical histories and socioeconomic information were obtained by trained interviewers in home interviews using standardized forms. Dietary questionnaires were also obtained and quantified using a current nutrient databank. Medical examinations were performed, and blood and urine specimens were collected. Blood lead levels were determined by the Centers for Disease Control (CDC) using a modified Delves Cup technique (Barthel et al. 1973), and both blind and bench quality controls were used. Specimens were analyzed in duplicate and the means of the two measurements were used in this analysis. There was no trend in the blind quality control data. Blood lead levels were obtained on 9,996 subjects. Occupational categories were obtained at the home interview, and subjects were coded exposed or non-exposed based on a list of exposed occupations provided by the National Institute for Occupational Safety and Health.

Gasoline lead usage in the U.S. was obtained by multiplying monthly sales of leaded gasoline (obtained from refiners reports to the U.S. Department of Energy) by quarterly concentrations of lead in gasoline (obtained from refiners reports to the U.S. Environmental Protection Agency).

The City of Chicago conducted an extensive blood lead screening program over the 1970s, with approximately 50,000 inner-city children screened each year. A 1 in 30 sample of the over 200,000 blood lead measurements obtained between 1976 and the middle of 1980 was keypunched, along with age, race, and sex. Because few whites were screened we have restricted our analysis to blacks. Blood lead was measured by modified Delves Cup atomic absorption spectrophotometry, using a single city laboratory in the CDC proficiency program. Monthly leaded gasoline sales in the Chicago Standard Metropolitan Statistical Area (SMSA) were obtained from reports of the Ethyl Corporation, and quarterly concentrations of lead in gasoline from the U.S. Environmental Protection Agency.

2.1.1. Analytical Approach

For the NHANES II survey, multiple regression analysis was used to assess the contribution of gasoline lead to blood lead, controlling for age, race, sex, income, degree of urbanization, region of the country, nutrient intake, head of household educational level, cigarette smoking, alcohol consumption, and occupational exposure. The model was

$$E(y_i) = \mathbf{X}_i'\beta$$

where y_i is the blood lead of the ith individual, and $\mathbf{X}_i$ is the vector of covariates for that subject, including gasoline lead, which was the same for all persons examined in the same month. Because the NHANES II survey is a stratified clustered random sample, the covariance may not be diagonal. We have used SURREGR, (Shah (1982)), a survey regression package in the SAS environment, to estimate the design effects for the variables in the regression models. The results shown reflect that adjustment. Partial regression plots were used to determine the need for transformation of variables. Insignificant factors were subsequently backwards eliminated. Analyses were performed on subgroups to test for interactions and to assess the robustness of the relationship. Logistic regression was used to model the percentage of subjects with blood lead levels in excess of 30 μg/dl, which was the CDC's definition of lead toxicity at the time of the survey. These regressions also allowed us to assess the change in the variance of the blood lead distribution, since two points determine a lognormal distribution such as the distribution of blood lead.

Gasoline lead may affect blood lead levels with some delay. We examined this possibility by looking at lags in the relationship. In addition, dummy variables for seasons and a linear time trend were tested.

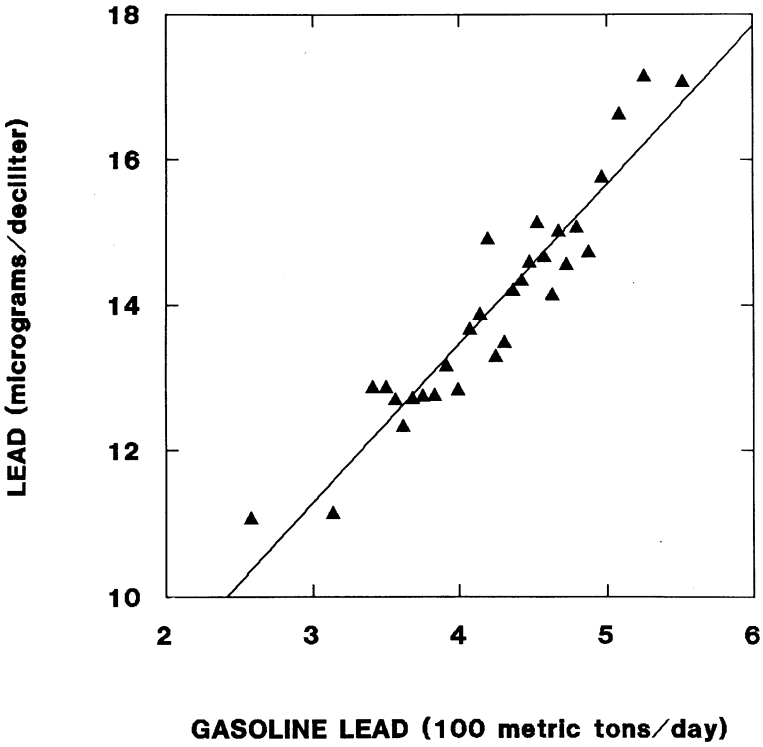

Fig. 1. Each point represents the mean blood lead level and mean gasoline lead usage for 310 consecutive observations, sorted in increasing order of gasoline lead, after adjusting by regression for age, race, sex, income, degree of urbanization, nutrient intake, smoking, alcohol consumption, occupational exposure, and other significant covariates. The regression line is the true regression line for all 9,987 subjects.

The relationship between gasoline lead use and mean blood lead levels in Chicago was also examined using both linear and logistic regressions. Since the Chicago data represents a time series of observations in the same population, an autoregressive model was used. We assumed that

$$y_i = \mathbf{X}_i'\beta + \sum_j \delta_j \{ y_{i-j} - \mathbf{X}_{i-j}'\beta \} + e_i.$$

Again, variables were eliminated by backwards elimination. The degree of autocorrelation was chosen using Akaike's information criteria (1973).

3. Results

Figure 1 shows the relationship between gasoline lead and blood lead in the NHANES II survey, after controlling by regression for age, race, sex, income, degree of urbanization, region of the country, head of household educational level, smoking, alcohol consumption, occupational exposure, dietary factors, and interactions between these variables. Each point represents the mean blood lead and mean gasoline lead for 310 subjects, sorted in increasing order of gasoline lead, and after adjusting for the other covariates (the grand means have been added back in). The regression which showed the strongest association was between blood lead levels and the previous month's gasoline lead use. The results of this regression (for whites) is shown in Table 1. Results

Table 1. Regression results for whites

Variable	Coefficient	Standard error	*p*-value
Intercept	6.15		
Gasoline lead	2.14	0.192	0.0000
Low income	0.79	0.243	0.0025
Moderate income	0.32	0.184	0.0897
Child (under 8)	3.47	0.354	0.0000
Number of cigarettes	0.08	0.001	0.0000
Occupationally exposed	1.74	0.251	0.0000
Vitamin C	−0.04	0.012	0.0010
Teenager	−0.30	0.224	0.1841
Male	0.50	0.436	0.2538
Male teenager	1.67	0.510	0.0026
Male adult	3.40	0.511	0.0000
Small city	−0.91	0.292	0.0039
Rural	−1.29	0.316	0.0003
Phosphorous	−0.001	0.0003	0.0009
Drinker	0.67	0.173	0.0007
Heavy drinker	1.53	0.316	0.0000
Northeast	−1.09	0.332	0.0028
South	−1.44	0.374	0.0005
Midwest	−1.35	0.500	0.0115
Educational level	−0.60	0.141	0.0000
Riboflavin	−1.88	0.071	0.0186
Vitamin A	0.018	0.009	0.0355

The coefficients of the dummy variables show how much blood lead (in μg/dl) is, on average, attributable to a specific effect. The coefficient of gasoline lead shows the number of μg/dl of blood lead attributable to each 100 metric tons per day of gasoline lead use.

Table 2. Mean dietary lead intake (micrograms/day)

Fiscal year	Infants	Toddlers	Males aged 15–20
1976	21	30	71.1
1977	22	28	79.3
1978	25	35	95.1
1979	36	46	81.7
1980	—	—	82.9

for blacks (not shown) were essentially identical.

The NHANES II survey was conducted over a four-year period, during which time blood lead levels and gasoline lead use fell. Several approaches were used to assure that the correlation found was not due to changes in other sources of lead exposure or to changes in the population sample (e.g., black vs white or urban vs rural) that might have created a change in blood lead levels in the survey when none occurred in the population. The first approach was to include all major correlates of blood lead levels and sources of lead exposure in the regression described in Table 1. Separate regressions were then run for males, females, blacks, whites, children, adults, and more and less urban areas. The regression coefficients for gasoline lead varied by less than ± 10 percentage points in these regressions from the coefficient including all subjects. The same covariates were included in each model (except for the covariates the models were stratified on).

Table 2 shows the mean dietary lead intakes for three age ranges obtained from market basket surveys by the U.S. Food and Drug Administration. No significant time trend is apparent, and forcing an (insignificant) term for this trend in overall food lead concentration into the regression did not change the coefficient of gasoline lead. In contrast, some of the individual dietary intake variables were significantly associated with blood lead levels, as shown in Table 1. The association with dietary vitamin C intake possibly reflects the heavy use of lead-based pesticides and growth regulators on citrus fruits in the United States.

The decline in gasoline lead levels during this period was caused by government regulations, which changed in 1978. Gasoline lead use fell much more rapidly after the change. An interaction term for gasoline lead after the regulatory change was insignificant, indicating that the sharp change in gasoline lead decline was mimicked by a similar change in the decline of blood lead, thus resulting in an unchanged slope. This result was already noted in time plots published by the Centers for Disease Control (1982).

Neither the linear time trend term nor the seasonal dummy variables were significant in the regression when gasoline lead was in the model, although they were significant in models without gasoline lead. Gasoline lead remained significant when they were inserted into the model, and the stepwise regression always selected gasoline lead, but not these terms, regardless of what other variables were considered.

Table 3 shows the results of stepwise logistic regression of the probability of a child's blood lead level exceeding 30 μg/dl for black and white children. Gasoline lead was highly predictive of the risk of lead toxicity.

Table 4 shows the results of regressing the quarterly average blood lead levels for black children in Chicago aged 0 to 5 years against quarterly gasoline lead use in the SMSA, for the 18 quarters from 1976 to mid-1980. Again, gasoline lead was an excellent predictor of children's blood lead level ($p < 0.0001$).

Autoregressive terms were not significant or indicated by Akaike's information

Table 3. Logistic regression on probability of blood lead > 30 μg/dl for children 6 months to 7 years

Black children under 8 years old, 479 observations

Model Chi square 39.63 with 5 D.F.

Variable	Beta	Standard error	Chi square	*p*-value
Intercept	−6.9468	1.2656	30.13	0.0000
Gaslead	0.8633	0.2452	12.40	0.0004
Poor	0.9815	0.2803	12.26	0.0005
Age < 2	1.1404	0.6246	3.33	0.0679
Age 2–4	1.1938	0.5696	4.39	0.0361
Age 5–6	0.5428	0.5728	0.90	0.3433

Fraction of concordant pairs of predicted probabilities and responses = 0.718.

White children under 8 years old, 2,225 observations

Model Chi square 33.58 with 5 D.F.

Variable	Beta	Standard error	Chi square	*p*-value
Intercept	−8.1667	1.2322	43.93	0.0000
Gaslead	0.6331	0.2160	8.59	0.0034
Poor	1.2174	0.2935	17.21	0.0000
Age < 2	1.4332	0.7978	3.23	0.0724
Age 2–4	1.7168	0.7415	5.36	0.0206
Age 5–6	1.1405	0.7503	2.31	0.1285

Fraction of concordant pairs of predicted probabilities and responses = 0.637.

Table 4. Regression results: Mean blood lead levels for black children in Chicago[1]

Variable	Coefficient	Standard error	*p*-value
Intercept	12.45	0.68	0.0001
Age 1	4.27	0.48	0.0001
Age 2	5.33	0.50	0.0001
Age 3	4.40	0.52	0.0001
Age 4	4.23	0.55	0.0001
Age 5	3.68	0.62	0.0001
Gaslead	16.12	1.37	0.0001

[1] Blood lead is measured in micrograms/deciliter and gasoline lead in 10^9 grams/calendar quarter.

criteria. Forcing their inclusion in the model did not produce any noticeable change in the gasoline lead coefficient. The gasoline lead coefficient in Table 4 is in billions of grams per quarter in the Chicago metropolitan area. To compare it to the NHANES II coefficient we scaled by the ratio of Chicago's gasoline use to the nation's, and converted to units of 100 metric tons per day (the units used in Table 2). This gave a gasoline lead coefficient of 1.97 for Chicago, quite close to the national results of 2.14. Gasoline lead was also a highly significant predictor of the probability that a child was

found lead toxic, after controlling for age ($\beta = 4.01$, $t = 11.14$, $p < 0.0001$).

Finally, we note that whenever a blood lead level over 30 μg/dl was found in Chicago, the child's home, as well as those of neighbors or relatives where the child played, where checked for the presence of lead paint. The percentage of lead-toxic children with and without lead paint exposure was available for each quarter. In addition, in 1977 all 80,000 children in the screening program had their homes checked for lead paint, providing the overall prevalence of lead paint exposure in the screening population. Using the quarterly probability of lead paint exposure given that a child is lead toxic, the quarterly probability of lead toxicity, and the overall rate of lead paint exposure, Bayes' theorem can be used to compute the probability that a child is lead toxic given no lead paint exposure. Gasoline lead was a highly significant predictor of this probability in a logistic regression ($\beta = 3.71$, $t = 8.62$, $p < 0.0001$).

Figure 2 shows the relationship between gasoline lead and blood lead for black children in Chicago. Each point represents the mean blood lead and mean gasoline lead for each calendar quarter, after adjusting for age (by regression). Figure 3 shows the results of a nonparametric three dimensional smoothing of the mean blood lead level versus age and gasoline lead in Chicago, using distance weighted least squares smoothing, based on an algorithm of McLain (1974). The essentially linear relationship

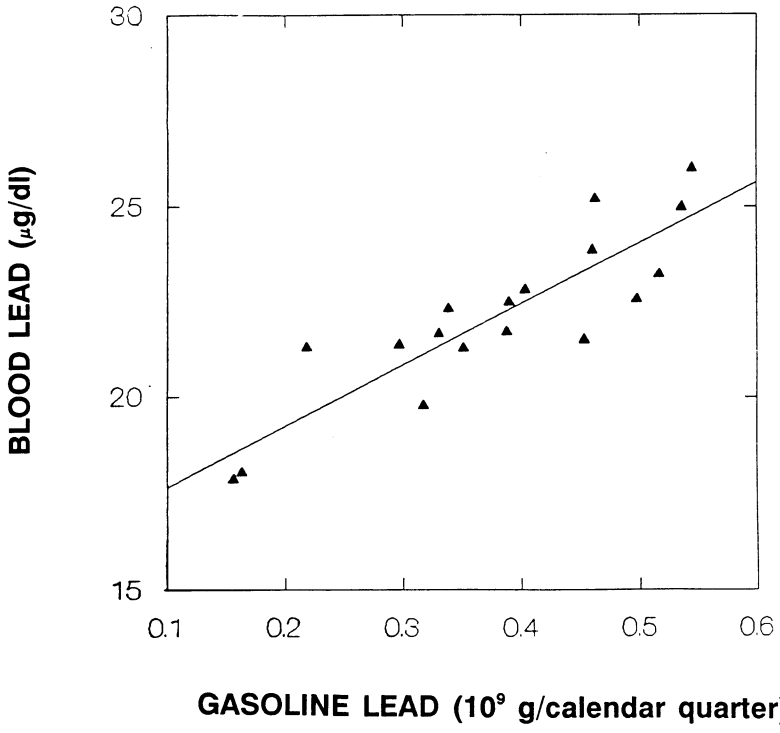

Fig. 2. Each point represents the mean gasoline lead and mean blood lead for each calendar quarter, after adjusting, by regression, for age. The line drawn is the true regression line for the 5,476 individuals in the sample.

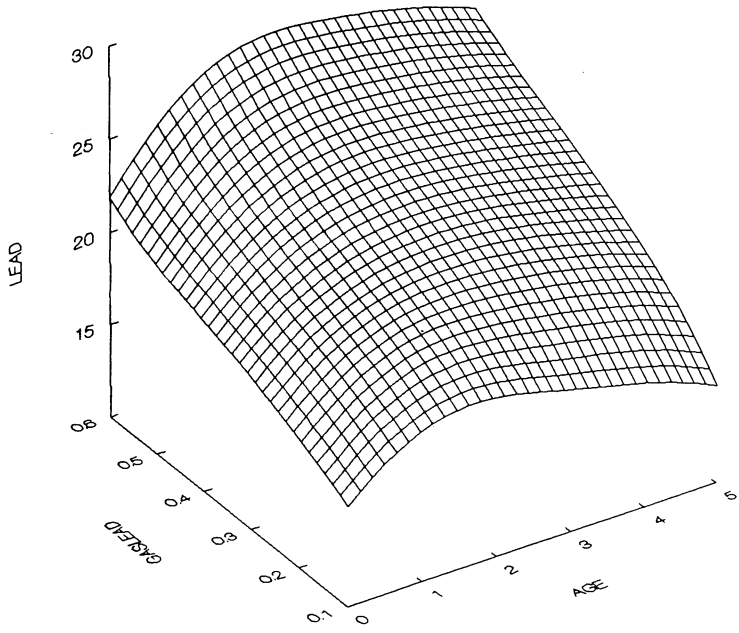

Fig. 3. This surface is the non-parametric smoothed surface of the simultaneous dependence of blood lead on gasoline lead and age

between gasoline lead and blood lead at all ages, as well as the peaking of children's mean blood lead levels between two and three years of age are apparent.

4. Discussion

A strong relationship was noted between gasoline lead and blood lead in the NHANES II data that was robust to specification and subgroup. Analysis in Chicago yielded an almost identical estimate of the increase in mean blood lead levels resulting from an additional 100 metric tons per day of gasoline lead use nationwide. Since mean gasoline lead use during this period was 426 metric tons per day, gasoline lead appears to have contributed over 8 μg/dl to mean blood lead levels during the period. The logistic regressions allowed the estimation for the change in the risk of blood lead levels in excess of 30 μg/dl as gasoline lead changed. Here, the coefficients in Chicago differed more substantially from those in NHANES II, where there was also a noticeable difference in the coefficients between whites and blacks. This reflects the substantial differences in mean blood lead levels between blacks in NHANES II, whites in NHANES II, and blacks in Chicago. The relative risk of exceeding 30 μg/dl as a result of additional exposure that increases the mean blood lead by 1 μg/dl would be expected to differ between groups whose mean initial blood lead levels are 16 μg/dl (whites in NHANES II) and those whose mean initial blood lead levels are 22 μg/dl (blacks in Chicago) unless there is a substantially lower variance in the second group. This was not the case.

What was apparent from comparing the results of the regressions predicting mean blood lead levels and those predicting the percentage over 30 μg/dl was that as mean blood lead levels fell, the geometric standard deviation of the blood lead distribution

appeared to increase. The effect of gasoline lead reduction was greater on the mean than on the tail of the distribution. We believe this result makes sense. A constant geometric standard deviation implies a constant percentage reduction across the range of the blood lead distribution. However, children with high blood lead levels because of excessive exposure to lead in paint or drinking water are unlikely to show larger absolute reductions in blood lead levels than those without such exposure, which is what a constant percentage model would imply.

An indirect test of the validity of the association is provided by the lag structure that was found. The half life of lead in the blood is about 30 days (Rabinowitz, Wetherhill, and Kopple (1976)). Since the average blood lead test occurred in the middle of the month, this lag would be consistent with a stronger association with the previous month's gasoline lead than with the current month's. If, in contrast, gasoline lead was merely acting as a proxy for the passing of time, the association with current month, previous month, and two month's prior gasoline lead should be comparable. In fact, previous months' gasoline lead was the best predictor of blood lead, and the coefficient of two month's prior gasoline lead was half the size of the coefficient of the last month. This is entirely consistent with the 30 day half life.

In addition to Billick's finding of a strong association between gasoline lead and blood lead in New York City, Rabinowitz and Needleman (1983) reported a strong ($p < 0.0001$) association between gasoline lead used in the Boston metropolitan area and umbilical cord lead measurements from 11,000 consecutive births at Boston Women's Hospital between 1979 and 1981. No significant monthly variation was noted in mother's education levels or in her smoking or alcohol consumption history. Water lead levels increased somewhat over the period, while blood lead and gasoline lead fell.

It is important to confirm epidemiologic studies with experimental data when possible, and experimental data do exist to indicate that this statistical relationship is reasonable. Facchetti and Geiss (1982) investigated the contribution of gasoline lead to blood lead in Turin, Italy during the late 1970s by changing the isotopic composition of lead added to gasoline, and monitoring the isotopic composition of blood lead. The isotopic composition of air lead changed rapidly with the change in the isotopic composition of gasoline lead. Changes in the isotopic composition of blood lead occurred more slowly, because of re-equilibration with bone lead stores. Since blood lead isotopic ratios were still changing when the gasoline isotopes were switched back, only a lower bound estimate of the contribution of gasoline lead to blood lead in Turin was possible. Facchetti and Geiss estimated this lower bound at 6 μg/dl blood lead.

Manton (1977) analyzed isotopic changes in blood lead in the United States and found contributions of airborne lead (predominantly from gasoline) of 5 to 10 μg/dl in most of his subjects. Tera, Schwartzman, and Watkins (1985) analyzed isotopic ratios of blood lead in children in Washington, D.C. as the isotopic ratio of air lead changed. Their data showed that as late as 1983 at least 38% of the lead in children's blood still came from gasoline, despite a 50% reduction in lead in gasoline since 1978. These isotope studies confirm the association found in the NHANES II and Chicago data, and yield similar effect size estimates.

4.1. Potential confounding factors

A great deal is known about the sources of

lead exposure and their correlates. The major sources of lead are gasoline, food, water, and paint. There are highly significant associations between exposure and socioeconomic factors, age, smoking, and alcohol consumption. Dietary consumption, smoking, alcohol use, age, socioeconomic factors, and occupational exposure were all included in our model. Water lead levels are determined by the corrosivity of the water and the amount of lead materials in the plumbing. Data from EPA indicates that the pH of the water supplies did not change during the course of the study, and plumbing systems change too slowly to be a source of confounding. In Chicago, there was no change in the pH of the water during the period analyzed.

Lead paint exposure was not available in the NHANES II survey, but is an unlikely source of confounding since the same relationship between blood lead and gasoline lead was found in adults who do not ingest lead paint chips and children who do. Moreover, as noted previously, gasoline lead did not have a disproportionate association with changes in children at high blood lead levels, and there was no significant removal of lead paint from housing during the survey. Lead paint exposure was available for the lead toxic children in Chicago, and as noted above, a strong association was found between lead toxicity and gasoline lead in children with no lead paint exposure in Chicago.

To assure that there was no confounding with any other sources of lead exposure that varied geographically, we inserted dummy variables into the NHANES II regression for each of the 48 counties in the survey that had populations in excess of 100,000. The coefficient of gasoline lead changes by less than 5 percentage points. We conclude that confounding by omitted covariates is unlikely.

5. Conclusion

Gasoline lead was a major source of blood lead in the U.S. population in the 1970s and the removal of lead from gasoline represented a major reduction in exposure to a toxic substance. Removing lead from gasoline has apparently reduced the mean blood lead level of the population by over 50 percentage points, and resulted in important, but not proportional, reductions in children with lead toxicity.

6. References

Akaike, H. (1973): Information Theory and an Extension of the Maximum Likelihood Principle. In B.N. Petrov and F. Csaki, eds. 2nd International Symposium on Information Theory, Budapest, Akademiai Kiado, pp. 267–281.

Annest, J.L., Pirkle, J.L., Makuc, D., Neese, J.D., Bayse, D.D., and Kovar, M.G. (1983): Chronological Trends in Blood Lead Levels Between 1976 and 1980. New England Journal of Medicine, 308, pp. 1373–1377.

Barthel, W.F., Smirek, A.L., Angel, G.P., Liddle, J.A., Landrigan, P.J., Gehlbach, S.H., and Chisolm, J.J. (1973): Modified Delves Cup Atomic Absorption Determination of Lead in Blood. Journal of the Association of Official Analytic Chemists, Vol. 65, No. 5, pp. 1252–1256.

Billick, I.N., Curran, A.S., and Shier, D.R. (1979): Analysis of Perdiatric Blood Lead Levels in New York City for 1970–76. Environmental Health Perspectives, Vol. 31, pp. 183–190.

Centers for Disease Control (1982): Blood Lead Levels in the U.S. Population. Morbidity and Mortality Weekly Reports, Vol. 31, pp. 132–134.

Fachetti, S. and Geiss, F. (1982): Isotopic Lead Experiment: Status Report. Luxem-

bourg: Commission of the European Communities; Publication No. EUR 8352 EN.

Manton, W.I. (1977): Sources of Lead in Blood: Identification by Stable Isotopes. Archives of Environmental Health, Vol. 32, pp. 149–159.

McLain, D.H. (1974): Drawing Contours from Arbitrary Data Points. The Computer Journal, Vol. 17, pp. 318–324.

Rabinowitz, M.S., Wetherhill, G.W., and Kopple, J.D. (1976): Kinetic Analysis of Lead Metabolism in Human Health. Journal of Clinical Investigations, Vol. 58, pp. 260–270.

Rabinowitz, M.S. and Needleman, H. (1983): Petrol Lead Sales and Umbilical Cord Blood Lead Levels in Boston, Mass. Lancet, #8314/5, Vol. 1, p. 63.

Shah, B.V. (1982): Standard Errors of Regression Coefficients from Sample Survey Data. Research Triangle Institute, Research Triangle Park, NC.

Tera, O., Schwartman, D.W., and Watkins, T.R. (1985): Identification of Gasoline Lead in Children's Blood Using Isotope Analysis. Archives of Environmental Health, Vol. 40, No. 2, pp. 120–123.

Received November 1988
Revised July 1989