# Low-Level Lead Exposure and the IQ of Children

## A Meta-analysis of Modern Studies

Herbert L. Needleman, MD, Constantine A. Gatsonis, PhD

We identified 24 modern studies of childhood exposures to lead in relation to IQ. From this population, 12 that employed multiple regression analysis with IQ as the dependent variable and lead as the main effect and that controlled for nonlead covariates were selected for a quantitative, integrated review or meta-analysis. The studies were grouped according to type of tissue analyzed for lead. There were 7 blood and 5 tooth lead studies. Within each group, we obtained joint *P* values by two different methods and average effect sizes as measured by the partial correlation coefficients. We also investigated the sensitivity of the results to any single study. The sample sizes ranged from 75 to 724. The sign of the regression coefficient for lead was negative in 11 of 12 studies. The negative partial *r*'s for lead ranged from −.27 to −.003. The power to find an effect was limited, below 0.6 in 7 of 12 studies. The joint *P* values for the blood lead studies were <.0001 for both methods of analysis (95% confidence interval for group partial *r*, −.15 ± .05), while for the tooth lead studies they were .0005 and .004, respectively (95% confidence interval for group partial *r*, −.08 ± .05). The hypothesis that lead impairs children's IQ at low dose is strongly supported by this quantitative review. The effect is robust to the impact of any single study.

(*JAMA*. 1990;263:673-678)

THE NEUROTOXIC properties of lead at high doses have been recognized for at least a century and are not a matter of dispute. In 1943, Byers and Lord[1] first suggested that childhood exposure to doses of lead insufficient to produce clinical encephalopathy was associated with deficits in psychological function. The question of low-level lead exposure has been studied widely over the past two decades and, in contrast to high-dose lead exposure, has been the source of considerable contention. Several methodological difficulties encountered in the conduct of these studies have contributed to the controversy. Among them are (1) selecting adequate markers of exposure or internal dose, (2) measuring outcome with instruments of adequate sensitivity, (3) identifying, measuring, and controlling for factors that might confound the lead effect, (4) recruiting and testing a sample large enough to provide adequate statistical power to detect a small effect, and (5) designing a study that avoids biases in sample selection.

From the Department of Psychiatry, University of Pittsburgh (Dr Needleman), and the Department of Statistics, Carnegie-Mellon University (Dr Gatsonis), Pittsburgh, Pa.

Reprint requests to Department of Psychiatry, University of Pittsburgh, 3811 O'Hara St, Pittsburgh, PA 15213 (Dr Needleman).

A number of reviews of studies on the effects of low-level lead exposure on the neuropsychological function of children have been published.[2-5] The outcome of major focus in these reviews has been psychometric intelligence. The general approach was to provide narrative summaries in which the epidemiologic and statistical issues often received limited critical attention. Where quantitative synthesis was attempted, it consisted of a simple tally of those studies showing statistically significant effects (at the .05 level) vs those that did not. This approach gives undue emphasis to the individual study's *P* value and attaches equal weight to all studies without regard to their specific merits or flaws. The size of the effect measured in each study is generally ignored in the process.

The statistical techniques that have been subsumed recently under the rubric of meta-analysis offer a framework within which formal research synthesis can be conducted with more clearly defined methods and criteria.[6-9] In this approach, individual studies are treated as data points in a larger "meta-study." Summary measures from each study are pooled by one of a number of techniques, and quantitative inferences are drawn about the research questions of interest. The difficulties entailed in combining dissimilar studies ("apples and oranges") is a concern for any meta-analysis.[6,8] It points to the need for some measure of commonality in the studies that are being combined. At the same time, the usefulness and novelty of meta-analysis lies in the fact that it enables the investigator to combine the results of studies that *differ* in several respects, while addressing the same research questions.

The first meta-analysis of six lead-IQ studies was reported by Schwartz et al[10] in 1985. They used Fisher's aggregation technique to calculate a joint *P* value of .004 for the effect of lead on IQ in the six studies. Needleman and Bellinger[11] extended the analysis by Schwartz et al and also used Fisher's technique on pooled tooth and blood lead studies.

In the last few years, a substantial number of new epidemiologic studies from various nations, using more refined designs, larger sample sizes, and more sophisticated statistical techniques, have been reported. This presents an opportunity for a more comprehensive meta-analysis. Herein, 12 recent studies are reviewed and a quantitative synthesis of their results is presented. The major outcome of interest is full-scale IQ, although many studies also examined the effects of lead exposure on important functions such as school performance, reading ability, and classroom behavior. All studies reviewed employed multiple regression analysis in which the dependent variable (IQ) was treated as continuous. Lead exposure was classified by one of two methods: blood or tooth lead level. In contrast to earlier attempts, this analysis divides the studies by tissue analyzed and combines inferences within tissue groups. The question of possible bias in the obtained sample of studies (known as the "file drawer" problem) is addressed. Moreover, the aggregate effect of the exclusion criteria is assessed by performing an analysis that combines all of the initial 24 studies. The sensitivity of the results of this meta-analysis is further investigated by eliminating each of the included studies, one at a time, from the analysis and observing how this affects the conclusions. The statistical power of each study to find an effect is also computed.

This article presents a discussion of some methodological difficulties encountered in the studies reviewed, examines the critical question of effect assessment in pollutant studies, and concludes with comments on the difficulties entailed in drawing causal inferences from observational studies of lead exposure and intellectual development.

Table 1.—Candidate Studies for Meta-analysis*

| Study | Year | No. of Subjects | Tissue | Lead Level† | Data Analysis | Included/Reason for Exclusion | Comments | Lead Effect, $P<.05$ |
|---|---|---|---|---|---|---|---|---|
| Kotok[12] | 1972 | C = 25; E = 24 | Blood | C = 38; E = 81 | *t* test | No/G,D | A, B | No |
| Perino and Ernhart[13] | 1974 | C = 50; E = 30 | Blood | C<30; E = 40-70 | Multiple regression | No/H | . . . | Yes |
| Rummo et al[14] | 1979 | C = 45; E = 45 | Blood | C = 23; E = 61-88 | ANOVA | No/G,D | A | Yes |
| de la Burde and Choate[15] | 1975 | C = 67; E = 70 | Blood | R = 30-100 | $\chi^2$ | No/D,E,G | C | Yes |
| Landrigan et al[16] | 1975 | C = 78; E = 46 | Blood | C<40; E = 40-68 | *t* test | No/D,E | . . . | Yes |
| McNeil et al[17] | 1975 | C = 37; E = 101 | Blood | C = 29; E = 59 | *t* test | No/D | E, G, H | No |
| Yamins[18] | 1976 | 80 | Blood | PbB$\bar{x}$ = 33.2 = 9.1 | Multiple regression | No/E | A, B | Yes |
| Kotok et al[19] | 1977 | C = 36; E = 24 | Blood | C = 38; E = 81 | *t* test | No/G,D | A, B | No |
| Ratcliffe[20] | 1977 | C = 23; E = 24 | Blood | C = 28; E = 44 | *t* test | No/E | A, B | No |
| Needleman et al[21] | 1979 | C = 100; E = 58 | Tooth | PbC = 24; PbE = 36; PbTC≤10; PbTE≥20 | ANCOVA | No/H | . . . | Yes |
| Yule et al[22] | 1981 | 166 | Blood | C≤13; E = 13-32 | Multiple regression | Yes | . . . | Yes |
| Winneke et al[23] | 1982 | C = 26; E = 26 | Tooth | PbTC$\bar{x}$ = 2.4; PbTE$\bar{x}$ = 9 | *t* test | No/D | A | No |
| McBride et al[24] | 1982 | 108 | Blood | C = 2-9; E = 19-29 | ANOVA | No/D,E | . . . | No |
| Smith et al[25] | 1983 | 402 | Tooth | PbT$\bar{x}$ = 5.1 = 2.8 | ANCOVA | No/H | . . . | No |
| Winneke et al[26] | 1983 | 115 | Tooth | PbT$\bar{x}$ = 6.2; PbB$\bar{x}$ = 14 | Multiple regression | Yes | . . . | No |
| Harvey et al[27] | 1984 | 48 | Blood | R = 6.2-26.8 | Multiple regression | No/F | A | No |
| Shapiro and Marecek[28] | 1984 | 193 | Tooth | R = 30-150 | Multiple regression | No/E | . . . | Yes |
| Needleman et al[29] | 1985 | 218 | Tooth | PbT$\bar{x}$ = 12.7 | Multiple regression | Yes | . . . | Yes |
| Ernhart et al[30] | 1985 | 80 | Blood | C = 30; E = 40-70 | Multiple regression | Yes | A | No |
| Schroeder et al[31] | 1985 | 104 | Blood | Median = 30 | Multiple regression | Yes | . . . | Yes |
| Hawk et al[32] | 1986 | 75 | Blood | PbB$\bar{x}$ = 21; R = 6-47 | Multiple regression | Yes | A | Yes |
| Lansdown et al[33] | 1986 | C = 80; E = 80 | Blood | C = 7-12; E = 13-24 | Multiple regression | Yes | . . . | No |
| Hatzakis et al[34] | 1987 | 509 | Blood | PbB$\bar{x}$ = 23; R = 7-63 | Multiple regression | Yes | . . . | Yes |
| Pocock et al[35] | 1987 | 402 | Tooth | PbT$\bar{x}$ = 5.1 = 2.8 | Multiple regression | Yes | . . . | Yes |
| Fergusson et al[36] | 1987 | 724 | Tooth | PbT$\bar{x}$ = 6.2 = 3.8 | Multiple regression | Yes | . .. | No |
| Fulton et al[37] | 1987 | 501 | Blood | GM = 11.5; R = 3-34 | Multiple regression | Yes | . . . | Yes |
| Hansen et al[38] | 1987 | 156 | Tooth | PbT$\bar{x}$ = 10.7; PbB$\bar{x}$ = 5 | Multiple regression | Yes | . . . | Yes |

*A indicates small sample; B, weak outcome measures; C, poor exposure measures; D, inadequate data analysis or reporting; E, inadequate or no covariate control; F, overcontrol; G, clinical levels of lead exposure (blood lead level >3.86 μmol/L); H, later reanalysis substituted (Needleman et al[29] [1985] for Needleman et al[21] [1979], Pocock et al[35] [1987] for Smith et al[25] [1983], and Ernhart et al[30] [1985] for Perino and Ernhart[13] [1974]; PbT$\bar{x}$, mean tooth lead value; PbB$\bar{x}$, mean blood lead value; R, range; PbTC, values for control group; PbTE, values for high-lead group; GM, geometric mean; ANOVA, analysis of variance; and ANCOVA, analysis of covariance.

†All tooth studies are measured in parts per million and all blood studies are measured in micrograms per deciliter.

## METHODS

### Data Collection

All studies on lead exposure and children's neurobehavioral development that were published since 1972 were examined for eligibility. The sources of candidate studies were a computerized MEDLINE subject search and a search of programs of meetings on metals, neurotoxicology, lead, pediatrics, and public health. Dissertation abstracts were also searched. Table 1 lists the studies identified in the search[12-38] and presents summary data.

Studies were excluded for the following reasons: (1) Inadequate control of covariates reflecting socioeconomic and familial factors.[12,16-20,24,28] (2) Overcontrol of factors that reflect exposure to the independent variable, lead. One study[27] controlled for pica and peeling paint. (3) Inclusion of subjects with defined clinical lead poisoning (ie, blood lead levels >3.9 μmol/L).[12,14,19] (4) Reported data either did not permit any further quantification[15] or did not enable us to calculate the coefficient of lead in a multiple regression model.[12,14,16,17,23,24]

Some studies were excluded on the basis of more than one of the above criteria. The first criterion effectively excludes most of the early studies in this area since these simply compared high- and low-lead groups, with limited or no control for relevant covariates. The second criterion was selected to avoid overcontrol. The one study[27] that was excluded on this basis also involved a very small sample (multiple regression with 17 covariates and complete data on 48 subjects).

Two of the studies[21,25] originally analyzed the data by dichotomizing lead exposure. The data were later reanalyzed by regression, treating exposure as a continuous variable. We used the results reported in the reanalyses. Supplementary information about the regression analysis was obtained from the authors of two studies.[26,38]

### Data Analysis

To achieve an acceptable level of homogeneity, the studies were divided into two groups according to the type of tissue analyzed for lead (blood or tooth). The $P$ values within each group were compared for homogeneity using the technique of Rosenthal,[8(p76)] which is based on the sum of the squared deviations of the $t$ values for lead from the group mean.

Joint $P$ values for lead were calculated for each of the two groups using two different approaches proposed by Fisher and by Mosteller and Bush.[8(p94)] In Fisher's procedure, the logarithm of the product of the individual $P$ values is multiplied by $-2$. The resulting quantity has a $\chi^2$ distribution with 2N *df*. In the procedure by Mosteller and Bush, the weighted sum of the $t$ values of the lead coefficient is computed, with each coefficient being weighted by its *df*. This method effectively weights each study by the number of subjects involved. It is particularly useful in this meta-analysis because of the wide range of sample sizes (75 to 724).

For each study, the partial correlation coefficient of lead was derived from the corresponding $t$ value and was used as a measure of effect size. These coefficients were transformed to $z$ scores using Fisher's transformation[8(p27)] and were then compared via a $\chi^2$ statistic.[8(p77)] When the hypothesis of homogeneity was not rejected, the values of partial $r$ from each study were treated as independent estimates of a common (group) partial correlation. Weighted $z$ score averages were computed and were used to construct 95% confidence intervals

Table 2.—Studies Included in the Meta-analysis*

| Study | Year | Exposure Measure | Outcome Measure | Publication Status | Subjects' Age, y | Country |
|---|---|---|---|---|---|---|
| Yule et al[22] | 1981 | Blood | WISC-R V, F | Journal | 6-12 | United Kingdom |
| Lansdown et al[33] | 1986 | Blood | WISC-R V, F | Journal | Preschool | United Kingdom |
| Winneke et al[26] | 1983 | Tooth | WISC-R V, F | Journal | 7-12 | Germany |
| Needleman et al[29] | 1985 | Tooth | WISC-R V, F | Journal | 7-8 | United States |
| Ernhart et al[30] | 1985 | Blood | McCarthy Scale | Journal | Preschool | United States |
| Schroeder et al[31] | 1985 | Blood | Bayley/Stanford Binet IQ Scale | Journal | 1-6 | United States |
| Hawk et al[32] | 1986 | Blood | Stanford Binet IQ Scale | Journal | 3-7 | United States |
| Fergusson et al[36] | 1987 | Tooth | WISC-R V, F | Journal | 8-9 | New Zealand |
| Fulton et al[37] | 1987 | Blood | British Ability Scale C | Journal | 6-9 | United Kingdom |
| Hatzakis et al[34] | 1987 | Blood | WISC-R V, F | PROC | 7-12 | Greece |
| Pocock et al[35] | 1987 | Tooth | WISC-R F | Journal | 6 | United Kingdom |
| Hansen et al[38] | 1987 | Tooth | WISC-R V, F | PROC | 7-8 | Denmark |

*WISC-R indicates Wechsler Intelligence Scale for Children–Revised; V, verbal; F, full-scale; and PROC, proceedings of meeting.

Table 3.—Covariates Entered Into the Final Multiple Regression Model*

| Study‡ | SES | Parental Factors | Perinatal Factors | Physical Factors | Gender | Parent IQ | Parental Rearing | Lead Coefficients† | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Unadjusted | Final Model |
| Yule et al[22] (2) | * | ... | ... | Age | ... | ... | ... | NA | −8.08 (4.63) |
| Lansdown et al[33] (2) | * | ... | ... | Age | ... | ... | ... | NA | 2.15 (4.48) |
| Winneke et al[26] (52) | * | ... | * | * | * | ... | ... | NA | −0.125 (466) |
| Needleman et al[29] (5) | * | * | ... | ... | ... | * | ... | NA | −0.21 (0.07) |
| Ernhart et al[30] (3) | ... | * | * | Age | ... | * | ... | NA | NA |
| Schroeder et al[31] (7) | * | ... | ... | ... | ... | ... | ... | NA | −0.199 (0.07) |
| Hawk et al[32] (1) | * | ... | ... | ... | * | * | * | −0.456 | −0.255 (0.15) |
| Fergusson et al[36] (7) | * | * | * | ... | * | ... | ... | NA | −1.46 (1.25) |
| Fulton et al[37] (21)§ | * | * | ... | * | * | * | * | −5.45 (1.5) | −3.70 (1.31) |
| Hatzakis et al[34] (10) | * | * | ... | ... | ... | * | * | −0.376 | −0.266 (0.07) |
| Smith et al[25] (18) | * | * | * | ... | * | * | * | −2.66 (0.86) | −0.77 (0.63) |
| Hansen et al[38] (6) | * | ... | * | ... | ... | ... | ... | NA | −4.27 (1.21) |

*Asterisk (*) indicates those factors entered into the model; and SES, socioeconomic status.
†NA indicates not available. Where available, coefficients for lead are given for the unadjusted bivariate model and the final multivariate model.
‡The number of coefficients entered into the initial model is in parentheses.
§The SE of the coefficients was estimated from the data.

for the group partial correlation coefficient.[7(p227)]

Power for each study to find a "small" effect was computed using the method (and program) described in Gatsonis and Sampson.[39] We used the definition by Cohen[40] of a "small" effect (partial $r = .14$).

Finally, to assess whether the exclusion of 12 of the original 24 studies had a biasing effect on our conclusions, we used Fisher's aggregation technique in an analysis that included all 24 studies. For most of the early studies, $P$ values were either given in the published reports or derived on the basis of the published data. In the few cases where a $P$ value was not available, we followed a conservative approach and assumed it was .5.

## RESULTS

All studies considered and reasons for exclusion are listed in Table 1. Of the 12 excluded studies, 5 reported an effect significant at the .05 level and 7 did not. Twelve studies were included in the meta-analysis; 7 of them measured exposure by blood lead and 5 by tooth lead values (Table 2). The two groups were analyzed separately. In 11 of the 12 studies reviewed, the $t$ value of the regression coefficient for lead was negative, ranging from −3.86 to 0.48 in the blood lead group and from −3.0 to −0.03 in the tooth lead group. The partial correlation coefficient of lead ranged from −.27 to .05 and from −.2 to −.003, respectively, for the two groups.

The dependent variable (IQ) was measured by the Wechsler Intelligence Scale for Children–Revised in eight studies. Two studies employed the Stanford Binet IQ Scale, one employed the British Ability Scale, and one employed the McCarthy Scale. The comparison of the distributions of lead exposure was hindered by two difficulties: (1) methods for measuring lead level differed, particularly in the tooth lead group, and (2) summary descriptions of the distribution of lead exposure also differed. In the blood lead group, the lead exposure in the study by Landsdown et al[33] (mean, 0.62 μmol/L) was among the lowest, while the exposure in the study by Schroeder et al[31] (median, 1.46 μmol/L) was among the highest. In the tooth lead group, where analytic methods were different, the lead exposure in the study by Smith et al[25] was among the lowest (248 of 402 children had tooth lead concentration <5.5 ppm), while the exposure in the study by Needleman et al[21] (mean, 12.7 ppm) was among the highest. The sets of covariates included in the regression equations differed for each study, although most covariates purported to measure factors that were similar across studies. It is impractical to present herein a detailed list of the covariates for each study. A condensed form of this information is in Table 3, in which we classified the various covariates into groups on the basis of seven factors. Where available, the unadjusted coefficient of lead is also included in Table 3, along with the coefficient of lead in the final model. In some studies the logarithm of the lead measurement was used in the regression equations.

The $P$ values for the common direc-

Table 4.—Results of Synthesis of 12 Studies

| Study | Weighted t Values: z | Weighted t Values: P (One-Sided) | Fisher's Technique: $\chi^2$ | Fisher's Technique: P |
|---|---|---|---|---|
| **Blood Lead Studies** | | | | |
| All studies | −5.46 | <.0001 | 61.29 | <.0001 |
| Eliminating one study at a time (study eliminated) | | | | |
| Hatzakis et al[34] | −3.88 | <.0001 | 42.87 | <.0001 |
| Hawk et al[32] | −5.34 | <.0001 | 55.3 | <.0001 |
| Schroeder et al[31] | −5.15 | <.0001 | 49.68 | <.0001 |
| Fulton et al[37] | −4.87 | <.0001 | 49.68 | <.0001 |
| Yule et al[22] | −5.25 | <.0001 | 54.86 | <.0001 |
| Lansdown et al[33] | −5.56 | <.0001 | 60.52 | <.0001 |
| Emhart et al[30] | −5.31 | <.0001 | 54.86 | <.0001 |
| Combining studies using log-transformed values (Fulton et al,[37] Yule et al,[22] and Lansdown et al[33]) | ... | ... | 18.83 | .005 |
| **Tooth Lead Studies** | | | | |
| All studies | −2.65 | .004 | 33.11 | <.0005 |
| Eliminating one study at a time (study eliminated) | | | | |
| Needleman et al[29] | −1.97 | .024 | 19.29 | <.025 |
| Hansen et al[38] | −2.3 | .011 | 23.9 | <.005 |
| Winneke et al[26] | −2.67 | .004 | 31.68 | <.0005 |
| Smith et al[25] | −2.36 | .009 | 28.69 | <.0005 |
| Fergusson et al[36] | −3.04 | .001 | 28.88 | <.0005 |
| Combining studies using log-transformed values (Smith et al[25] and Fergusson et al[36]) | −1.61 | .001 | 8.66 | <.0005 |

Table 5.—Lead Coefficients for Full-scale IQ Scores*

| Study | Coefficient | SE | t | P (One-Sided) | Sample Size | Partial r | Total $R^2$ |
|---|---|---|---|---|---|---|---|
| **Blood Lead Studies** | | | | | | | |
| Hatzakis et al[34] | −0.27 | 0.07† | −3.86† | .0001 | 509 | −.17 | 0.25 |
| Hawk et al[32] | −0.25 | 0.15 | −1.67 | .05 | 75 | −.20 | 0.21 |
| Schroeder et al[31] | −0.2 | 0.07† | −2.78 | .003 | 104 | −.27 | NA |
| Fulton et al[37]‡ | −3.7 | 1.37 | −2.77 | .003 | 501 | −.12 | 0.46 |
| Yule et al[22]‡ | −8.08 | 4.63 | −1.75 | .04 | 129 | −.16 | NA |
| Lansdown et al[33]‡ | 2.15 | 4.48† | 0.48 | .68 | 86 | .05 | NA |
| Emhart et al[30] | NA | NA | −1.8† | .04 | 80 | −.20 | NA |
| (Average weighted partial r = −.152; 95% confidence interval, −.2 to −.1) | | | | | | | |
| **Tooth Lead Studies** | | | | | | | |
| Needleman et al[29] | −0.21 | 0.07 | −3 | .001 | 218 | −.20 | 0.35 |
| Hansen et al[38] | −4.27 | 1.91 | −2.23§ | .01 | 156 | −.18 | 0.2 |
| Winneke et al[26] | −0.13 | 4.66 | −0.03§ | .49 | 115 | −.003 | 0.13 |
| Pocock et al[35]‡ | −0.77 | 0.63 | −1.22 | .11 | 388 | −.06 | NA |
| Fergusson et al[36]‡ | −1.46 | 1.25 | −1.17 | .12 | 724 | −.04 | NA |
| Average weighted partial r = −.08; 95% confidence interval, −.13 to −.03) | | | | | | | |

*NA indicates not available.
†Estimated from data in article.
‡Log transforms.
§Obtained from the author.

tional hypothesis that lead is negatively correlated with IQ were tabulated. Before combining the probabilities, the homogeneity of the $P$ values was assessed. The $\chi^2$ statistics were 11.02 ($df=6$, $P=.09$) and 5.13 ($df=4$, $P=.26$) for the blood lead and tooth lead group, respectively. Thus, the hypothesis of homogeneity cannot be rejected for either group.

Combined $P$ values in the blood lead group were less than .0001 for both methods of combining probabilities. The corresponding combined $P$ values for the tooth lead group were less than .0005 and .004, respectively.

## Sensitivity Analysis

The sensitivity of the findings was examined by removing the studies one by one from the analysis and recalculating combined $P$ values (Table 4). For the tooth lead group the highest combined $P$ value was .025 and the lowest was .0001. The corresponding figures for the blood lead group were below .0001. The overall finding of a significant lead effect is supported by both methods of combining the data. No single study seems to be responsible for the significance of the final finding.

## Effect Size

In the case of multiple regression/correlation studies, the usual measure of effect size is the partial correlation coefficient (partial $r$).[7,8,40] Derived partial $r$'s for the 12 studies under review are given in Table 5.

Each partial $r$ was converted to a $z$ score using Fisher's $z$ transform. The $\chi^2$ statistics for homogeneity were 5.78 ($df=6$, $P>.4$) for the blood lead group and 6.44 ($df=4$, $P>.1$) for the tooth lead group. The hypothesis of homogeneity of the effect sizes cannot be rejected for either of the two groups. The weighted $z$ score averages were −.152 (SE = .027) and −.08 (SE = .025), respectively. In the original scale, the approximate 95% confidence intervals for the group partial $r$ were $-.15 \pm .05$ for the blood lead group and $-.08 \pm .05$ for the tooth lead group.

The results of the analysis in terms of the partial $r$'s support those obtained from the analysis of the $P$ values. Neither approach provides an overall estimate of the raw effect size, ie, of the average change in IQ units per unit change in lead exposure. A meaningful attempt to arrive at such an overall estimate is precluded by the substantial differences in model specification among the studies, as well as in units and methods of measuring lead exposure and outcome.

## Selection Bias and the File Drawer Problem

There were two basic steps in the selection of studies for this meta-analysis: (1) the retrieval of studies and (2) the formulation and application of exclusion criteria to the retrieved studies. The possibility of bias in both steps was investigated. In particular with respect to the second step, calculations with all the original 24 studies included showed that Fisher's statistic was 93.8 ($df=34$, $P<.0001$) for the blood lead group, 42.5 ($df=14$, $P<.001$) for the tooth lead group, and 136.4 ($df=48$, $P<.0001$) for all studies together. This is evidence that the application of the exclusion criteria was not an important source of bias in this meta-analysis.

The possibility of bias resulting from the first step has been termed the *file drawer problem.*[8(p107)] Such bias may result from at least two sources (beyond faults in the retrieval process): the failure of all investigators to report their results or the failure of journals to publish all results submitted. Studies that show a statistically significant result do tend to be published more frequently.

We estimated the magnitude of the file drawer problem by calculating the number of unpublished nonsignificant

Table 6.—Power Calculations for "Small" Effects of Lead ($\alpha = .05$; Partial $r = .14$)

| Study | Sample Size | No. of Covariates (Final) | Power |
|---|---|---|---|
| **Blood Lead Studies** | | | |
| Fulton et al[37] | 501 | 14 | 0.87 |
| Hatzakis et al[34] | 509 | 8 | 0.88 |
| Hawk et al[32] | 75 | 5 | 0.21 |
| Schroeder et al[31] | 104 | 7 | 0.28 |
| Yule et al[22] | 129 | 2 | 0.35 |
| Lansdown et al[33] | 86 | 2 | 0.25 |
| Ernhart et al[30] | 45 | 4 | 0.23 |
| **Tooth Lead Studies** | | | |
| Needleman et al[29] | 218 | 5 | 0.53 |
| Fergusson et al[36] | 724 | 8 | 0.96 |
| Smith et al[25] | 388 | 10 | 0.78 |
| Winneke et al[26] | 115 | 4 | 0.31 |
| Hansen et al[38] | 156 | 7 | 0.40 |

studies that would be necessary to bring the overall $P$ value to greater than .05. Using the procedure of Rosenthal,[8(p108)] we found that 26 null result studies would be necessary to dilute the finding for the tooth lead group and that 67 would be necessary for the blood lead group. This procedure assumes that the mean $z$ score of the unseen studies is 0. A more stringent procedure is suggested by Iyengar and Greenhouse,[41] which assumes that all unseen studies simply are not significant at the .05 level. Under this assumption, it would require 16 and 35 studies to dilute the finding for the tooth lead group and the blood lead group, respectively. Given the expense of conducting human studies of lead exposure and the amount of attention directed to this question, it is unlikely that this number of negative studies have escaped notice.

## Power Calculations

The studies included in this meta-analysis are observational. The values of the covariates cannot be fixed in advance by design but are themselves outcomes of the study. Any calculation of power must account for this extra variability.[39] Table 6 presents the a priori power of each study to detect a partial $r$ of .14 (denoted as a "small" effect[40]). "Small" in this sense does not mean biologically unimportant, it means difficult to identify. Cohen[40] has pointed out that a result of this size "all too frequently in practice represents the true order of magnitude of the effects being tested." As can be seen from Table 5, a partial $r$ equal to $-.14$ is near the center of the values for the partial correlation coefficient that were derived from the studies under review. Of the 12 studies, 8 had power below 60% to detect an effect of this magnitude.

The power figures given here are optimistic: they are calculated on the number of covariates present in the final model reported in each study. Most studies, however, initially controlled for many more covariates than those in the final model. As few articles gave information about missing values in the data, it is possible that some of the sample sizes used to calculate power are larger than the effective sample sizes of the studies.

## Some Methodological Issues

The inclusion criteria ensured that the studies analyzed provided an acceptable minimum control for relevant covariates. In two studies,[22,33] control was done only for social class. Multiple regression analysis was employed in all studies, usually in stepwise form. No study reported any analysis of residuals, model checking, and detection of possible outliers in the data. Only two studies attempted to select an "optimal" regression model in a formal way. No study addressed the issue of errors in measurement of the independent variables. The question of errors in variables is particularly relevant when measuring exposure at low levels. Other covariates that represent arbitrary constructs (eg, marital relationships, parental interest, parental involvement in school, and so on) are also particularly vulnerable to errors-in-variables problems.

## COMMENT

The overall evidence from our meta-analysis establishes a strong link between low-dose lead exposure and intellectual deficit in children. A natural question that arises at this point is whether the link is a causal one. The answer to this question goes beyond the formal meta-analytic method. Some of the epistemological issues encountered in making causal inferences are discussed below.

The effects of lead on the central nervous system are embedded in a complex process involving biologic, environmental, familial, and socioeconomic factors. Epidemiologic studies cannot, by themselves, establish causal relationships. Causality is not subject to empirical proof, whether in the field or in the laboratory.[42] Given that direct demonstration of proof of a low-dose lead effect in a naturalistic setting is not achievable, epidemiologists rely on canons[43] that, if satisfied, permit the conservative drawing of causal inferences. They are (1) time precedence of the putative cause, (2) biologic plausibility, (3) nonspuriousness, and (4) consistency.

Cross-sectional studies such as those reviewed herein cannot establish the time precedence of lead exposure; the level of lead was measured at the same time as IQ. The claim has been made that neurobehavioral deficits result in excess lead intake, ie, deficient children mouth more leaded substances. This assertion has been effectively refuted by forward studies of lead exposure beginning at birth. These studies have shown a clear relationship between umbilical cord blood lead levels and later development at 6 to 24 months.[44-46]

Biologic plausibility demands that mechanisms at a lower biologic level have been demonstrated to explain the phenomenon under examination. Lead is a thoroughly investigated neurotoxin.[47] Among many effects that have been demonstrated, lead has been shown to affect neurotransmitter activity, brain adenyl cyclase activity, and dendritic complexity.[48-50] Demonstration of dose-response relationships strengthens the plausibility of the relationships studied. Convincing demonstrations of dose-related behavioral effects have been made in animal studies.[51,52] Epidemiologic demonstrations of association between dose (blood or tooth lead levels) and response (teachers' ratings of classroom behavior and reaction time under varying intervals of delay) also have been published.[21,34]

Nonspuriousness means that the relationship put forth in the causal claim is not due to a confounder or a set of confounders. Complete confounder control is impossible in real world studies. In most studies reviewed, control of confounders has reduced the magnitude of the lead-IQ effect but has not obliterated it. The argument for nonspuriousness is further strengthened by the evidence provided by animal studies in rodents and subhuman primates, which produced cognate outcomes in cross-fostered litter mates.[51,52]

Finally, consistency requires that the phenomenon be demonstrated in different studies under similar but not identical circumstances. The statistical nature of these investigations requires an extended notion of consistency. Even if the effect under study exists in nature, the $P$ values and effect sizes reported in investigations of the question will vary in magnitude, and not all studies will give a significant result.

A different type of evidence for consistency was offered in the study by Wallsten and Whitfield.[53] This evidence is based on the probabilistically encoded opinions of six lead experts of widely ranging viewpoints about the dose-response relation between lead exposure and IQ. Five of the six experts' estimates of the dose-response curve were convergent, leading the authors to state: "In view of the extensive debate concerning the effects of lead on IQ, the degree of consensus reflected in the

study's results is notable, especially since the experts were selected so as to span the full range of opinion."

The four previously cited reviews of the studies of lead at low dose differed in their evaluation of essentially the same evidence. One review came to a qualified negative conclusion,[5] one came to a positive conclusion,[4] and two found the evidence inconclusive.[2,3] This difference of opinion partly proceeds from a limitation inherent in the method of narrative reviewing; it essentially evaluates each study in isolation and is unable to achieve a systematic synthesis. Meta-analysis avoids this limitation and includes all studies in a joint inference. Using this method, and incorporating into our review a number of recent studies that were not available to the earlier reviewers, we found that although the sample of studies varied widely in their individual power to find an effect, and not all found an effect by the conventional rule of $P<.05$, 11 of 12 studies reviewed reported a negative coefficient for lead. The joint probability of the findings reported occurring by chance under the null hypothesis was quite small, and this was not materially influenced by any single study. The estimated effect sizes for the two groups were both significantly different from zero. These findings, taken in sum, permit a strong inference that low-dose lead exposure is causally associated with deficits in psychometric intelligence.

Dr Gatsonis was supported by grant MH 15758 from the National Institute of Mental Health, Bethesda, Md.

We thank Joel Greenhouse, PhD, Satish Iyengar, PhD, and David Bellinger, PhD, for critical review of the manuscript.

**References**

1. Byers RK, Lord E. Late effects of lead poisoning on mental development. *AJDC*. 1943;66:471-494.
2. Rutter M. Raised lead levels and impaired cognitive/behavioral functioning: a review of the evidence. *Dev Med Child Neurol*. 1980;22(suppl 42):1-26.
3. Bornschein R, Pearson D, Reiter L. Behavioral effects of moderate lead exposure in children and animal models. *CRC Crit Rev Toxicol*. 1980;8:43-99.
4. Needleman HL, Bellinger D. The developmental consequences of childhood exposure to lead: recent studies and methodological issues. In: Lahey BB, Kazdin AE, eds. *Advances in Clinical Child Psychology*. New York, NY: Plenum Publishing Corp; 1984;7:195-220.
5. Pocock SJ, Ashby D. Environmental lead and children's intelligence: a review of recent epidemiological studies. *The Statistician*. 1985;35:31-44.
6. Glass GV, McGaw B, Smith ML. *Meta-analysis in Social Research*. Beverly Hills, Calif: Sage Publications Inc; 1981.
7. Hedges LV, Olkin I. *Statistical Methods for Meta-analysis*. Orlando, Fla: Academic Press Inc; 1985.
8. Rosenthal R. *Meta-analytic Procedures for Social Research*. Beverly Hills, Calif: Sage Publications Inc; 1984.
9. Sacks HS, Berrier J, Reitman D, Ancoma-Berk VA, Chalmers TC. Meta-analyses of randomized control trials. *N Engl J Med*. 1987;316:450-455.
10. Schwartz J, Pitcher H, Levin R, Ostro B, Nicholas AL. *Cost and Benefits of Reducing Lead in Gasoline: Final Regulatory Impact Analysis*. Washington, DC: Office of Policy Analysis, Environmental Protection Agency; 1985.
11. Needleman HL, Bellinger D. Type II errors in studies of low level lead exposure: a critical and quantitative review. In: Smith M, Grant L, Sors A, eds. *Lead Exposure and Child Development: An International Assessment*. Lancaster, England: MTP Press Ltd; 1988.
12. Kotok D. Development of children with elevated blood levels: a controlled study. *J Pediatr*. 1972;80:57-61.
13. Perino J, Ernhart C. The relation of subclinical lead level to cognitive and sensory impairment in black preschoolers. *J Learn Disab*. 1974;7:26-30.
14. Rummo JH, Routh DR, Rummo NJ, Brown JF. Behavioral and neurological symptomatic and asymptomatic lead exposure in children. *Arch Environ Health*. 1979;34:120-124.
15. de la Burde B, Choate MS. Early asymptomatic lead exposures and development at school age. *J Pediatr*. 1975;87:638-664.
16. Landrigan P, Baloh R, Whitworth R, Staeling N, Rosenblum B. Neuropsychological dysfunction in children with chronic low level lead absorption. *Lancet*. 1975;1:708.
17. McNeil J, Ptasnik J, Croft D. Evaluation of long-term effects of elevated blood lead concentrations in asymptomatic children. *Arch Ind Hygiene Toxicol*. 1975;26:97-118.
18. Yamins JG. *The Relationship of Subclinical Lead Intoxication to Cognitive and Language Function in Preschool Children*. Long Island, NY: Hofstra University; 1976. Thesis.
19. Kotok D, Kotok R, Heriot T. Cognitive evaluation of children with elevated blood levels. *AJDC*. 1977;131:791-793.
20. Ratcliffe JM. Developmental and behavioral functions in young children with elevated blood lead levels. *Br J Prev Soc Med*. 1977;31:258-264.
21. Needleman H, Gunnoe C, Leviton A, et al. Deficits in psychologic and classroom performance in children with elevated dentine lead levels. *N Engl J Med*. 1979;300:584-695.
22. Yule W, Lansdown R, Millar I, Urbanowicz M. The relationship between blood lead concentration, intelligence, and attainment in a school population: a pilot study. *Dev Med Child Neurol*. 1981;23:567-576.
23. Winneke G, Kramer G, Brockhaus A. Neuropsychological studies in children with elevated tooth lead concentration, I: pilot study. *Int Arch Occup Environ Health*. 1982;51:169-183.
24. McBride WG, Black BP, English BJ. Blood lead levels and behaviour of 400 preschool children. *Med J Aust*. 1982;2:26-29.
25. Smith M, Delves T, Lansdown R, Clayton B, Graham P. The effects of lead exposure on urban children: The Institute of Child Health/Southampton Study. *Dev Med Child Neurol*. 1983;25(suppl 47):1-54.
26. Winneke G, Kramer G, Brockhaus A, et al. Neuropsychological studies in children with elevated tooth lead concentration. *Int Arch Occup Environ Health*. 1983;51:231-252.
27. Harvey PG, Hamlin MW, Kumar R, Delves HT. Blood lead, behaviour and intelligence test performance in preschool children. *Sci Total Environ*. 1984;40:45-60.
28. Shapiro IM, Marecek J. Dentine lead concentration as a predictor of neuropsychological functioning in inner-city children. *Biol Trace Elem Res*. 1984;6:69-78.
29. Needleman HL, Geiger SK, Frank R. Lead and IQ scores: a reanalysis. *Science*. 1985;227:701-704.
30. Ernhart CB, Landa B, Wolf AW. Subclinical lead level and developmental deficit: re-analyses of data. *J Learn Disab*. 1985;18:475-479.
31. Schroeder SR, Hawk B, Oto DA, Mushak P, Hicks RE. Separating the effects of lead and social factors on IQ. *Environ Res*. 1985;91:178-183.
32. Hawk BA, Schroeder SR, Robinson G, et al. Relation of lead and social factors to IQ of low SES children: a partial replication. *Am J Ment Def*. 1986;91:178-183.
33. Lansdown R, Yule W, Urbanowicz MA, Hunter J. The relationship between blood-lead concentrations, intelligence attainment and behavior in a school population: the second London study. *Int Arch Occup Environ Health*. 1986;57:225-235.
34. Hatzakis A, Kokevi A, Katsouyanni K, et al. Psychometric intelligence and attentional performance deficits in lead exposed children. In: *International Conference on Heavy Metals in the Environment*. Edinburgh, Scotland: CEP Consultants; 1987:204.
35. Pocock SJ, Ashby D, Smith V. Lead exposure and children's intelligence. *Int J Epidemiol*. 1987;16:57-67.
36. Fergusson DM, Fergusson JE, Harwood LJ, Kinzett NG. A longitudinal study of dentine lead levels, intelligence, school performance and behavior, III: dentine lead levels and cognitive ability. *J Child Psychol Psychiatry*. 1988;29:793-809.
37. Fulton M, Raab G, Thomson G, Laxen D, Hunter R, Hepburn W. Influence of blood lead on the ability and attainment of children in Edinburgh. *Lancet*. 1987;1:1221-1226.
38. Hansen ON, Trillingsgard A, Beese I, Lyngbye T, Grandjean P. A neuropsychological study children with elevated dentine lead levels. In: *International Conference on Heavy Metals in the Environment*. Edinburgh, Scotland: CEP Consultants; 1987:54-56.
39. Gatsonis C, Sampson A. Multiple correlation: exact power and sample size calculations. *Psych Bull*. 1989;106:516-524.
40. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Orlando, Fla: Academic Press Inc; 1977.
41. Iyengar S, Greenhouse J. Selection models and the file-drawer problem. *Stat Sci*. 1988;3:109-135.
42. Rothman KJ. *Modern Epidemiology*. Boston, Mass: Little Brown & Co Inc; 1986.
43. Kenny DA. *Correlation and Causality*. New York, NY: John Wiley & Sons Inc; 1979.
44. Bellinger D, Leviton A, Waternaux C, Needleman H, Rabinowitz M. Longitudinal analyses of prenatal lead exposure and early cognitive development. *N Engl J Med*. 1987;316:1037-1043.
45. Deitrich K, Krafft K, Beir M, et al. Early effects of fetal lead exposure: developmental findings at 6 months. In: *International Conference on Heavy Metals in the Environment*. Edinburgh, Scotland: CEP Consultants; 1987:63-65
46. McMichael AJ, Baghurst PA, Wigg NR, et al. Port Pirie cohort study: environmental exposure to lead and children's abilities at four years. *N Engl J Med*. 1988;319:468-475.
47. *Air Quality Criteria for Lead*. Research Triangle Park, NC: Environmental Protection Agency; 1986. Publication EPA/600/8-83/028af.
48. Hrdina PD, Hanin I, Dubas TC. Neurochemical correlates of lead toxicity. In: Singhal RL, Thomas JA, eds. *Lead Toxicity*. Baltimore, Md: Urban & Schwarzenberg; 1980:273-300.
49. Nathanson JA. Lead-inhibited adenylate cyclase, a model for the evaluation of chelating agents in the treatment of lead toxicity. *J Pharm Pharmacol*. 1977;29:511-513.
50. Averill D, Needleman HL. Neonatal lead exposure retards cortical synaptogenesis in the rat. In: Needleman H, ed. *Low Level Lead Exposure: The Clinical Implications of Current Research*. New York, NY: Raven Press; 1988.
51. Rice DC, Willes RF. Neonatal low level lead exposure in monkeys *(Macaca fasciculosis)*: effect on two-choice non-spatial form discrimination. *J Environ Pathol Toxicol*. 1979;2:1195-1203.
52. Cory-Schlecta DA, Bissen ST, Young AM, Thompson T. Chronic post-weaning lead exposure and response duration performance. *Toxicol Appl Pharmacol*. 1981;60:78-84.
53. Wallsten TS, Whitfield RG. Assessing the risks to young children of three effects associated with elevated blood-lead levels. Argonne, Ill: Argonne National Laboratory; 1986. Report ANL/AA-32.