UNITED STATES ENVIRONMENTAL PROTECTION AGENCY

3/85

THE RELATIONSHIP BETWEEN BLOOD LEAD LEVELS AND BLOOD PRESSURE

Joel Schwartz

Concerns about the health effects of ambient exposure to lead traditionally have focused on children. Although lead has a variety of adverse effects on the health of adults, most of them appear not to be of substantial concern except at very high blood-lead levels. Recently, however, two new and extensive analyses of the NHANES II data set have shown a strong and robust relationship between blood lead levels and blood pressure. That finding has important implications for the benefits of reducing lead in gasoline, because high blood pressure, in turn, is linked to a variety of cardiovascular diseases.

This paper analyzes the statistical relationship between blood lead and blood pressure. The first part provides a brief overview of earlier studies on the subject, while the second part provides a detailed discussion of a recently completed statistical analysis of the NHANES II data.

A.1. Earlier Studies

Lead has long been associated with effects on blood pressure and the cardiovascular system, including a paper in the British Medical Journal by Lorimer in 1886 that found that higher blood-lead levels increased the risk of hypertension. Most of the studies have focused only on hypertension and relatively high lead-exposure levels, and have not looked for a continuous effect of lead on blood pressure. Investigators reporting such an effect include Beevers et al. (1980), Morgan (1976), Richet et al. (1966), and Dingwall-Fordyce and Lane (1963). Others have failed to find effects of lead on hypertension that were significant at the 95 percent confidence level, although most of them did find a positive association. These include Ramirez-Cervantes et al. (1978) and Fouts and Page (1942).

More recently, Batuman et al. (1983) found an association between chelatable body-lead levels and hypertension in veterans, and several recent general population studies and lower lead-exposure studies (Beevers et al., 1976; Kromhout and Coulande, 1984) have found a significant association with blood lead. Moreau et al. (1982) also found a significant relationship ($p < 0.001$) between blood lead levels and a continuous measure of blood pressure in 431 French policemen, after controlling for age, body mass index, smoking, and drinking.

An even more recent British study (Pocock et al., in press) found blood lead significantly related to blood pressure at the 99 percent confidence level, but the authors felt that the small size of their correlation coefficient suggested no noticeable effect. However, that conclusion appears to reflect a misunderstanding of statistics. It is the regression coefficient that indicates the size of an effect. A correlation coefficient confounds that measurement with the variances of the dependent and independent variables. While their full data set was not available to us, Pocock et al. presented their grouped data, and we were able to perform a regression of blood pressure versus the log of blood lead on their group averages, both before and after adjustment for confounders. The regressions were weighted by the inverse of the variance of each group, and confirmed their finding that blood lead was a significant predictor of blood pressure in their data, both before and after adjusting for covariates. Moreover, the regression coefficient indicated that the size of the effect was significant, suggesting a change of 3 mm Hg (millimeters of mercury, the standard measure of blood pressure) as blood lead goes from 5 to 15 ug/dl.

Weeden (1975) found lead associated with the vascular renal changes linked to essential hypertension, indicating a possible causal pathway. Cooper and Gaffey (1974) analyzed mortality data from 1,267 death certificates for 7,032 lead workers employed between 1900 and 1969, and found a significant excess of deaths from hypertension disease and renal disease. A later analysis of similar data from 1971 to 1975 also found an increase in cardiovascular and renal disease, but it was no longer significant at the 95 percent confidence level (Cooper, 1981).

Animal data also link lead to hypertension. Victery (1982) found lead associated with a significant elevation of blood pressure in rats with blood lead levels of 41 ug/dl. Importantly, this study confirms Beevers et al.'s finding of a sex differential, with male but not female rats becoming hypertensive. Webb (1981) examined the vascular responsiveness of tail arteries in rats exposed to blood lead levels in the 40 ug/dl range that had suffered increases in systolic blood pressure, and found that the arteries in exposed rats had increased contraction in response to stimulation by neurotransmitters.

Iannaccone et al. (1981) also reported increased blood pressure ($p < 0.001$) in rats at blood lead levels of 38.4 ug/dl, as well as significant increases in the blood pressure response to noradrenalin. Perry and Erlanger (1979) found that low level exposure of rats to lead produced increases of 15-20 mm Hg in systolic blood pressure. Kopp (1980) repeated those findings and found electrocardiogram changes, indicating an effect on the heart itself. Subsequent tissue analysis of the heart showed reduced levels of ATP in the heart muscle, indicating that heme synthesis inhibition by lead was affecting energy availability in the heart itself.

The direct cardiological effects of lead are also indicated by electrocardiogram changes in lead-poisoned children, which are reversed by chelation therapy (Freeman, 1965; Silver and Rodrigues-Torres, 1968). Williams (1977, 1978, 1979) has shown persistant increased susceptibility to norepinephrine-induced arrhythmias in rats exposed to lead in the first three weeks of life.

A.2. Analysis of NHANES II Data

In light of these indications of potential effects of lead on the cardiovascular system, the relationship between blood lead levels and blood pressure has recently been explored (Harlan et al., 1985; Pirkle et al., 1985) using the NHANES II data. The NHANES II is an excellent data base for this analysis because of the care given to accurate measurements, the great range of information on possible confounding factors, and because it is a representative sample of the U.S. population. As such it avoids the problems of selection bias, healthy-worker effect, other occupational exposures, and the choice of controls that confound many occupational studies. Harlan et al. found blood lead related to blood pressure for males aged 12 to 74 after controlling for the traditional variables associated with blood pressure (age, age-squared, body mass index, race) as well as alcohol consumption, socio-economic factors, and all nutritional variables suspected of affecting blood pressure. Moreover, this relationship held in each year of the NHANES II sample, when analyzed separately, and this relationship held for both blacks and whites.

Pirkle et al. found that blood lead levels were a statistically significant predictor of blood pressure in adult males. This relationship held not only when blood lead was evaluated in a regression with all known factors that have previously been established as correlated with blood pressure, but also when that relationship subsequently was tested against 87 additional variables representing linear and nonlinear functions of every dietary and serologic variable in the NHANES II survey.

A.2.a. Blood Pressure Measurements

Three blood pressure measurements were taken during NHANES II. A seated measurement was taken as soon as the examinee entered. Later, a recumbent measurement was taken. A second seated measurement was taken just before the end of the examination. It is standard medical practice to prefer the second seated measurement, because nervousness on just entering a medical examination center makes the first seated measurement less stable. All of the results presented are for the second seated measurement. However, almost all of the regressions and robustness tests described were performed on all three measurements, and on the average of the first and third seated measurements; all of the conclusions concerning lead's significance held for all eight regressions (four diastolic, four systolic).

A.2.b. Initial Analysis

After replicating the Harlan et al. results for all adult males, the first goal was to determine if blood lead levels were related significantly to blood pressure in white males, 40 to 59 years old. This subgroup was chosen because at lower ages both blood pressure and blood lead vary with age. This collinearity could artificially mask or enhance the correlation between blood lead and blood pressure. Between 40 and 59 years of age, however, blood pressure is essentially independent of age. Choosing this subgroup avoids any collinearity problems. We focused on whites because data relating cardiovascular disease to blood pressure are less extensive for nonwhites. The established correlates of blood pressure are age, sex, race, and one of the indices of relative height-to-weight. Body mass index (BMI = weight/height^2) was used in this analysis. By limiting attention to 40 to 59 year old white males, there was no need to control for race, sex, or, to a large degree, age. Although age was only occasionally significant in the stepwise analysis, both age and age-squared were forced into each multiple regression model to be certain any effect of lead was independent of age.

The natural log of blood lead was more normally distributed, was more statistically significant, and gave a higher R^2 than untransformed blood lead, blood-lead-squared, blood lead plus blood-lead-squared, the square root of blood lead, or blood lead to other fractional powers (0.15, 0.2, 0.3, 0.4). All of the results reported here are for the natural log of blood lead, but regressions using lead on the untransformed scale gave very similar results.

The initial regressions analyzed systolic and diastolic blood pressures for white males, 40 to 59 years old, with a model consisting of age, age-squared, BMI, and blood lead. These regressions were done to determine whether blood lead levels were significantly associated with systolic and diastolic blood pressures after controlling for age, sex, race, and BMI, which are well-documented correlates of blood pressure. Lead was statistically significant ($p < 0.01$) for both systolic and diastolic blood pressures in all the regressions (unweighted, weighted, and weighted with design effects). The regressions also tested whether this relationship held up when other potentially confounding variables were considered.

A.3. Tests of Robustness

The regression models were expanded to incorporate additional variables, with particular attention directed to the stability and significance of the lead coefficient in the presence of nutritional factors and blood biochemistries.

A large set of nutritional and biochemical variables from NHANES II was included in the stepwise regressions. Additional regression analyses considered potential problems of interaction terms. To ensure the robustness of the relationships, further analyses were done to address marginally insignificant variables and nonnutrition variables. Although our analysis focused on males aged 40 to 59, additional regressions were also performed considering all males over age 20.

A.3.a. Nutritional and Biochemical Variables

To provide an unusually rigorous test of the independent significance of blood lead, almost all of the nutritional and biochemical variables in the NHANES II were included in stepwise regressions. In addition, to account for possible curvilinear relationships, squared and natural logarithmic transformations of almost all of these variables were also included. The variables are listed in Table 1. The objective was not to evaluate the possible association of nutritional or biochemical measurements with blood pressure, but rather to conservatively estimate the strength and independence of the relationship between blood pressure and blood lead.

Including these additional 87 variables increases the probability of variables being found statistically significant due to chance alone. This complicates the interpretation of nutritional and biochemical factors, but not the interpretation of the lead variable; it only makes it more difficult for lead to maintain its significance.

The general procedure for variable selection was as follows. First, weighted stepwise multiple linear regression was used to determine which variables were significantly related ($p < 0.05$) to blood pressure (using the Stepwise and MAXR options of the SAS procedure, STEPWISE). The MAXR procedure was the principle one used; it determines for any given model size (i.e., number of variables) the variables that explain the greatest amount of the variance (i.e., maximize R^2). We chose the largest model with all variables significantly related to blood pressure ($p < 0.05$). The Stepwise option, which uses forward selection with backwards elimination, chose very similar models, and also always chose blood lead. From the 87 nutritional and biochemical variables, the weighted stepwise regression selected five additional variables for diastolic pressure and six additional variables for systolic pressure using a 5 percent significance test. These were used

TABLE 1. Variables Included in the Stepwise Regression Analyses

age *	dietary iron †
age-squared *	dietary vitamin A †
body mass index	dietary vitamin C †
dietary sodium †	dietary thiamine †
salt shaker sodium	dietary riboflavin †
dietary sodium X salt shaker sodium	dietary niacin †
	serum cholesterol+
dietary potassium †	serum vitamin C †
dietary sodium - potassium ratio	serum iron †
	serum transferrin saturation
dietary calcium †	serum zinc †
dietary phosphorus †	serum copper †
dietary protein †	serum albumin †
dietary fat †	hemoglobin †
dietary carbohydrate †	red blood cell count
dietary cholesterol †	ethanol consumption / week †
dietary saturated fatty acids †	cigarettes smoked / day
	total dietary grams †
dietary oleic acid †	total dietary calories †
dietary linoleic acid †	cigar or pipe smoking

* forced into each regression to remove any possible age effects on blood pressure.

† the natural log and squared transformation of these variables were also included in the stepwise regression.

as the starting model for the SAS procedure SURREGR, which additionally incorporated the survey design effects. For both systolic and diastolic blood pressures, one variable from the weighted stepwise regression failed to maintain significance at the 5 percent level after the design effects were incorporated. The final regression results for systolic and diastolic pressures, after accounting for the weighting and design effects, are given in Table 2.

The multiple logistic regressions (of the probability of hypertension) were also performed using programs from SAS. The procedure LOGIST was used for the stepwise unweighted multiple logistic regression, and the procedure NLIN (nonlinear regression) was used for the weighted logistic regression calculations. The selection process again chose the largest significant model that explained the greatest amount of the variance. Calculations of threshold levels for effects were made using the procedure NLIN on segmented regression models, which finds the threshold point that minimizes the sum of the squares of the error terms. The results of the logistic regression on hypertension are shown in Table 3. Note that the logistic regressions included blacks as well as whites, because these regressions were used only to predict the effect of lead on the probability of having hypertension and were not used to estimate the number of cardiovascular diseases and deaths. As noted earlier, blacks were not included in the linear regressions because the best available coefficient for predicting cardiovascular risks included insufficient numbers of blacks.

TABLE 2. Regression of Diastolic and Systolic Blood Pressures in White Males Aged 40 to 59

Variable	Coefficient	t-Statistic	Probability
Diastolic			
Age	0.2768	0.17	0.8636
Age^2	-0.0014	0.10	0.9321
Body Mass Index	1.131	8.55	0.0001
Log(blood lead)	3.954	2.85	0.0080
Dietary Potassium	-0.0018	4.92	0.0001
Hemoglobin	1.548	3.90	0.0005
Albumin	3.587	2.50	0.0179
Log(dietary vitamin C)	1.838	4.65	0.0001
Systolic			
Age	1.311	0.57	0.5720
Age^2	-0.0068	0.30	0.7706
Body Mass Index	1.736	9.42	0.0001
Log(blood lead)	8.436	3.24	0.0028
Albumin	7.088	2.50	0.0178
Log(dietary Vitamin C)	2.411	3.84	0.0005
Log (dietary riboflavin)	-5.509	3.07	0.0044
Log(dietary oleic acid)	3.992	2.49	0.0183
Log(serum vitamin C)	-3.472	2.47	0.0184

TABLE 3. Weighted Logistic Regression on Probability of Diastolic Blood Pressure Greater Than or Equal to 90 mm Hg in Men Aged 40 to 59

Variable	Coefficient	t-statistic	p-Value
Constant	-16.41	10.13	0.0000
Log(Blood Lead)	0.693	3.96	0.0000
Albumin	0.0873	3.70	0.0001
Body Mass Index	1.700	9.34	0.0000
Hemoglobin	0.0329	5.25	0.0000
Log(Vitamin C)	0.3585	5.98	0.0000
Dietary Potassium	-0.00058	7.47	0.0000
Total Carbohydrates	0.00246	3.09	0.0010

After including the nutritional variables, the blood analytes, and their curvilinear transformations, lead remained significantly associated ($p < 0.01$) with both systolic and diastolic blood pressures. The magnitude of this relationship, adjusted for the other significant variables, is shown graphically in Figures 1 and 2. Furthermore, segmented regression analyses indicated there was no threshold blood lead level in the data.

These segmented "hockey stick" regressions fit two regression lines to the data. One, below the putative blood lead threshold T, depends on all the variables except lead. The other, for blood lead levels above T, includes lead. An iterative technique is used to find the value of T that minimizes the sum of the squares of the error terms over the full range of both regression lines. In this case, the error in the regression was minimized at a threshold of zero; that is, lead was significantly related to blood pressure at all levels down to zero.

A.3.b. <u>Interaction Terms</u>

In multiple regression analysis, another consideration is the possibility of significant interaction terms. To evaluate this possibility, an additional weighted stepwise regression analysis was done for systolic and diastolic blood pressures. The variables consisted of the linear interaction terms between the final variables in the model (shown in Table 2) and the linear form of all the other variables originally selected for the initial stepwise regression, including their log and square transforms (Table 1). This meant running a stepwise regression with 162 interaction terms added to the final regression models for systolic and diastolic pressures. Using such a large set of variables gave a high probability that some variables would enter at the 5 percent level by chance. However, the purpose was not to determine if those variables were independently significant, but, rather, to further test the significance and independence of the relationship between blood pressure and blood lead. As expected, several interaction variables entered the systolic and diastolic regressions, but in each regression the lead coefficient varied less than 10 percent and remained significant ($p < 0.015$).

A.3.c. <u>Marginally Insignificant Variables</u>

Three other analyses were done to ensure that this relationship was robust. First, the original weighted stepwise regression was extended to include variables significant through the 15 percent level to see if marginally insignificant variables influenced the significance of lead. For both systolic and diastolic pressures, lead remained significant and there was little change in the magnitude of the coefficient.

Second, for diastolic blood pressure, all the variables were included that were significant between the $p = 0.05$ and the 0.15 levels, and every possible combination of those variables was considered. All 255 combinations were added to the variables that were statistically significant, and a regression was performed on each one. The coefficient of the log of blood lead varied by only plus or minus 10 percent from the value we obtained when we included only significant variables, and the highest p-value for lead was still less than 0.01.

The last analysis was the most demanding test of the independence of the relationship between blood pressure and blood lead. Models for diastolic and systolic blood pressures were fit by weighted stepwise regression to the original model variables (Table 1), <u>excluding</u> lead. This gave all of the other variables and their curvilinear transformations the maximum opportunity to explain variation that could also be explained by lead. After obtaining this new final model without lead, a single regression was run adding the lead variable to the variables of this new final model. For both systolic and diastolic pressures, lead was still statistically significant ($p < 0.016$) and the magnitude of the lead coefficient changed less than 10 percent from those obtained in the original analysis. The results of all these analyses indicated that the strength and independence of the relationship between blood pressure and blood lead were remarkably stable.

Because some people have found small amounts of ethanol associated with reduced blood pressure, ethanol was also modeled as a quadratic function of consumption, and with two dummy variables for light and heavy drinking. The stepwise regression was repeated, with no change.

A.3.d. <u>Nonnutrition Variables</u>

Pirkle et al. then considered nonnutrition variables that might be associated with blood pressure. In additional runs completed since then, we have added several other variables. The complete set is shown in Table 4. Socio-economic and demographic factors as well as additional medical history variables were included.

TABLE 4. Nonnutrition Variables Tested in the Stepwise Regression

Demographic Variables	Other Personal-History Variables
Family Income	Tricep Skinfold
Poverty Index	Subscapular Skinfold
Region of the Country	Recreational Exercise
Season of the Year	Work-Related Exercise
Degree of Urbanization	Recent Weight Loss
Residence Inside Central City	Family History of Hypertension
Educational Level	Kidney Disease
	Serum Creatinine
Hypertension Variables	
Hypertensive Medication	
Low Salt Diet	

Hypertension medication and low salt diet were tested -- not for inclusion in a final model, as they are essentially indicators of high blood pressure, but rather to see if the response to lead differed in those groups. The coefficient of lead did not change appreciably, and lead interaction terms with the two variables were insignificant.

The other variables in Table 4 were tested in two ways. First, the stepwise regression procedure was repeated with them using all nutritional and serum measurements that were significant at the p = 0.15 level. The nutritional factors were limited to those significant at the 0.15 level to give the nonnutritional factors a greater chance to enter the model. Again, lead was selected ($p < 0.005$) and its coefficient changed by less than 10 percent from the original model that included only age, age^2, and body mass index.

The variables in Table 4 were then added to <u>all</u> of those on Table V-1 (including their nonlinear transforms) and the stepwise process was repeated -- with the same results. Finally, the stepwise procedure was rerun using all of the variables in Tables 1 and 4 <u>except</u> lead; lead was then inserted into the model resulting from this procedure. It was still significant ($p < 0.006$), with less than a 10 percent change in its coefficient. Because the presence of two terms to describe the curvilinear dependence of blood pressure on age might reduce the chances of variable correlated with age achieving significance, age was modeled as a single curvilinear function (sine of age), and the stepwise regression repeated; the results were the same. In addition, smoking and drinking were forced into the regression, and lead was still significant ($p < 0.01$), with only a 3 percent change in its coefficient.

Our previous studies have shown that about half of the lead in people in the NHANES II sample came from gasoline. Tetraethyl lead has very little cadmium in it, so confounding with cadmium (which is also suspected of affecting blood pressure) is unlikely. However, we repeated the regression excluding occupationally exposed workers, who may also have cadmium exposure. Lead remained significant ($p < 0.01$), and its coefficient increased somewhat. We also regressed gasoline lead directly on blood pressure, and it was significant.

Although all of these analyses make it clear that collinearity is not a problem in these regressions, variance inflation factors were computed; no significant variable had a variance inflation factor above 1.4. (Variance inflation factors below 4 are considered acceptable in multiple regression analyses.)

To ensure that the significance of lead in the regression was not due to the presence of a few influential observations, influence diagnostic procedures were run. Studentized residuals were plotted for all the observations, and the largest residuals were clustered near the middle of the data, where their influence is slight. Cook's D statistics also were computed for each observation. The highest Cook's D was 0.029, and the second highest was 0.023, both of which are very small. Moreover, of the 10 observations with the largest Cook's D statistics, six had positive residuals and four had negative residuals, indicating that the most influential observations split almost evenly on which way they would influence the lead regression coefficient.

A.3.e. Other Age Groups

The 40 to 59 year old age group represents about one-third of adult males, and is the only one where the confounding of age and blood lead can be eliminated unambiguously. Additional regressions were performed, however, to confirm the Harlan et al. finding of an effect in all adult males. Tables 1 and 4 contain several variables that Harlan et al. did not consider in their analysis. Therefore, the stepwise regression analysis was repeated using all of the variables in both tables, and their square and natural log transforms as indicated. All males over the age of 20 were considered. Lead was selected by the regression, with a p-value less than 0.01. To check whether the relationship might be substantially different for different age groups, dummy variables for each 10-year age group (between 20 and 70 years), and interaction terms between lead and those dummy variables, were inserted in the stepwise regression. Such interaction terms check for differences in the lead/blood pressure relationship without having to subdivide the sample. None of the interaction terms was significant at even the p = 0.15 level.

A.4. Summary of Blood Lead - Blood Pressure Results

The final models for blood pressure, including all statistically significant variables, are shown in Table 5. The final logistic model for the probability of hypertension is shown in Table 6.

TABLE 5. Regression of Diastolic and Systolic Blood Pressures in White Males Aged 40 to 59

Variable	Coefficient	F-Statistic	Probability
Diastolic			
Age	-0.210	0.02	0.8960
Age-squared	0.003	0.04	0.8373
Body Mass Index	1.082	67.88	0.0000
Blood lead†	$\boxed{4.609}$	12.19	0.0014
Potassium	-0.002	25.30	0.0000
Hemoglobin	0.151	16.81	0.0003
Albumin	0.354	7.42	0.0104
Dietary Vitamin C†	1.886	23.67	0.0000
Family history of hypertension	2.085	4.37	0.0446
Recreational exercise	-1.851	9.48	0.0042
Systolic			
Age	1.142	0.25	0.6226
Age-squared	-0.005	0.05	0.8208
Body Mass Index	1.710	85.90	0.0000
Blood lead†	$\boxed{8.510}$	10.54	0.0027
Albumin	0.695	6.09	0.0192
Dietary Vitamin C†	2.458	13.78	0.0008
Dietary Riboflavin†	-5.101	8.14	0.0075
Dietary Oleic Acid†	3.650	5.34	0.0275
Serum Vitamin C†	3.365	5.81	0.0218
Family history of hypertension	3.653	4.59	0.0399

† log transform

TEH 0412826

DUP050453882

TABLE 6. Logistic Regression on Probablity of Blood Pressure Greater Than or Equal to 90 mm Hg in Men Aged 40 to 59

Variable	Coefficient	t-Statistic	p-Value
Constant	-15.40	7.0	0.0000
Log(Blood Lead)	0.793	3.20	0.0014
Albumin	0.650	2.06	0.0399
Body Mass Index	0.1571	6.57	0.0000
Hemoglobin	0.0265	3.19	0.0015
Log(Vitamin C)	0.3593	4.22	0.0000
Dietary Potassium	-0.00053	5.33	0.0000
Total Carbohydrates	0.00286	2.86	0.0080
Recreational Exercise	0.3864	0.128	0.0026

It is noteworthy that the logarithmic form of the dose-response relationship suggests a large initial effect, leveling off at higher blood-lead levels. This may explain why only about 60 percent of the occupational studies (i.e., high lead-exposure studies) have found an effect that was significant at the 95 percent confidence level, while almost all of the studies of lower lead levels have found the relationship to be significant.

The other low-exposure studies, the animal data, and the robustness of these results suggest that the relationship is causal. Moreover, specific analyses to determine whether there is a lower threshold below which lead has no effect on blood pressure showed that the data were fit best with a threshold of zero.