#589

CUYAHOGA COUNTY HOSPITAL

• CLEVELAND METROPOLITAN GENERAL HOSPITAL •
3395 SCRANTON ROAD CLEVELAND, OHIO 44109 • 216-398-6000

HENRY E. MANNING
PRESIDENT

216: 398 6000
3633

WILLIAM H. ANDREWS, F.A.C.H.A.
VICE PRESIDENT OPERATIONS

February 9, 1979

Dr. Herbert Needleman
Director, Lead Exposure Studies
Children's Hospital Medical Center
300 Longwood Ave.
Boston, Mass. 02115

Dear Herb,

I was pleased to note by your listing in Science that you are going on to work on prenatal lead effects. I hope that you are exploring other contaminants, as smoking, drinking, marijuana, etc. in the same protocol. My own data, however, showed no relation of smoking in pregnancy or smoker currently in home to preschool blood lead level or to school age blood lead, FEP, or dentine lead.

On somewhat of an impulse, I'm sending along my CV, although I don't think of myself as being in the job market. I would be an interesting program to work on, particularly if it can be set up on a fairly large scale in a well controlled manner and I'd enjoy the challenge of tackling it. I could not, of course, consider such a move without an appointment at least equivalent to what I now have (rank, tenure, some opportunity to teach, etc.). If this is the level at which you are looking, I'd like to talk with you about it.

My own most recent lead study is still in writing. Part of the problem is time; writing always gets put aside in the press of other demands. It's not an easy one either. The data aren't as definitive as I had hoped, although the overall thrust is significant.

Best regards,

Claire

Claire B. Ernhart, Ph.D.

#589 Ref. II 4-3-87

BEFORE THE
ENVIRONMENTAL PROTECTION AGENCY

TESTIMONY OF CLAIRE B. ERNHART, PH.D.
ASSOCIATE PROFESSOR, DEPARTMENT OF PSYCHIATRY, AND
ASSOCIATE PROFESSOR OF REPRODUCTIVE BIOLOGY
CASE WESTERN RESERVE UNIVERSITY SCHOOL OF MEDICINE
ON REGULATION OF FUEL AND FUEL ADDITIVES

included parents intelligence
Psychologist

Low Level Lead Exposure: Its Effects on Intelligence and Behavior of Children

Claire B. Ernhart, Ph.D.

This paper is prepared at the request of the Lead Industries Association, Inc. as part of the testimony presented to the Environmental Protection Agency at hearings on April 15-16, 1982. The topic of inquiry is the desirability of changing regulations regarding permitted levels of lead in gasoline.

The issue I am discussing is the possible effect of low level lead exposure on the intelligence and behavior of children. After careful review of the published studies of others, and my own data, I have reached the opinion that if there are such effects, they are minimal. The effects that have been reported are small and have not been obtained consistently across studies.

More importantly, there are a number of methodological problems in these studies, including my own. These can lead to biases or distortions in the findings. Some of these biases are one-sided, i.e. they lead to finding an effect which might not be there or to exaggerating a real effect. When there is a true effect which is very robust, we might be able to ignore these biases. However, when a reported effect is small or is not replicated by others, it is imperative that we scrutinize the design and procedures of the studies to identify such problems. The methodological issues may be classified into five categories:

1. Control of Confounding Variables
2. Selection and Recruitment of Sample
3. Indices of Lead Exposure
4. Choice and Reporting of Performance Measures
5. Study-Wise Error and Sample Size

Control of Confounding Variables

Control of confounding variables is the single most important problem in this area of research. Confounding variables are other conditions which may be associated with the study variable, in this case lead, and which may influence the specified performance variable. Higher lead levels are often found in children whose parents have lower intelligence and less education and who are less able to provide a stimulating home and child rearing atmosphere. These conditions can and do influence performance measures.

A particularly important confounding variable is parent intelligence. It has long been known that this is the single most important correlate of child intelligence. Let us look at what happens when this is omitted or included in a study of lead effects.

Percent of Variance in Child Intelligence Accounted for by Lead Indicator Without and With Control of Parent Intelligence

	Preschool Intelligence Measure	School Age Intelligence Measure			
	Preschool Blood Lead	Preschool Blood Lead	School Age Blood Lead	School Age Erythrocyte Protoporphyrin	School Age Dentine Lead
ut Control of t Intelligence	10.2**	6.3*	17.6**	8.4*	3.1
Control of t Intelligence	1.2[a]*	1.4	7.7*	4.5	0.0

irth weight is also controlled in this figure.
$^{*}p < .05$
$^{**}p < .01$

The data are from two studies (1,2) in which I have participated. To simplify the table, I am here presenting only the results for the General Cognitive Index of the McCarthy Scales. (This corresponds to I.Q.) The figures represent portion of the variance in child intelligence related to variance in the indicated lead index. Except for the dentine lead results, all variances in which parent intelligence is not controlled are statistically significant. With inclusion of parent intelligence, the variances decrease dramatically and only two of the five barely reach simple (single-variable) statistical significance. In the large majority of studies, parent intelligence is not considered. It is apparent that this confounding variable can lead to a false conclusion that there is an effect or the exaggeration of an existing effect.

Using a similar mode of analysis but controlling for age, sex and a simple index of social class, Yule, Lansdown, et al. (3) reported that only 2 1/2 to 4 1/2 percent of the variance in intelligence is associated with lead. This is statistically significant. I wonder, however, if it would have been if parent I.Q. had been included in their design. Yule, et al. do carefully qualify their results and express a concern for the appropriate control of social factors. In their ongoing studies they are including intelligence test data from both the father and mother as well as detailed information on social and family factors. They are quite knowledgeable about methodological issues and I look forward to reading their further reports.

I hope that the importance of parent intelligence as a confounding variable will be recognized by other investigators. In other published studies to date, only the 1979 Needleman, et al. report (4) includes this as a covariate.

As suggested by Yule, et al., parent intelligence is not the only potentially confounding variable in these studies. For example, Milar and

colleagues (5) have reported that quality of stimulation and form of discipline in the home (Caldwell's HOME Scale (6)) are related to lead exposure. Rutter (7) has commented that even with a more comprehensive measurement and control of parental characteristics, family circumstances, and social environment there may still remain a small effect attributed to lead which is actually due to social or genetic factors.

Moving away from social variables, we find another possibly important confounding variable that has not been mentioned in any published study of cognitive and behavioral effects. Iron deficiency is not unusual in the population of children most at risk of lead exposure, that is, the urban inner city child. Incidentally, according to one of my colleagues in hematology, we will see an increased incidence of iron deficiency in our inner-city hospital. This prediction is based upon reduced support of the WIC program. Iron deficiency has been shown to increase susceptibility to lead. Furthermore, high levels of lead can lead to iron deficiency - the two conditions are not independent. Oski and Honig (8) and my colleagues at Case Western Reserve University, Lozoff and Brittenham (9), have related iron deficiency to developmental deficits in young children. Iron deficiency in elevated lead level children may well be producing effects sometimes attributed to lead. We are now including measures of serum ferritin in our research protocols.

Another aspect of the control of confounding variables has not been mentioned in investigations other than my own. It is well known in psychometrics that the measurement of psychological and social factors includes error. For most purposes, it is not biasing or distorting; it merely increases the difficulty of seeing an effect. When we use such a measure as a control variable in analyses of covariance, regression models, etc. or even in matching, the correction is not complete. In other words, when we think we have controlled for, say, parent I.Q., we have only partially controlled for it. The direction of the bias is toward undercorrection. I know of no effective satistical solution for this. What is important is that investigators be aware of this when they make interpretations of the small effects we see in these studies.

Summary. (a) Control of confounding variables is probably the single most important problem in this area of research. (b) Major factors that could be related both to lead level and to child performance are not included in most studies: these include parent intelligence, a number of social factors, and possibly, iron status. (c) Statistical techniques used to control confounding variables almost invariably undercorrect. (d) The direction of bias in this research is toward finding an effect that might not be there or increasing the apparent magnitude of an actual effect.

Selection and Recruitment of Sample

Two issues are of concern here. The first is the choice of the sample, the second is differential or biased loss of eligible cases.

Once an effect has been well established, the extent of the effect can be determined by selecting samples representing the entire population. Until then, it is more efficient to study the population most at risk for the target condition - one is more likely to find the effect, given it exists. Aside from proximity to an inadequately controlled industrial source, the children who are most apt to have elevated lead levels are from inner-city, low socioeconimic

status families. It was a sample of disadvantaged urban black children that we used in the research mentioned previously.

Selection from a more restricted source has another advantage – the sample is more homogeneous and the effects of social factor confounding are lessened.

Let us consider a discrepancy between two sets of data. These are teacher ratings of children using almost identical items. Within the Needleman, et al. study (4) there are significant differences between groups. In mine (2) there are no significant differences and the overall level is less favorable.

Comparison of Relationship of Lead Level Classifications to Teacher Rating Items in Two Studies*

Teacher Rating Item	Ernhart, et al. 1981		Needleman, et al. 1979	
	Low Lead (%)	Moderate Lead (%)	Low Lead (%)	High Lead (%)
Easily frustrated	23	19	11	25
Hyperactive	20	9	6	16
Impulsive	37	19	9	25
Distractible	33	31	14	36
Daydreams	13	25	15	34

* Lead levels as defined within respective studies

A major difference between the studies was that the Needleman, et al. sample was more heterogeneous with respect to social factors. Since these are correlated wih lead exposure, and have not been controlled in these data, it is possible that they caused the appearance of a lead effect.

A serious problem in some studies is failure to recruit subjects identified for inclusion in a study or attrition in longitudinal study. Such loss is more likely to occur among families who are socially and economically disadvantaged. These families are often quite mobile and tracing is difficult. Parents are often not motivated to bring children to a research center. With considerable effort and examination in homes or schools one can keep loss to a low level. When extensive loss occurs it is not random with respect to lead level. In the Needleman et al. paper (4), the loss rate due to lack of interest and mobility was appreciably and significantly greater for the high lead group than for the low lead group.

The direction of bias associated with extensive differential loss rate is not always easy to determine. Since only minimal information about these cases can be obtained, the effect of exclusion in a given study is unknown. A higher

loss rate in higher lead groups reduces the available data for the identification of an effect, given it exists, or for the identification of a threshold level, if there is one.

Summary. (a) One is more likely to find an effect, given there is one, by studying a population with a higher exposure level. (b) Confounding variables may be less relevant in a more homogeneous population. (c) Without extensive effort to minimize loss rate, loss due to lack of parental interest will be higher in high lead groups.

Measurement of Lead Exposure

The most frequently used index of lead exposure is blood lead level. This is considered to be a rather unstable index. Systematic repetition of blood lead measurement has not been used.

Dentine lead from shed deciduous teeth has been the index in the Needleman, et al. report (4), in a recent study by Winneke and colleagues (10), and it was one of several indices in my work. Variability (or nonconcordance) within and between teeth from a given child is noted in all three studies. A recent report of a study with dogs (11) indicated a discrepancy between central and lateral incisors. This descrepancy increased as dosage increased.

Erythrocythe protoporphyrin is a sensitive indicator of tissue effect and is now used in some screening programs. Interestingly, it also reflects iron deficiency and high values require further tests to differentiate conditions.

The indices are correlated, but the correlations we obtained, which are shown in the table, are not high enough for interchangeability. The table also includes preschool blood lead values obtained five years prior to the other indices. Although blood lead decreased significantly, the between years correlations did not differ much from the between indices correlations.

Means, Standard Deviations, and Intercorrelations of Lead Values

Variable	Descriptive Statistics*	Intercorrelations			
		Blood Lead Measures			
		Pre-school	School	FEP	Dentine
Lead					
Preschool blood lead	32.5 ± 12.7		.46†	.46†	.52†
School-age blood lead	26.9 ± 7.2			.51†	.43†
Free erythrocyte protoporphyrin (FEP)	27.2 ± 19.6				.66†
Dentine	6.0 ± 5.9				

* N = 63 except for dentine data for which N = 34

† P < .01

One solution to the problem of choice of an index, given the instability of each, is the use of several indices in a given study. To date, this approach has been used only in our work. (Needleman, et al. had earlier (4-5 year) blood lead values for 81 of their children. They did not report the relations between these lead levels and any outcome measures for their high and low lead groups. These data could have provided another test of the important question of persistence of effect.)

Summary: (a) Most studies use a single index of lead level although each has some degree of unreliability and the intercorrelations among these are not high. (b) Use of several indices in the same study will result in more precise identification of lead effects, if any.

Choice of Performance Measures

Performance measures selected for study should meet two criteria: (a) they should be psychometrically sound and, (b) they should be indicators of whatever attributes one expects to be affected by lead. Types of measure may be hierarchically organized according to the importance and psychometric merit of available procedures.

Some investigators have reported lead effects with measures that are not psychometrically standardized. In fact, for some, published procedures are not available. While the inclusion of a few experimental procedures is reasonable, reporting the results of these in the absence of positive findings with measures with known characteristics simply clouds the field. The reader does not know just what is measured, let alone how well it is measured or how the data compare with data from other research uses. With measures of unknown psychometric characteristics, a strong possibility of chance findings must be entertained. Furthermore, as discussed later, there are problems that are not ususally recognized in having a large battery of procedures. One should have a strong rationale for the inclusion of experimental procedures at the cost of standardized procedures.

The more important measures include intelligence, academic achievement (for school age children), behavior, and perceptual-motor skills. This order reflects both availability of psychometrically sound measures, and, probably, importance for the developing child.

Most investigators include reliable and valid individually administerd tests of intelligence. Where positive findings are reported, they are most likely to be obtained with these instruments. Effects are small and, as indicated above, most investigators fail to include any consideration of parental intelligence and have only limited information regarding pertinent social factors.

If children are of school age, standardized measures of school performance are indicated. Yule, et al. (3) reported significant effects of lead on several academic achievement tests, but once again their study is limited in its control of social factors including parent intelligence. De la Burde and Choate (12), McNeil and Ptasnik (13), and Needleman, et al. (4) have failed to find significant lead effects with reliable standardized measures of academic achievement. If lead has an effect that is more than minimal, it should have been detected with these measures.

Behavior (hyperactivity, short attention span, impulsiveness, etc.) can be rated by teachers, parents, or examiners. De la Burde and Choate (12) reported significant lead effects with examiner ratings but did not control social factors. Milar and associates (14) report no evidence of lead effects on two parent rating scales of activity and one free-field experimental measure of gross activity. These results, considered with the previously mentioned Needleman, et al. teacher ratings, which lack control of confounding variables and our negative findings lead to the conclusion that the evidence for lead effects on behavior is quite tenuous.

There is little consistency in results relating lead to psychomotor function. Positive findings with the Bender Gestalt test are reported by two research groups (de la Burde and Choate (12) and Winneke et al. (10)) but four other groups failed to find significant effects. Positive findings have not been obtained with other standardized procedures. The occasional significant findings reported with experimental procedures have not replicated by other investigators.

Summary. (a) Some performance measures are unstandardized experimental procedures of unknown psychometric quality. Nevertheless, positive results with these are sometimes emphasized while negative findings with well standardized procedures are ignored. (b) Measures may reasonably be ordered: intelligence, academic performance, behavior disturbance, and psychomotor skills. Positive findings with the better standardized measures of intelligence and academic achievement are more critical than those with the more experimental behavioral rating and psychomotor procedure.

Study-Wise Error and Sample Size

One might think that the inclusion of a very large number of performance measures improves a study since it decreases the liklihood that something important is overlooked. Some investigators, however, have not been aware of the associated problems. When there are a large number of statistical tests on a set of data there is a drastically increased liklihood that some of the results will be statistically significant in the absence of a true effect. If, for instance, one had statistical analyses of five independent (uncorrelated) performance measures in the absence of a true effect, the probability of obtaining at least one analysis "significant at the .05 level" is .23, not .05. One could reduce this risk to .05, study-wise, by requiring an .01 level for the individual analyses. If the number of individual statistical analyses is increased to, say, 50, the probability of having one or more individual analyses of independent variables reach the .05 level in the absence of true effect is .92. In order to reach a study-wise error rate of .05 one would have to ignore any statistical analysis for which the calculated probability is greater than .001. This situation is not farfetched. The lead study that appears to have the largest number of individual (or univariate) statistical tests is that of Needleman et al. (4) with at least 52 performance measures. By these calculations only two statistical tests in that study should be considered. These are two trial blocks of the experimental unstandardized reaction time procedure.

The above is somewhat oversimplified since the performance measures are certainly not independent. Failure to consider the intercorrelations among performance measures can lead to distortions in the statistical analyses. Unlike some of the other biases in these procedures, the directions of

distortion are not always predictable.

There are effective procedures to reduce the number of statistical analyses in a large data set. These inlcude data reduction techniques and the use of multivariate statistics. As a bare minimum, the investigator should interpret very conservatively a very few "statistically significant" tests in group of tests of outcome measures. This caution is not common in this area of research, yet failure to consider the issue is probably as important as inadequate control of confounding variables.

A related technical issue is the relation of sample size to the size of the set of variables. A reasonable rule of thumb is that the ratio of number of subjects to number of variables should be about 10/1. Most of us have violated this to some extent, but the most flagrant violation is in the Needleman et al. report (4) with its very large number of variables. The problem is that if the data had been analysed by appropriate use of multivariate procedures, there would not have been enough statistical power in the analysis and true effects, if any, might not be detected. The direction of distortion here is not easily predicted but statisticians warn that distortions may be quite serious.

Summary. An investigator with a large number of performance measures must carefully consider several major technical difficulties: (a) the univariate statistics which are most frequently used will greatly increase the liklihood of concluding that there are effects which are not true, (b) intercorrelations among performance measures leave one unable to make adequate corrections for these error rates and distort the study-wise probability situation in ways that can't be predicted and, (c) in order for the appropriate multivariate statistical tests of a large data set to be reasonably powerful, a very large number of subjects must be enrolled.

Implications

The methodological issues raised are not trivial. Where a direction of bias can be identified, as for inadequate control of confounding and inadequate handling of study-wise error, the direction is toward exaggerating an existing effect or reporting an effect which is not there. When we considered all published studies and our own data in light of these issues we found that we could not support a statement that low level lead exposure influences the intelligence or behavior of children.

Infants as a Special Group

The conclusions drawn so far have dealt with preschool and school age children. It has been suggested that infants prior to birth and in the early period of infancy might be particularly vulnerable to toxic effects since this is the time of maximal brain development. The hypothesis is being studied at several centers, including ours. The prospective study design appropriate for this takes time; at the moment all I can provide is preliminary data reflecting exposure of the fetus.

Levels of lead in cord blood are low. We collected maternal and cord bloods from a preliminary sample of 67 mother-infant pairs. This is from a population at risk for high lead values. The sample is of low socioeconomic level. The women reside, often in older housing, in the inner city area of an industrial city. The incidences of poor nutrition, smoking, alcoholism, and polydrug use are considerably higher than for the general population. It is possible that women occupationally exposed or who engage in unusual behavior

(pica of pregnancy, drinking moonshine liquor, etc.) will have higher lead values; aside from aberrant conditions our means will be about as high as one might find. The means are about 16 µg/dl whole blood for both maternal and cord samples. The highest cord lead level was 23.9 µg/dl whole blood. These figures are appreciably below the lead levels of 30-35 µg/dl suggested in several sources as being associated with some risk to the fetus or newborn.

We have not completed collecting the data to determine whether there are cognitive and behavioral effects associated with these levels. My expectation is that effects, if any, with neonates, will be minimal and nonpersistent. This expectation is based in part on the evidence available for older children and, in part, on experience in infancy studies. Newborns are more robust than is usually thought. If we exclude markedly damaged infants, those of us doing research on infant development, find it difficult to demonstrate persisting effects of neonatal risk conditions such as anoxia or prematurity.

In studies of the development of infants at risk, parental, social, home stimulation, and nutrition factors overwhelm other variables. When we do lead research we are trying to detect a possible faint signal through continuing noise. The children most likely to be exposed to lead are these most likely to be poorly nourished, to live in substandard housing, to have a mother who is herself a child, to have no visible father, to have one or both parents actively abusing drugs, to witness or be victims of stress and violence, and are least likely to have remedial or enriching education opportunity... We are here considering the possibility of a just noticable difference associated with lead. The real concern should be with the reduction of programs to ameliorate the real problems of these children.

I recently wrote that there may be a risk of frank lead poisoning in vulnerable children and that this risk justifies limitations on lead in gasoline. The statement was made out of an interest in being very conservative pending continuing research results. Three conditions, which I had not completely reviewed, would support this conservative position. First, there must be a reasonable number of children of elevated lead level who might be near a threshold of effect. Second, these children would not have been identified by screening programs with the consequent removal of primary sources of exposure. Third, the increase in airborne lead associated with changes in gasoline additives must be great enough to increase the lead level in these children to the point that they cross a threshold of effect. Information on these condtions is beyond my expertise.

I'd like to quote from an editorial by Martin Bax in Developmental Medicine and Child Neurology.

"It is very tempting to accept some physical aspect of the environment - be it blood lead, food additives or too much sugar - as the cause of children's problems. By doing so, we need look no further at the child and his family and can avoid uncovering the uncomfortable fact that adults are not invariably sensitive to children's needs. This of course extends beyond meeting their simple physical needs to providing an environment in which cognitive and social skills can develop. Knowing one s own frailties, it is unwise to point the finger at another human being and easy to elect the petrol tanker as the source of all the mischief."

References

1. Perino, J. and Ernhart, C.B. The relation of subclinical lead level to cognitive and sensorimotor impairment in black preschoolers. Journal of Learning Disabilities, 1974, 7, 616-620.

2. Ernhart, C.B., Landa, B. and Schell, N.B. Subclinical levels of lead and development deficit - a multivariate follow-up reassessment. Pediatrics, 1981, 67, 911-919.

3. Yule, W., Lansdown, R., Millar, I.B. and Urbanowicz, M. The relationship between blood lead concentrations, intelligence and attainment in a school population: A pilot study. Developmental Medicine and Child Neurology, 1981, 23, 567-576.

4. Needleman, H.L., Gunnoe, C., Leviton, A., Reed, M., Peresie, H., Maher, C. and Barrett, P. Deficits in psychological and classroom performance of children with elevated dentine lead levels. New England Journal of Medicine, 1979, 300, 689-695.

5. Milar, C.R., Schroeder, P.M., Dolcourt, J.L. and Grant, L.D. Contributions of the caregiving environment to increased lead burden of children. American Journal of Mental Deficiency, 1980, 84, 339-344.

6. Bradley, R. and Caldwell B. Early home environment and changes in mental test performance from 6 to 36 months. Developmental Psychology, 1976, 12, 93-97.

7. Rutter, M. Raised lead levels and impaired cognitive/behavioral functioning: A review of the evidence. Developmental Medicine and Child Neurology, 1980, 22, 1-25.

8. Oski, F.A. and Honig, A.S. The effects of therapy on the developmental scores of iron-deficient infants. Journal of Pediatrics, 1978, 92, 21-25.

9. Lozoff, B., Brittenham, G.M., Viteri, F.E., Wolf, A.W. and Urrutia, J.J. The effects of short-term oral iron therapy on developmental deficits in iron-deficient anemic infants. Journal of Pediatrics, 1982, 100, 351-357.

10. Winneke, G., Brockhaus, A., Kramer, U., Ewers, U., Kujanek, G., Lechner, H. and Janke, W. Neuropsychological comparison of children with different tooth-lead levels. Preliminary report. Unpublished paper presented at the International Conference on Heavy Metals in the Environment. Amsterdam, September, 1981.

11. Kaplan, M.L., Peresie, H.J. and Jeffcoat, M.K. The lead content of blood and deciduous teeth in lead-exposed beagle pups. In H.L. Needleman (Ed.), Low Level Lead Exposure: The Clinical Implications of Current Research, New York: Raven, 1980.

12. De la Burde , B. and Choate, M.S. Early asymptomatic lead exposure and development at school age. Journal of Pediatrics, 1975, 87, 638-642.

References (continued)

13. McNeil, J.L. and Ptasnik, J.S. Epidemiological study of a lead contaminated area. Draft final report, Project LH-208. International Lead and Zinc Research Organization, 1975.

14. Milar, C.R., Schroeder, S.R., Mushak, P. and Boone, L. Failure to find hyperactivity in preschool children with moderately elevated lead burden. Journal of Pediatric Psychology, 1981, 6, 85-95.

15. Bax, M. Editorial. Lead and impaired abilities. Developmental Medicine and Child Neurology, 1981, 23, 565-566.

CHILDREN'S HOSPITAL OF PITTSBURGH
125 DESOTO STREET
PITTSBURGH, PA. 15213

BEHAVIORAL SCIENCE DIVISION
Telephone: 412/647-5100

June 30, 1986

Claire B. Ernhart, Ph.D.
Department of Psychiatry
Case Western Reserve University
Cleveland Metropolitan General Hospital
Cleveland, OH 44109

Dear Dr Ernhart:

I have read your reponse to my letter in the Journal of Learning Disorders with considerable interest. I believe that careful review of your statements will show that you have made a rather serious error in calculating the power of your study. I bring it to your attention so that you can check it and then perhaps correct your published statement.

In your letter, you state that by calculating your power using a T-test model, I underestimated the true power of your last reanalysis. I have recalculated the power in your second reanalysis, using the values you supplied. This time, I used a regression model. I returned to Jacob Cohen's book, Chapter 9 : F tests of variance proportions in multiple regression/ correlation analysis.

Assuming that the main effect (lead) to be measured is "moderate to small, but not trivial" $f^2=0.1$, n=45, alpha=0.05

then $L= f^2 \times n=4.5$

Using table 9.3.2 from Cohen's book, I find the power to be .37 for u=4; .41 for u=3. I enclose a marked copy of that table for your information.

Perhaps even more to the point are the words of Professor Cohen himself, contained in the ILZRO report you cite. Referring to your study and the conclusions you drew, he states:

> Positive conclusions of "no effect are drawn for all instances [by Ernhart] in table 5 where the 2-tailed significance is not met. This reasoning is always formally incorrect, and can be particularly misleading when the statistical power of the test is not high, as is the case here. In other words, it is a far cry from failing to find an effect to finding that there is no effect, and particularly when the a priori chance ofdetecting a real (nontrivial) effect was not very good... One can appreciate how weak the [Ernhart] argument is by performing a power analysis on the crucial tests in table 5. If one assumes the population

> r in these tests is .3 (operationally defining a medium degree of relationship) so that R^2 = .09, a two-tailed .05 test with n=63 has only a .65 (power) chance of yielding a significant result. For the n=33 cases with dentine lead under these conditions, the power is only .37!

Since you cited this report, I assume you have read it. In these statements, Professor Cohen is referring to your first reanalysis published in Pediatrics,but the same argument applies. Since the n in your second reanalysis is even smaller, it applies, I believe even more stringently. It is difficult to see how reducing the sample by 16 more subjects could increase the power.

You state in your letter that covariate control (of socioeconomic status) is not essential when dealing with an "essentially invariant condition (SES in our study)." but I want to again remind you of your first published description of your sample in J. Learn. Dis. 1974. Referring to the parents of the subjects in your study you stated:

> They ranged from managers, clerical workers, skilled and unskilled workers to service workers and welfare recipients. Thus it might be reasonable to conclude that, at least among the black community, children in the lower middle or lower-lower classes have an equal chance of being affected by moderate amounts of lead.

The use of the words "ranged" and "lower-middle or lower-lower classes" in your first article conveys a very different picture than "essentially invariant condition (SES in our study)" used in your most recent letter. I find the two apparently opposing descriptions of your sample difficult to reconcile, and do not believe the locution "essentially" carries enough explanatory power to cover the gap.

Finally, one of the conventional canons of scholarship, in addition to accuracy, is completeness. In this regard, quoting the EPA expert Committe's Draft report's criticisms of my work, when you know that CASAC directed that they be withdrawn from the Critieria Document, is a serious misrepresentation of the status of my studies and the regulatory process. I commend to you Page 12-79 of Volume IV of the last draft of the EPA Criteria Document which refers to my work as follows:

> Reanalyses carried out in response to the Committee's recommendations have been reported by Needleman (1984) and EPA' Office of Policy Analysis as confirming the published findings on significant associations between elevated dentine lead levels and decrements in IQ, after correcting errors in data calculations detected in earlier published analyses and using alternate model

specifications that incorporated better control for potentially confounding errors.

In the future, if you cite the EPA's Committee Report, I request that you accurately convey the the entire verdict, and cite the final EPA statement. The ethics that govern scholarly conduct should compel you to this step without advice from me. I think you should also read and reflect on the EPA's evaluation of your work contained on pages 12-74, 12-75. Here the document states that your studies do not allow for a definitive conclusion of "no effect."

When there are differences in opinion about computation and quotation, there always exists the possibilty of resolution of the conflict. I would ask you to recalculate the power analysis, consult with your statistical consultants, perhaps even the frequently quoted Professor Jacob Cohen. If you still believe your power analysis is correct, please let me know. If on the other hand, you find you have misrepresented the power to find a lead effect when you stated it was .71, I believe you are obliged to bring it to the attention of the editors of the Journal. Scholarly completeness would dictate that you also address the discordance between your two descriptions of the social class distribution of your two communications. Finally since you quoted the EPA Panel, the same need for comprehensiveness as opposed to selective quotation should require you to acknowledge that the EPA has expunged the Panel report you quoted, and has accepted the conclusions of my study showing an effect of lead on IQ as valid, while expressing scepticism about your "no-effect" conclusion.

Yours sincerely,

Herbert L. Needleman M.D.

HLN/mr

TABLE 9.3.2

POWER AS A FUNCTION OF L AND u AT a = .05

	L											
u	2.00	4.00	6.00	8.00	10.00	12.00	14.00	16.00	18.00	20.00	25.00	30.00
1	29	52	69	81	89	93	96	98	99	99	*	*
2	23	42	58	72	82	88	93	96	97	99	*	*
3	19	36	52	65	76	84	90	93	96	98	99	*
4	17	32	47	60	72	80	87	91	94	96	99	*
5	16	29	43	56	68	77	84	89	93	95	98	*
6	15	27	40	53	64	74	81	87	91	94	98	99
7	14	25	38	50	61	71	79	85	89	93	97	99
8	13	24	36	48	59	68	77	83	88	92	97	99
9	13	23	34	45	56	66	74	81	86	90	96	99
10	12	21	32	43	54	64	72	79	85	89	96	98
11	12	21	31	42	52	64	70	78	83	88	95	98
12	11	20	30	40	50	60	69	76	82	87	94	98
13	11	19	29	39	49	58	67	74	80	85	93	97
14	11	18	28	37	47	57	65	73	79	84	93	97
15	11	18	27	36	46	55	64	71	78	83	92	97
16	10	17	26	35	45	54	62	70	76	82	91	96
20	10	16	23	31	40	49	57	65	72	78	88	94
24	09	15	21	29	37	45	53	60	67	74	85	92
28	09	14	20	27	34	42	49	57	64	70	82	91
32	08	13	18	25	32	39	46	53	60	67	80	88
40	08	12	17	22	28	40	41	48	55	61	74	84
50	08	11	15	20	25	31	37	43	49	55	69	80
60	07	10	14	18	23	28	33	39	45	50	64	75
80	07	09	12	16	20	24	28	33	38	43	56	67
100	07	09	11	14	18	21	25	29	34	38	50	61

* Power greater than .995.

#589

Jacob Cohen, Ph.D.
45 West Tenth Street
New York, NY 10011

(212) 254-0283

October 23, 1986

Dr. Claire B. Ernhart
Highland View Hospital
3395 Scranton Road
Cleveland, OH 44109

Dear Dr. Ernhart,

I hope you will understand my intention to studiously avoid embroilment in any part of your controversy with Dr. Needleman other than the discrete issue of the power analysis for the Case 1-1 specifications you described in your published Reply and again in your letter to him of July 23 last.

To begin with, I believe that it was (9.3.8), not (9.3.9) to which you meant to refer in the letter. An $f^2 = .10$ implies that the numerator of the l.h. term in (9.3.9), i.e., the squared multiple partial correlation equals .0909, since $f^2 = .0909 / (1 - .0909) = .10$. In (9.3.8) the numerator of the l.h. term is the square multiple semipartial correlation, which for the data given implies the value .06636, which, in turn, implies that the denominator of (9.3.8) should equal 1 - (.27 + .06636) = .66364, with the result that $f^2 = .06636 / .66364 = .10$, as specified. The incorrect substitution you give in the letter would make $f^2 = .10 / (1 - .37) = .1587$, a much larger value. When $f^2 = .10$ is used, (9.3.8) or (9.3.9) yield L = .10 (45 - 1 - 3 - 1) = 4.00 (not 6.35), and entry into Table 9.3.2 produces power = .52.

Now, unhappily, I must tell you that in the course of preparing a new edition of the power book, I discovered an error in the formula for L (9.3.1): L equals f^2 times (u + v + 1), not v. (This of course changes all the other L formulas.) I also expanded the tables to get more accurate power values. Now, v = N - u - w - 1) = 45 - 1 - 3 - 1 = 40. Thus, the correct L for your problem is .10 (1 + 40 + 1) = 4.20. Interpolating in Table 9.3.2 yields power = .54. Using my new expanded tables yields power = .50.

Although I make no claim for any special competence in the substantive area involved, I would nevertheless hazard the opinion that the $f^2 = .10$ predicated is a relatively large effect for the circumstances at issue. It is, for example, much larger than than my operational definition of "small" (.02). An $f^2 = .10$ implies a partial R^2 (here, a partial r^2, since u = 1) = .10 / (1 + .10) = .09, hence a partial r = .30. However one may characterize an effect of this size, it is certainly not negligible (see, for example, pp. 16-17 in the power book). The

analysis then indicates that the detection of an effect of this magnitude would be a fifty-fifty proposition and leaves the possibility of a nontrivial effect a very real one.

While working on this material I received a phone call from Dr. Needleman to whom I expressed the feelings in the first paragraph above. I am sending him a copy of this letter which I hope ends my involvement in this affair.

Sincerely yours,

cc: Herbert L. Needleman, M.D.

Jacob Cohen

#589

Cleveland Metropolitan General Hospital

Highland View Hospital

3395 SCRANTON ROAD • CLEVELAND, OHIO 44109 • 216-398-6000

HENRY E. MANNING
PRESIDENT

JAMES KRETZSCHMAR
SENIOR VICE - PRESIDENT
OPERATIONS

July 23, 1987

Herbert L. Needleman, M.D.
Children's Hospital of Pittsburgh
125 Desoto Street
Pittsburgh, Pennsylvania 15213

Dear Dr. Needleman:

This is written in response to your letter of June 30, 1986. The slurs on my integrity and scholarship are noted, but I will not join you in that kind of behavior. The following is limited to factual material in the hope that some of the issues will be clarified. I have no obligation to do this other than my belief that openness about one's work is one of the hallmarks of science.

Your second attempt to use power analysis with respect to my data is also wrong, and wrong both with respect to the model chosen and to the execution of the formula for the chosen model. You ignored the very important point made in my reply in J. Learning Disabilities, i.e., "...the correct power analysis should include the covariates of the statistical models..." You selected the general model (p.414, Cohen, 1977) rather than the model appropriate to the design, that for the increment in variance. The equation was then used incorrectly in that you substituted **N** for **v**. You then entered Table 9.3.2 with **u** = 4 (covariates + IV) or 3 (covariates only).

The appropriate model for determining the power of an increment in variance is 9.3.2 Case 1, (more specifically Case 1-1) described on p. 423-425 of Cohen. The covariates are represented as set **A**, with **w** = 3 (3 covariates). Lead is represented as set **B**, with **u** = 1 (1 IV). Now solve the correct equation, (9.3.9), for **L**.

$$L = \frac{.10}{1 - .37} \times (45 - 1 - 3 - 1) = 6.35$$

The first part is the f^2 of (9.2.3); the second is the v of (9.3.7). One then enters the table with $L = 6.35$ and $u = 1$ (not 3 or 4). Interpolation yields power = .71. The same procedure applies for the power (.85) for the even more important analysis relating preschool Pb to schoolage GCI.

I don't think you understand power analysis. Covariate control is usually used to adjust for other factors in a design with confounding, but it can also be very effective in increasing statistical power. I do consult with statisticians for complex analyses, but I teach power analysis in seminars on research design for residents.

Cohen's comments on the power of my analyses are correct as far as they go, but he didn't incorporate the covariate data in his cursory review of the subject. Cohen was incorrect when he attributed a "no effect" inference to me. Shortly after Cohen provided this review to ILZRO, Ms. Volpe of the ILZRO staff discussed it with me. She then sent me a copy of her notes of that discussion (April 1983). The relevant section states;

> "Ernhardt (sic) says that she has never stated there is 'No effect.' She is not sure that there is no effect, therefore, she has said that if there is an effect, it is minimal. Ernhardt says that her study is not sufficient for a no effect conclusion."

Maybe I should have written to Cohen and demanded a correction. I don't believe that I have anywhere made an unqualified inference of no effect for the overall results of a study. I do agree that there may be a very small and difficult-to-detect effect. The point that I think is bothering you is that if one takes into account the methodological limitations of the studies published to date, the evidence of more than a minimal effect is very thin indeed and that this is inconsistent with the position you have espoused.

You advised me to "read and reflect on" pages 12-74, 12-75 of the draft of the EPA Criteria Document. The wording "...results of the reanalyses do not allow for a definitive conclusion of 'no-effect,' either (as stated by Ernhart, 1963)" can be interpreted to mean that I said that there is no effect or that I said that the reanalyses do not allow for a definitive conclusion of no effect. Obviously, you chose the incorrect interpretation. My 1983 EPA file includes detailed correspondence but nowhere do I say that there is no effect, except insofar as the term is qualified by "statistically significant," "detectable," etc. The issue is addressed: "In conclusion, these sets of data do not support an inference of low level lead effects. Unfortunately, the rather low statistical power prevents a definitive conclusion of no effect." (Memo to EPA dated August, 1983.) You used a less than reliable secondary source to attribute to me an improper inference.

July 23, 1986

You wrote about a lack of covariate control of SES in my work. Of course SES among my subjects was not perfectly homogeneous (in the sense of homogenized milk) but if you ever visit the Jamaica and Far Rockaway sections of Queens you will see pervasive socioeconomic distress. But why is this point so important to you? Do you hope that continued sniping on this issue will lead either me or other objective researchers to forget that some of your analyses and graphic presentations fail to incorporate any control of confounding? This issue appears to be nothing more than an attempt to divert attention from the serious problems in your work. I trust you realize that if you quote the Perino & Ernhart article again, it would be a serious error to fail to take into account the change in inference associated with the correction of df, as noted in my reanalysis paper.

You wrote about the "conventional canons of scholarship" and yet almost every aspect of your work is biased either by methodological naivete' (as in testing low lead children first with differentially experienced examiners and with all of the other potential problems inherent in such a protocol), by the appearance of opportunism in the selection of cases (including the failure to mention the implications of excluding the high IQ child with the history of plumbism), opportunism in the selection of variables (example: reporting Verbal IQ rather than the more reliable and valid Full Scale IQ), by clear errors in conclusions presumably based on statistical analyses, by possible problems in the validity of assessment shown by Verbal - Performance discrepancies, by failure to control for or to recognize the possible influence of uncontrolled confounders, by selection of those portions of your results that are most consistent with the position you espouse, by failure to make available results not consistent with your position, by failure to recognize the problem of mulitple statistical tests, and so on and on.

Some of these were noted in the Cohen review which you have; some were identified by the EPA Expert Committee; some have been described by me in several sources. Thus you are aware of these problems and I need not take valuable time to document them for you. The renalyses that you reported (Science, 1985) did not, could not, address many serious problems noted by the Committee in their short visit to your office and review of materials supplied. (By the way, the reanalyses look strange. Did you have typing errors, analytic errors or a suppression effect?)

Given the kind of advice that you have given me, you should consider writing a paper in which you acknowledge the problems in your work and discuss the ways in which they affect the inferences drawn. Even though your attention has been directed to many of these problems, you have continued to act as if they don't exist. As a fairly obvious example, you continued to copy in publication, and to permit the copying, of the graph of teacher ratings, yet you never discuss the limitations of this with respect to confounding or the fact that your unpublished replication didn't replicate.

While it's much less important, I frequently note carelessness in your work. For example, in your letter to J. Learning Disabilities you indicated that the children in the Perino & Ernhart study were males, but we had both sexes. You cited J. Educational Psychiatry. There is no such journal. It should have been J. Educational Psychology.

July 23, 1986

You wrote about the EPA Expert Committee. The work of this Committee was a valid review by objective, unbiased and knowledgeable individuals. I cooperated fully, including the provision of copies of all data, efforts to locate additional requested information (as hematocrit data for the correction of PbB for the children of my study and the Yamins study) and full reanalysis. I did find the one serious error in the df in the Perino & Ernhart analyses. I found it, I reported it, and I assume responsibility for not having detected it earlier. This is ethical behavior. Nothing was hidden; nothing was held back.

You apparently think that I should accept and cite the conclusions drawn at EPA. EPA can accept, reject, modify, or expunge materials in its documents to suit its various purposes. How can it possibly expunge the errors detected in your work? I put considerable effort into weighing available evidence, including biases that might bear on that evidence and I think I would be most remiss in my work if I simply accepted as dogma a statement from EPA, or anyone else.

I have no interest in carrying this matter further since it takes time from my research. As suggested in my reply in the J. Learning Disabilities, it is time to leave behind the limited and problematic old studies and to emphasize the more methodologically sound investigations now ongoing. However, if you misquote any part of this letter, or if you quote out of context in such a way as to distort anything said above, I will copy this letter and yours of June 30 to a lengthy list of people working in this field.

Sincerely,

Claire B. Ernhart, Ph.D.
Associate Professor
Departments of Psychiatry
and Reproductive Biology

CBE/dhp