ERC

Energy and Resource Consultants, Inc.

P.O. Drawer O, Boulder, CO 80306 • (303) 449-5515

March 18, 1983

Mr. Dave Weil
Environmental Criteria Assessment Office
MD-52
U.S. Environmental Protection Agency
Environmental Resource Center Annex
Beaunit Building
Research Triangle Park, NC 27711

Dear Mr. Weil:

We are sending the enclosure at the request of Joel Schwartz. It is an extension of an earlier review of the ICF and Ethyl analyses by Dan Violette and myself. The various reports and the comments by Bradley and Draper depict a lively debate, and it has been an interesting exercise trying to separate the fact from fiction and the relevant from the spurious.

We hope that our comments are of some use to you in your decision making.

Sincerely yours,

Craig Miller, Ph.D.
Principal, Systems Engineering

gwb
Enclosure

E R C Energy and Resource Consultants, Inc.
P.O. Drawer O, Boulder, CO 80306 • (303) 449-5515

COMMENTS ON STUDIES USING THE NHANES II DATA TO RELATE HUMAN BLOOD LEAD LEVELS TO LEAD USE AS A GASOLINE ADDITIVE

Prepared for:

Environmental Protection Agency
Office of Policy Analysis

March 1983

Contact:

Dr. Craig Miller
Dr. Daniel Violette

LEADE-Mem2 (104)

MEMORANDUM

Comments on Studies Using the NHANES II Data

To Relate Human Blood Lead Levels to Lead Use

as a Gasoline Additive

Date: **March 18, 1983**

To: **Joel Schwartz, Office of Policy Analysis, U.S. EPA**

From: **Dr. Craig Miller and Dr. Daniel Violette**

In an earlier communication (8), we reviewed the use of the NHANES II data to examine the relationship between lead use in gasoline and human blood lead levels by the U.S. EPA Office of Policy and Resource Management (OPRM) and ICF, Inc. This memo is intended as an extension and enhancement of that review, incorporating some additional comments on Ethyl Corporation's analysis and review of the ICF/OPRM work and a brief review of some work currently underway by ICF/OPRM.

The Ethyl Analysis

In this communication, we will focus on what we feel are the major shortcomings of the Ethyl report and note a few detail errors which have not previously been raised. Ethyl Corporation's analysis of the NHANES II data and review of the October 1982 ICF study (2) was critiqued by ICF (3). That review was reviewed in turn by Bradley (6) and Draper (5). To continue the debate presented in this series of documents would be relatively fruitless; much of what was said by all parties is correct and most of the controversies were due to misunderstandings of the methods used.

Ethyl's analysis of blood levels consisted of a two-stage regression. The rationale behind this bifurcation was to segregate factors specific to the individual in the first stage and factors specific to the individuals' environment in the second stage. Included in the first stage area 35 "personal identifiers" representing combinations of age, income, and race, and 64* dummy variables representing each of the NHANES II sites. A proxy for exposure to lead from gasoline was used in the second stage.

As noted in our earlier communication (8, p. 7) the two-stage approach attributes any common covariation between the second-stage gasoline lead variables and the independent variables in the first stage to the first stage variables. Upon reviewing this comment, we feel it is necessary to make this point more strongly. The coefficients on the lead proxy variable will be biased downward if any one of the first stage variables is correlated with the second stage proxy variable for lead in gasoline. Further, this bias may be severe even if the simple correlation between any one variable in the first stage and the lead proxy variables are small. Instead, if there is a high degree of multiple-correlation between the first stage variables as a set and the second stage lead proxy, the second stage coefficients will be biased. Given the very large number of first stage variables, this is not an unlikely condition. Subsequent work by ICF/OPRM has examined this problem. The preliminary results are discussed in the next section.

A more fundamental problem with the Ethyl model specification is that they did not take advantage of all the available information. Ideally, ICF and Ethyl would like to have

---

* Ethyl states on (2, p. A-1): "The 64 locations were assigned numbers 1 to 64." Some readers of the Ethyl data have taken this to mean that the value inserted of each of the variables representing a location was a number other than 0 or 1. Though we have not yet been able to verify precisely what was done, we assume that Ethyl followed correct procedures and used a 0-1 dummy variable to represent each site. The fact that neither Draper nor Bradley objected to the Ethyl procedure strongly indicates that no error was made in this regard. However, if they did not use 0-1 dummies but instead entered nubmers 1 through 64 as a variable representing the site, they would have entered a meaningless variable with the potential for biasing the estimates. This was discussed in our March 4, 1983 communication.

known the actual levels of environmental exposure to lead from gasoline experienced by the NHANES II individuals. Unfortunately, these data were not available. In lieu of the ideal, ICF and Ethyl had to derive a proxy for exposure. ICF used the lead content of gasoline on a national basis.* One problem with this measure is that it does not capture the variation in exposure between areas. This could be due to differences in traffic volume and other factors that would affect the exposure of a local population to lead. Ethyl attempted to address this problem by using a proxy for exposure which assumes that traffic density (and, hence, the concentration of atmospheric lead from gasoline) is proportional to population density. Ethyl did not test this assumption, but it would seem to offer one approach to dealing with the lack of spatial resolution inherent in ICF's formulation. However, Ethyl's use of population density as a proxy for lead exposure for each site, in contrast to the ICF proxy, does not reflect the fact that the lead content of gasoline changed substantially during the NHANES II sampling period.

It would seem that the Ethyl analysis could have been improved by calculating the lead exposure proxy for each individual rather than for each site. This could be done by dividing the lead in gasoline variable by population or population density, although the latter variable would be in units of exposure per person per square mile. Ethyl was correctly concerned with exposure to lead in food and the possible contribution of lead in gasoline to exposure by this mode. They postulated that the "contribution of lead to the diet can be represented by the gasoline lead use per square mile for each year." (2, p. 22) Gasoline usage per square mile was estimated by dividing gasoline lead usage for a state by the area of the state. Again, it appears that no attempt was made to determine if the assumptions inherent in the use of this proxy are valid. In any case, Ethyl should have

* ICF had data on lead in gasoline for 135 SMSAs, however, these data was based on very small samples within SMSA (e.g. 10 service stations) and approximately half the NHANES II sample sites were not among the 135 SMSAs. For these reasons, ICF used the national data to estimate lead in gasoline.

taken advantage of published information concerning the lead content of food, which shows an increase during the period of time during which blood lead levels were decreasing (2, p. 18), indicating the dietary lead could not have contributed to the observed direction of effect.

A final criticism of the Ethyl report is, perhaps, a fine point. A statement is made in the opening critique of the CDC analysis (2, p. 7 ff), which one might expect would be supported by a more extensive discussion later in the text. The support was not forthcoming, and, as the statement rests on several assumptions, it should not have been made without establishing that those conditions necessary for the statement to be true do, in fact, hold. the statement in question is as follows:

> "Another aspect of the NHANES II data, which argues against the decline in blood lead being caused by reduced lead in gasoline, is that the percent decline in blood-lead levels in the rural population over the four-year period was 33%, which was almost twice the percent decline in the urbanized population of greater than 1,000,000 persons. But the decline in air-lead levels over the period of NHANES II, of course, was much greater in urban areas that in rural areas. Thus, attributing the decline in blood-lead levels to the decline in gasoline lead usage is not proven by the simplistic approach of CDC and EPA." (2, p. 10)

Implicit in this statement are two assumptions: (1) that the relationship between exposure to blood lead levels and exposure to lead is linear and (2) that the same relationship hold for the urban and rural populations. The linearity assumption means that a reduction in exposure of, say, 20% would lead to twice the reduction in blood lead levels or a reduction of 10%. ICF, in their analysis, did not assert that such a linear relationship exists. Urban areas have much higher air-lead levels than rural areas, and have seen a correspondingly higher decline in concentrations over the past decade in absolute terms. However, because the starting points were so radically different, it is not necessarily the case that one would expect the percentage reduction in blood lead levels to be greater in urban areas. To establish whether this should be the case, it would be necessary to derive a dose-response curve for airborne lead. ICF wisely concentrated first on the simpler problem of establishing that there is a relationship and did not present a dose-response curve.

The second assumption implicit in the Ethyl statement is that rural and urban populations respond identically to exposure to lead. It is generally recognized, and has been demonstrated in both the ICF and Ethyl studies, that race is an important determinant of blood lead levels and in the ICF analysis of the relationship between lead in the environment and blood lead levels. Thus, the assumption that urban and rural populations would respond similarly to exposure to lead requires that the racial mix be the same in the two groups. This is not the case.

Additional Work by ICF and the Office of Policy and Resource Management

One comment we made on the December ICF/OPRM work was that the inclusion of dummy variables could improve the analysis by at least partially controlling for variables specific to each site but omitted from the equation. In recent and ongoing work, ICF/OPRM have performed this analysis. The results of the new analysis have not yet been presented and more work is being done, so it is not appropriate to discuss the results at length in this memo. A brief description would be useful, however, as the preliminary results relate specifically to one of the criticisms we offered of the Ethyl analysis -- that there is a correlation between the Ethyl first-stage variables and second-stage variables.

ICF/OPRM performed two two-stage regressions. The dependent variable in the first stage was blood lead levels. The dependent variable in the second stage was the set of residuals from the first stage regression. The independent variables in the two regressions were as follows:

Independent Variables Regression Analysis 1	Independent Variables Regression Analysis 2
First Stage	**First Stage**
Teen	Teen
Kid	Kid
Male	Male
Adult Male	Adult Male
Rural	Rural
Small	Small
Income 1	Income 1
Income 2	Income 2
64 0-1 variables for sites	National lead in gas
Second Stage	**Second Stage**
National lead in gas	64 0-1 variables for sites

Teen and Kid are 0-1 variables to model the age of the person in the sample. Small and Rural are 0-1 variables, denoting the size of the sample community. Small represents urban areas with an SMSA population under 1 million. Rural represents areas defined as rural in the Census. Income 1 and Income 2 are 0-1 variables representing an annual income in the ranges 0 - $6,000 and $6,000 - $15,000, respectively. National lead in gas is the national average concentration of lead in gasoline in grams per gallon.

In the Ethyl model, each individual is an observation in the first stage, but in the second stage it is necessary to group all of the individuals within a site because of the problem of collinearity. The remaining dependent variables have precisely the same value for each individual within a site.* In contrast, the ICF/OPRM analyses do not require this type of aggregation, since different individuals have different values for the dependent variables due to the inclusion of individual data.

* Recall that in the Ethyl analysis, all personal variables are taken out in the first stage and the second stage uses only site variables which will be the same for all individuals sampled at that site.

The new regression runs were run for whites only. In the first run, the coefficient of the lead variable had a value of 1.95* The T ratio was 26.1635, but this was calculated without compensation for the unusual structure of the NHANES II data set, so it may be high by as much as a factor of 2. Regardless, the coefficient is still significant.

In the second run, with gasoline lead in the second stage, the coefficient dropped to 0.971 with a T ratio of 13.4499. This substantial drop on the lead coefficient, with the reversal of the order of the two stages, is a clear substantiation of the hypothesis we presented earlier that there may be some multiple correlation between the Ethyl first stage variables and gasoline lead levels. Calculation of the multiple correlation coefficient using the Ethyl data would show this more unambiguously.

Causality

Ethyl Corporation's report (2) attacks the relationship between lead use in gasoline and human blood levels reported by ICF (1) by attempting to establish that factors other than lead in gas may account for the demonstrated decrease in blood lead levels seen in the NHANES II data. To succeed in this attempt would tend to cast doubt on the validity of ICF's inference that there is a causal relationship between gasoline lead and blood lead levels. However, just as showing that there is a correlation between two factors does not prove that one causes the other, proving that one factor (in this case, lead in gas) is not uniquely correlated with another (in this case, blood levels) does not show that there is no causal relationship. Such a demonstration does show something of the complexity of the issue and clearly indicates that great care should be taken that the dependent variable

* The units of the coefficients are: $\frac{\text{ug of lead/dl of blood}}{\text{g of lead/gallon of gas}}$

under study is related to blood levels by a physically and scientifically reasonable causal mechanism.

The Ethyl report succeeds in this regard. Despite the problems with the study, it does show that other factors may explain some part of the decrease in blood lead levels during the period of the NHANES II survey. This serves to focus the study where it belongs — on determining whether the statistical relationship between lead in gas and human lead levels, which has been demonstrated clearly in the ICF/OPRM work, is supported by a clear and logical explanation of how lead in gasoline can cause higher blood lead levels.

Ralph A. Bradley, in a review of the ICF work (6) for Ethyl Corporation, makes the following statement:

> "At best, regression analysis may suggest a relationship for further investigation, perhaps through an experiment designed to confirm an hypothesized causal effect if it exists. If similar time trends exist for blood lead, lead use in gasoline, average weight of new automobiles sold, and the rate of inflation, there is no more justification for suggesting a causal effect for one of these variables than another in the absence of additional information."
>
> Bradley, (6) p.5

These statements are certainly true, and perhaps reflect the caution of an experienced statistician in resisting the temptation to infer causality from correlation. However, the suggestion that such factors as the decline in the average weight of automobiles and the decline in the lead content of gasoline may be judged as equally important factors in the decline in blood lead levels given current information is sure hyperbole. There is a sound basis for believing that lead in gasoline may contribute to lead in blood.

Simply, the major uses of lead (in decreasing annual volume by weight) are:

1. automobile batteries
2. gasoline additives
3. paints
4. electrical apparatus
5. construction (primarily for sound proofing)
6. ammunition

Source (9)

Of these, only the use of lead in gasoline and paint are likely to introduce lead into the environment in a way that could conceivably lead to human exposure. Certainly, humans are exposed to lead from other sources, including food, but there is a very good reason to believe that lead in gasoline can threaten human health. The "additional information" which Bradley sought exists; the ICF/OPRM analysis is based on a reasonable premise.

Uses of the Ethyl and ICF Studies and Associated Comments

It should be clear to anyone who has read this communication and earlier ones by the authors (12), Draper (5), Bradley (6), and ICF (13) that none of the studies done to date using the NHANES II data are perfect. However, all of the work so far has been useful, and each study has advanced the understanding of the role of the factors which may contribute to lead in blood. None of the studies should be used to the exclusion of the others. The studies should be regarded and used as an entity, a body of research viewed against the context of other epidemiological and toxicological studies of lead exposure.

REFERENCES

1) "The Relationship between Gasoline Lead Usage and Blood Lead Levels in Americans: A Statistical Analysis of the NHANES II Data," prepared for the Office of Policy and Resource Management, U. S. EPA by ICF, Inc., December 1982.

2) Ethyl Corporation's Comments following EPA's Public Hearing on Proposed Rule for Lead in Gasoline, Washington, D. C., September 7-8, 1982 - Submitted to EPA's Docket No. A-81-36, October 8, 1982.

3) "A Review of the Ethyl Corporation's Comments on EPA's Proposed Rule for Lead in Gasoline," prepared for the U. S. Environmental Protection Agency, Office of Policy and Resource Management, Office of Policy Analysis, by ICF Incorporated, November 1982.

4) "The Relationship between Gasoline Lead Emissions and Blood Poisoning in Americans," prepared for the Office of Policy and Resource Management, Office of Policy Analysis, by ICF Incorporated, October 1982.

5) "Comments on the Documents: Ethyl Corporation's Comments Following EPA's Public Hearing on Proposed Rule for Lead in Gasoline and Review of Ethyl Corporation's Comments by ICF," by Norman Draper for Ethyl Corporation, undated.

6) "A Discussion of Issues and Conclusions on Gasoline Lead Use and Human Blood Levels" by Ralph A. Bradley, for the Ethyl Corporation, undated.

7) Statement by Dr. Vernon N. Houk, Acting Director, Center for Environmental Health, Center for Disease Control, before the Subcommittee on Environment, Energy and Natural Resources, Committee on Government Operations, U.S. House of Representatives, Washington, D.C., April 14 1982.

8) "The Usefulness of the NHANES II Data for Discerning the Relationship Between Gasoline Lead Usage and Blood Lead Levels in Americans and A Review of ICF's Analysis Using the NHANES II Data," Prepared for the Environmental Protection Agency Office of Policy Analysis, by Daniel Violette and Craig Miller, March 1983.

9) U.S. Bureau of Mines, Mineral Fuels and Problems, 1975 Edition, p. 607.