Pergamon

Archives of Clinical Neuropsychology
16 (2001) 359–374

Archives
of
CLINICAL
NEUROPSYCHOLOGY

Studies of lead exposure and the developing central nervous system: a reply to Kaufman

Herbert L. Needleman[a,b,*], David Bellinger[c]

[a]*Department of Psychiatry, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA*
[b]*Western Psychiatric Institute and Clinic, Pittsburgh, PA, USA*
[c]*Neuroepidemiology Unit, Department of Neurology, Children's Hospital and Harvard Medical School, Cambridge, MA, USA*

Accepted 28 August 2000

Abstract

Kaufman's critique of the lead-cognition hypothesis is a reiteration of well worn and weary claims raised many times in the past, primarily by spokespersons for the lead industry. They have been fully rebutted in the literature. The essence of these claims is that those studies showing an association between lead and IQ are flawed by uncontrolled confounding, multiple comparisons, and errors in measurement. Any effect of the lead, Kaufman asserts, if present, is small. This response examines each of these issues and shows that they lack substance. Lead's negative impact on IQ persists in most modern studies after confounding has been controlled in many different statistical models. At least three metaanalyses have confirmed an effect of lead at low dose. Animal studies in which lead is given systematically, and the issue of confounding thereby avoided, demonstrate an unequivocal lead effect at similar doses to the human studies. The criticism of multiple comparisons similarly does not withstand examination. Measurement errors are nonsystematic and nondifferential. They are, therefore, null biasing. The actual size of the lead effect has been shown to be substantial, and to be found most prominently at the ends of the distribution. Kaufman says that lead requires study of diverse dimensions of intellect, but he restricts his scope to a sample of studies of lead and IQ, ignoring recent high quality studies that show a clear lead effect, and in those studies that he critiques he ignores data that contradict his position. His article adds nothing to the dialogue on lead neurotoxicity. © 2001 National Academy of Neuropsychology. Published by Elsevier Science Ltd.

* Corresponding author. Suite 310, 3520 Fifth Avenue, Pittsburgh, PA 15213, USA. Tel.: +1-412-383-3100.
E-mail address: hlnlead@pop.pitt.edu (H.L. Needleman).

0887-6177/01/$ – see front matter
PII: S0887-6177(00)00087-1

> In order to have societal impact, lead level would need to be shown conclusively to negatively affect children's functioning in diverse dimensions of intellect.
>
> Alan Kaufman

This is the final sentence in Alan Kaufman's critique of the widely accepted hypothesis that lead damages children's brains at silent doses. Kaufman begins with a paradox: he argues for the need to examine the diverse functions of intellect, but his paper ignores those studies that have done precisely that. His self-titled "Careful Examination of the Literature," apart from its unfortunate implications regarding the quality of others' efforts, restricts itself to only those reports or parts of reports that focus on a single outcome: IQ; and his survey ends with papers published in 1993.

The issue at stake is: Does lead at low doses affect children's brains, and is this evidenced in their behavior? Like an one-eyed astronomer, Kaufman sees nothing outside of the narrow scope of psychometric IQ tests. His universe of discourse is restricted to those studies that were the subject of two meta-analyses. He appears oblivious to the human literature on lead and attention, lead and school failure, lead and aggression, lead and social adjustment and behavior, lead and delinquency, the cognate studies of behavior in lead-exposed primates and rodents, and the vast experimental literature on the neurochemistry of lead. All of these data have important societal implications, all form the framework of the low lead-brain damage hypothesis, and all have been used by regulators in risk analyses to reach important regulatory decisions. But even within the sample of studies he selects, there are data on the effects of lead on social adjustment and behavior. They are ignored.

Epidemiologists have a word for this tactic: selection bias. Apart from this solecism, Kaufman's review contains little that is new; it is one more recital of a set of criticisms raised primarily by spokespersons of the lead industry over the past 20 years. They are:

- Studies of lead and IQ have not adequately controlled for other confounders. Therefore, the effect assigned to lead more properly belongs somewhere else.
- Studies of lead have tested many outcomes but have not adjusted for multiple comparisons.
- Measurements of maternal IQ were inadequate.
- IQ measurements are subject to error.
- The lead–IQ relationship is not linear.
- The size of any lead effect, if present, is small, and its societal consequences dubious.

1. Evolution of knowledge of lead toxicity

Since childhood lead poisoning was first reported over 100 years ago, scientific understanding of its nature and importance has evolved at an accelerating pace. With each major discovery, skeptics rose to question its validity. When it was first reported by pediatricians in Brisbane, Australia in 1892, established physicians disputed the existence of the disease. Once its reality was established, most asserted that there were only two outcomes: death or complete recovery. Children who did not die in the acute toxic phase were said to show no

effects of their illness. This myth persisted until 1943 when it was shown conclusively that children who did not die were left with severe psychological and behavioral problems upon recovery (Byers & Lord, 1943). It was then asserted that neurobehavioral deficits only occurred in children who displayed obvious signs of acute illness (Smith, 1964). In the 1970s, 1980s, and 1990s, studies from around the world refuted this in reports of intellectual and behavioral deficits in lead-exposed children with no overt signs of their exposure (National Academy of Sciences/National Research Council, 1993). Kaufman's paper is the latest in a string of challenges to the reality of central nervous system (CNS) damage and neurobehavioral effects in asymptomatic lead-exposed children, and it echoes the same claims as his predecessors.

There are now approximately 30 published reports of the harmful effects of lead on children's intelligence from investigators around the world. There is no other chemical that owns as sizable a body of investigative data demonstrating its neurotoxic properties.

2. The issue of confounding

Failure to control for confounders is a traditional charge raised against those studies that show a lead effect on intelligence. Kaufman offers a farrago of statements to buttress his argument that lead studies are confounded by other factors, stating, for example, that the Port Pirie study found that breast-fed babies in their study had higher IQ scores; that Hatzakis found a relationship between alcoholic mother and IQ; that Fulton found "noteworthy correlations" between IQ and parent–child interaction; and Lansdown found correlations between blood lead level and thumb-sucking, age at weaning, quarrels/bickering, and family cleanliness.

Kaufman acknowledges that a variable, in order to be classified as a confounder, is required to possess two properties: it must have an impact on the outcome under study and also be correlated with the main risk factor. Then, apparently unfamiliar with the strictures of causal inference, he ignores this definition to indicate that every variable that affects IQ should be entered into the model.

> But it is important to control for variables such as otitis media during infancy that are believed or known to vary alongside an outcome variable (such as IQ) *even if their relationship to blood lead is non-significant or unknown* . . . how can one feel confident that it is blood lead, and not an uncontrolled variable, that is primarily responsible for an observed IQ loss in lead-exposed children. [emphasis added]

Unless children that are exposed to more lead also tend to have more otitis media, it is difficult to see (and Kaufman fails to demonstrate) how a failure to include otitis media in the regression model will create a spurious association between lead and IQ.

Control of demonstrated true confounders is a fundamental part of any study, and most modern studies have reported that after they have done this, the effect of lead endured, although in some cases diminished somewhat in magnitude. The mere presence of a confounder or multiple confounders does not invalidate the role of the main effect (e.g., lead) under study. It becomes the task of investigators to measure the influence of

confounders and to estimate their impact on the main effect. To quote Schlesselman (1978): "... questions of spurious effects can be removed from the realm of 'expert opinion' to the realm of 'plain fact.'"

Kaufman discusses otitis media at some length in this context, stating that only two lead studies controlled for this variable. He is unaware that our group published a paper specifically addressing the association between lead and the risk of otitis media in the subjects of our prospective study (Rabinowitz, Allred, Bellinger, Leviton, & Needleman, 1990). We found that a child's history of otitis media and other infectious disorders was unrelated *to either blood or tooth lead levels*. Thus, otitis media did not meet the criteria for a confounder in this data set and did not require control.

Kaufman's paper focuses on 26 papers, not 26 investigations, and as a consequence, overlooks information that deals with precisely those methodological issues that he criticizes investigators for ignoring. Many investigators published detailed statements on the epidemiological issues under discussion in other papers than those he read. We, for example, published several primarily methodological papers in meeting proceedings (Bellinger, Leviton, Waternaux, & Allred, 1985), book chapters (Bellinger, Leviton, Waternaux, Needleman, & Rabinowitz, 1989), and in methodology journals (Bellinger, Leviton, & Waternaux, 1989; Waternaux, Laird, & Ware, 1989). The first report on our forward study included long tables listing the variables evaluated as potential confounders. In later papers, we limited the reports to those variables included in the final models and made reference to the earlier papers. There is no evidence that he is aware of these contributions.

Kaufman suggests that maternal smoking should be controlled as a confounder. This factor has been shown to influence child development, but where lead is concerned, the issue is considerably more complex than he realizes. Lead arsenate was used as a pesticide in growing tobacco, and many studies have shown that smokers have higher blood lead levels than nonsmokers. In addition, one important effect of smoking is impairment of the mucociliary processes involved in clearing inhaled particulates from the tracheobronchial tree. As a result, absorption of lead-containing particles is enhanced. Smoking, it can be seen, is a risk factor for lead exposure. From this standpoint, controlling for smoking would be analogous to controlling for lead in the paint of the home or the lead content of the formula consumed.

Kaufman continues:

> A more correct statement is the significant increase in prediction is due not only to lead level but also to all other variables, known or unknown, not controlled in the study. As the history of lead–IQ research has revealed, the more pertinent confounding and contaminating variables are controlled, the smaller the relationship between lead level and IQ loss; it is a common finding for significant relationships between BLL and IQ to become much smaller, or to disappear altogether when controlling for a diversity of relevant variables.

This is Kaufman's restatement of the phantom, unmeasured covariate argument: the more covariates in the model, the smaller the effect size. But his reading of our paper, which he deals with in some length, is wrong. In our forward study, because of the sociology and geography of Boston and its environs, increased prenatal lead levels were found in children of more favored family backgrounds. These were children of higher

socioeconomic status who lived in Cambridge, Newton, and Beacon Hill, MA. After adjustment for covariates, the effect size of lead *increased*. This finding, which we think is important to the issue of confounding (see later in Hill's Canons on the role of natural experiments), also escaped his notice.

Because multivariate space is infinite but the number of subjects in a study is necessarily finite, complete control of all confounders is an unattainable goal in real-world epidemiology. The investigator is consequently always confronted with an unsaturated structural model. This means that any number of "best fit" models can be constructed. Each individual study is in effect a model. When a large number of studies report a significant lead effect after control of different sets of covariates, it requires a unique type of cognitive amblyopia to deny the evidence. Of the 27 lead studies reported in one of the meta-analyses he cites (Needleman & Gatsonis, 1990), 15 reported P values $< .05$. Could such a result be due to chance? Yes. And if there were no effect of lead on IQ, what would be the odds of this outcome? Under the null hypothesis, the probability of finding such a distribution of P values can be calculated. It is 2.8×10^{-13}.

Another way to evaluate the threat to validity from confounding is to examine the impact of measurement and specification error on the inferences drawn. This approach permits estimation of the magnitude of mismeasurement of variables including potential confounders, which would be required to invalidate a causal inference. The procedure is a form of sensitivity analysis, and estimates the range over which a given confounder can vary without obviating the main effect under study (Leamer, 1978).

This has, in fact, been done twice: once in the case of the 1979 paper of Needleman et al. (Atkinson, Crocker, & Needleman, 1987), and more recently by Greene and Ernhart (1993) with their own data. Both of these measurement error papers support the robustness and validity of the lead–IQ association.

In the case of the Needleman et al. (1979) paper, the authors used the SEARCH procedure of E. Leamer, and after a rigorous estimation of the bounds of sampling and specification error, come to this conclusion:

> In this Bayesian analysis of the reconstituted Needleman et al. data set, we replace such conjectures by empirical tests. The possibility is seen to be remote that the introduction or omission of other covariates will significantly alter the estimated influence of lead exposure upon children's psychometric intelligence. Similarly, the Bayesian analysis demonstrates that other covariates are likely to affect the estimated influence of lead exposure only if they are quite poorly measured.

Kaufman cites Ernhart as one of the chief critics of the lead–IQ association. In 1974, Ernhart reported that children with elevated blood lead levels had significantly lower McCarthy scores than controls after covariate adjustment, and concluded, "While the effects of subclinical lead intoxication may not be noted in the individual cases seen in a pediatric clinic, analysis of group data indicate quite clearly that performance on an intelligence test is impaired." (Perino & Ernhart, 1974)

In 1981, she recanted her statement and said that if there were any lead effect, it was minimal (Ernhart, Landa, & Schell, 1981). In her latest 1993 paper, she applied different errors in measurement procedure to her study of dentine lead levels and children's IQ. They

found a significant association between lead and verbal IQ that was robust to large errors in covariate measurement. To quote the authors: "The findings of this article relating dentine lead level to IQ near age 5 years are consistent with previous research describing a negative correlation between children's lead levels and intelligence ..." This finding receives the briefest mention by Kaufman, but the authors' conclusion is ignored.

3. The issue of multiple comparisons

Kaufman's comments on this issue are once again colored by his failure to understand the context from which the 26 studies he reviews were drawn.

Examining only our 1992 publication (Bellinger, Stiles, & Needleman, 1992), he cites our forward study as an example of this fallacy. He argues that we placed undue importance on the significant association our group found between children's blood lead levels at 24 months of age and their IQ scores at 10 years of age. In fact, we followed 249 infants from birth to 10 years of life, and measured the association between blood leads at birth, 6, 12, 18, 24, 57, and 120 months and children's scores on various age-appropriate neurobehavioral tests (Bellinger, Leviton, Waternaux, Needleman, & Rabinowitz, 1987; Bellinger et al., 1992). In our 1992 paper, we reported that postnatal blood lead levels were significantly and inversely related to WISC-R IQ and Kaufman-TEA scores at 10 years of age. Earlier, in 1991, we had reported the significant association between 24-month blood lead and general cognitive index scores at 5 years. Kaufman ignores this and claims that we made 21 postnatal IQ comparisons and found only two (blood lead at 24 months) significant. He accuses us of capitalizing on chance, in his words, using "a shotgun approach." In fact, rather than being an isolated finding, the associations we reported in our 1992 paper neatly replicate at 10 years of age the associations we noted when the study cohort was 5 years of age. He also ignores the fact that the association between blood lead and outcome was the inverse for 20 of 21 lead exposure measures.

Nor does he give any weight to the well-known epidemiological and biological characteristics of lead exposure. All students of lead exposure know that 2 years of life is precisely when lead levels peak, and those who study child development know that this is the time of explosive development of cognitive competence, especially in language. The two-sided P value for the association between blood lead and verbal IQ at 24 months was .004. Given the large and growing literature on lead and cognition, we were entitled to strong Bayesian priors about the relationship between early exposure and later IQ. These were shared by the most active investigators of the question.

Instead of incorporating any of these factors into his judgement, Kaufman treats the blood lead measurements as if they were separate studies and prescribes a rote application of a Bonferroni correction, dividing each probability by 21. We are not alone in believing that a Bonferroni or other adjustment is unnecessary or at least needs to be applied with judgement. In recent years, the issue of multiple comparisons and the need to apply adjustments such as the Bonferroni correction has stimulated spirited debate, with some, such as the editor of *Epidemiology* and others, arguing that no adjustments at all are necessary (Poole, 1991; Rothman, 1990).

4. The issue of flawed measurement of maternal intelligence

It is true that the instruments used to measure parental intelligence in many lead studies are neither the most sensitive nor accurate. In what direction does this measurement error operate? There is no evidence that the error is systematic; that is, that it distorts IQ in a nonrandom fashion, or that it is differential, meaning the error is distributed differently between high- and low-lead subjects. A measurement error, which is nonsystematic and nondifferential, is null biasing; it will tend to *underestimate* the effect of lead (Thomas, Stram, & Dwyer, 1993). In this section of his review, Kaufman fails to identify any measurement error in the studies he reviews that are systematic and differential.

5. The issue of inaccurate measurement of children's IQ

Kaufman criticizes our studies and others because the administration of Wechsler instruments "... requires professional, graduate-level training in psychometrics, clinical skills, child development, neuropsychology, and so forth." This is a formidable criterion that most researchers would find difficult to summon the budget and personnel to meet. We agree that it is important to put detailed protocols in place to insure quality control in collecting and scoring psychometric data. However, Kaufman's strictures on training are disputed by a recent statement by the President and Chairman of the Practice Committee of the American Academy of Clinical Neuropsychology (Brandt & van Gorp, 1999). In their statement, they say that using individuals with baccalaureate degrees in psychology are acceptable, provided they receive specific training relevant to the tests to be administered and are supervised by a licensed doctoral level psychologist.

In discussing error in measuring IQ, he states that it could be null biasing. He does not seem to understand that this applies not only to the measurement of IQ but to covariates as well. To reiterate, mismeasurements of covariates, if nonsystematic and nondifferential, will underestimate the effect of lead.

Kaufman claims that "... the quality of the IQ test administrations and the concomitant validity of the obtained IQs are rarely assessed thoroughly or systematically." It would be more accurate to say that the methods used to establish and monitor the quality of these measurements are not fully reported. Again, the space limitations imposed by most journals limit the detail that can be provided. Using our studies once again as an example, we devoted considerable care to the training and monitoring of tester reliability and described our procedures in detail when space was available (e.g., Bellinger, Needleman, Bromfield, & Mintz, 1984). We also described the impact of including dummy variables coding for tester in our regression models. Although it would have been somewhat burdensome for him, it would have been appropriate and rigorous for Kaufman to contact investigators to ask about these issues. Had he done this, he would have been forced to change his blanket indictment that they failed to appreciate the importance of quality control in outcome assessments.

Kaufman criticizes our handling of WISC-R scores, arguing that because the norms were outdated by the late 1980s, we "... should have incorporated the spuriously high WISC-R IQs into [our] interpretation of data." The key finding driving the inferences we drew was

that, within the study sample, blood lead level at 2 years of age was significantly and inversely associated with IQ at age 10 years. Even if all children's scores were biased upward by 5 points, the internal validity of the dose–effect relationship we found would be unaffected. His charge is irrelevant to the inferences we drew.

Kaufman argues that the diversity of settings in which lead studies have been conducted represents a weakness of this literature insofar as the results of each may lack generalizability. Many see this diversity of settings in which the association between lead and child development has been demonstrated as a strength. In other words, the fact that studies conducted in different cultural settings or in different social strata within a given cultural setting tend to converge on a similar dose–effect estimate can be viewed as strong evidence that the association is a robust one. This is especially striking insofar the pattern of confounding of this association by other variables is likely to differ across settings, as Kaufman suggests.

Kaufman laments that restricting a study sample, in the case of Bellinger et al. (1992) to White middle and upper-middle class children, represents a weakness, limiting the generalizability of the results. In epidemiology, restriction of the study sample in such a way that the influence of powerful potential confounders is reduced is a common strategy (Rothman & Poole, 1988). Investigators of the relationship between particulate air pollution and lung function frequently restrict the study sample to non-cigarette smokers, thereby reducing the possibility that the adverse effect of smoking on lung function would overwhelm the lesser impact of air pollution on lung function. A similar line of reasoning motivated our decision to restrict our study sample to relatively advantaged families, thereby dissociating lead exposure from other powerful developmental risk factors at the stage of study design rather than attempting to do so solely by statistical means at the stage of data analysis.

6. The issue of linearity of the lead–IQ relationship

Kaufman comments on the absence of evidence for linearity in the association between children's lead levels and their IQ scores, but neglects to mention the relevant literature. In a paper Kaufman does cite for other reasons, Schwartz (1993) applied a statistical technique that is better suited than multiple regression or ANOVA to evaluating the functional form of a dose–effect relationship, namely nonparametric regression. Applying this method to the data reported by Bellinger et al. (1992), he found no evidence for a threshold. In another paper not cited by Kaufman, Schwartz applied the same method to the data of Needleman et al. (1979) with similar results.

7. The issue of societal effects of lead at low dose

Kaufman takes issue with the thesis that lead at low dose will increase the need for special services. In 1982, in response to numerous statements that the effect of low-level lead exposure on IQ scores was relatively small, we published a short note showing that the shift of IQ scores in our 1979 study took place across the entire distribution of IQ

scores (Needleman, Leviton, & Bellinger, 1982). Because of the sigmoid shape of the cumulative frequency distribution, a shift in median scores of 6 points was associated with a four-fold increase in children with IQ scores below 80, and a truncation of scores at the upper end of the distribution (Fig. 1). This means that 5% of the population at that time were deprived of achieving superior IQ scores (>125) because of their exposure to lead. Kaufman challenges this on the basis that it assumes that the IQ effect, if true, is linear, extends across the range of IQ scores, that IQ is unchangeable, and that IQ is the sole cause of special education placement.

The shift in distribution across the entire range is more than a theoretical concept; it is an empirical fact. We counted actual IQ scores in our study to draw the graph. The need for special placement in relation to lead exposure is also empirically based. We counted children and reported it (Bellinger et al., 1984). In his challenge to the need for special services, Kaufman also ignores one widely recognized outcome we published in the *New England Journal of Medicine* in 1990. Higher exposures to lead in early childhood were associated with a seven-fold increase in high school failure and a six-fold increase in reading disabilities

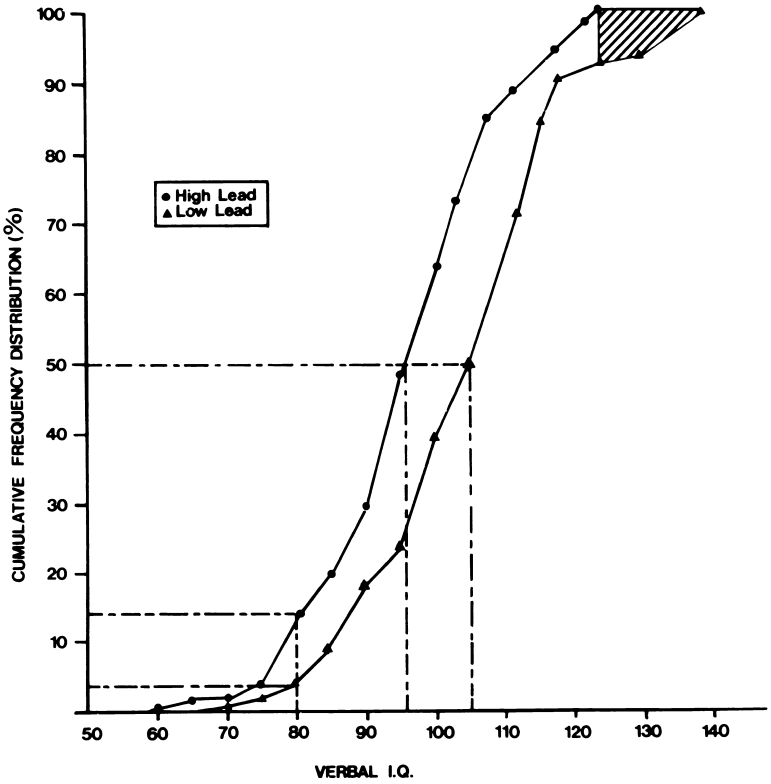

Fig. 1. The real meaning of a small change in mean or median scores. This graph displays the cumulative frequency distribution of verbal IQ scores in high and low lead groups. A relatively small shift in the median (6 points) is associated with a four-fold increase in the rate of severe deficit (IQ < 80). The impact at the upper end of the IQ distribution is noteworthy. Five percent of the low lead children display superior scores (IQ's > 125). Reprinted with permission from *The New Englnd Journal of Medicine, 306*, 367, 1982.

(Needleman, Schell, Bellinger, Leviton, & Allred, 1990). He asserts that IQ is not the only determinant of need for special services while overlooking our published data on two unequivocal outcomes that beg for remedial education: failure to graduate high school and inability to read at grade level. Recent findings by Fergusson, Horwood, and Lynskey (1997) from the Christchurch child development study in New Zealand, a close replication of our 1990 follow-up report not reviewed by Kaufman, provide strong support for these findings, in fact demonstrating that these adverse educational outcomes are apparent at even lower lead doses than we reported.

8. Lead and children's attention/behavior

The brains of children are central in their ability to regulate attention, and the ability to attend is a strong predictor of school and social success. Yet this area, which has been studied in relation to lead in this country and elsewhere, is completely neglected by Kaufman.

Interference with attentional function in lead-exposed children has been demonstrated in a number of studies. Our study of 2146 first- and second-grade children showed a lead-related, dose-dependent increase in poor classroom behavior. This is shown in Fig. 2 reproduced from the *New England Journal of Medicine*, 1979.

Yule, Lansdowne, Urbanowicz, and Miller (1961), using the same teachers' rating scales displayed in Fig. 2, reported that asymptomatic children with elevated blood lead levels had dose-related impairments in behavior. In Yule and Lansdown's (1981) study, high-lead children significantly scored more deviantly on 20 items of the Rutter B2 Scale. Lead was also related to scores on every item of the Connors' Scales, particularly the Conduct Disorder and also the Taylor–Sandford Hyperactivity Scales.

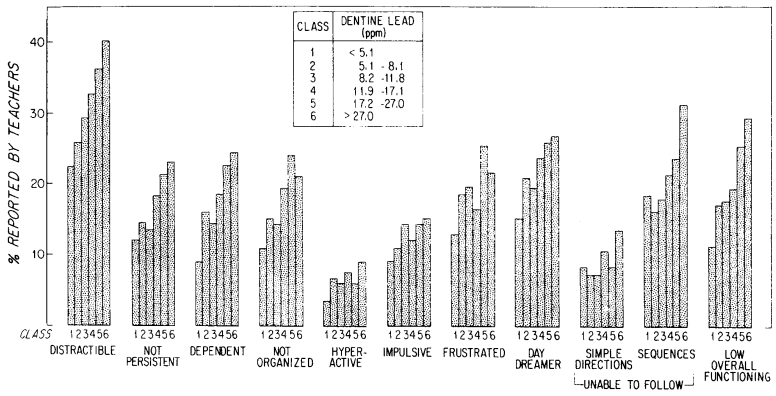

Fig. 2. Classroom behavior in relation to dentine lead levels. Children in the first and second grades (N = 2146) were evaluated on an 11-item forced choice questionnaire. Subject were arranged by ascending dentine lead concentration, and the number of unfavorable grades for each item tabulated. Reprinted with permission from *The New England Journal of Medicine, 300*, 689–695, 1979.

Hatzakis, Kokkevi, Katsouyanni et al. (1987), studying Greek children from Lavrion, reported similar data to those reported by Needleman and Yule on the Boston Rating Scale. Silva, Hughes, Williams, and Faid (1988) studied New Zealand children with extremely low body burdens of lead. Lead levels were significantly related to scores on the Parent Rutter and the Teacher Rutter Questionnaires and on the hyperactivity scale of both measures. Thomson et al. (1989) followed a sample of middle-class Edinburgh children who had demonstrated lead-related deficits in IQ and found that scores on both Parent and Teacher versions of the Rutter scales were significantly related to lead levels after covariate adjustment. Most recently, Fergusson, Horwood, and Lynskey (1993), who has been following a large cohort of New Zealand children classified by dentine lead levels, reported statistically significant relationships between lead and reading, scholastic ability, and measures of inattention/restlessness. Sciarillo (1992) studied 201 African American children aged 2–5 using the CBCL scales. The 78 high-lead subjects (blood lead >15 μg/dl) had higher scores on the Externalizing, Internalizing, and Total scales. In addition, 16.7% of the high-lead girls reached clinical range on the hyperactivity scale as opposed to 4% of the low lead (< 15 μg/dl) girls. Of the males, 29% of the high-lead subjects scored above the clinical threshold for aggressiveness compared to 13% of the low-lead subjects. Bellinger, Leviton, Allred, and Rabinowitz (1994) evaluated Teachers' CBCL ratings in a large sample of 8-year-olds. Dentine lead levels were related to Total Behavioral Problem, Internalizing, and Externalizing Scores.

9. Experimental studies of lead at low dose

We have said those epidemiological studies cannot achieve complete control of confounding or unequivocally establish causal direction. It is here that students of lead and other toxins turn to the experimental laboratory to fill in the gaps in causal inference.

Of the formidable literature in this area, we select the work of Rice (1985) and Cory-Slechta. In the rodent, Cory-Slechta's group has produced a number of studies showing interference with passive avoidance behavior and learning (Cohn, Cox, & Cory-Slechta, 1993). Rice has studied a group of monkeys given lead before birth and through childhood, and found changes in learning and attention at blood lead levels in the range experienced by American children (Rice, 1985).

10. The neurotoxicology of lead

It may be argued that this area is beyond the scope of the analysis presented. We believe that any serious student of lead toxicity must be familiar with the influence of lead on brain chemistry and function. Any attempt to discuss such questions as causal direction, linearity of effect, and reversibility of effect without knowing something about the neurochemistry and neurotoxicology of lead is to accept a serious and avoidable handicap. There is no mention of this literature in Kaufman's paper although there are literally thousands of studies demonstrating that lead alters essential neurochemical functions.

While the specific biological mechanisms underlying lead's effect on cognition, aggression, and impulsivity are not known, the metal is associated with a large number of alterations

of CNS function, some of which are involved in impulse control. Lead interferes with synaptogenesis (Averill & Needleman, 1980), diminishes the inhibition of brain phosphokinase C (Markovac & Goldstein, 1988), decreases norepinephrine-induced inhibition (Taylor, Nathanson, Hoeffer, Olson, & Steiger, 1978), and lowers brain levels of serotonin or 5-HIAA (Lasley, Greenland, Minnema, & Michaelson, 1984; Widmer, Butikofer, Schlumf, & Lichtenseiger, 1991). Lead exposure is associated with increased levels of D-aminolevulinic acid, which may antagonize GABA inhibition (Meredith, Moore, Campbell, Thompson, & Goldberg, 1978). Lead also enhances both D_1 and D_2 dopamine sensitivity, and alters NMDA-receptor sensitivity (Cory-Slechta, Pokora, & Widzowski, 1992).

11. The issue of causal direction

Kaufman mentions this obliquely, quoting Pocock; we confront it because it is one of the oldest criticisms of the lead–IQ hypothesis. For years, critics have said that because mentally retarded children tend to display more mouthing behavior, it is not lead that lowers IQ, but it is low IQ that causes children to take in more lead. Specifically to address this claim, we designed the first forward study of lead and demonstrated that umbilical cord blood lead levels were associated with infant developmental scores through 24 months of age. One of the axioms of causal inference is that cause must have temporal precedence over effect. In addition, there is the indisputable evidence from the work of Rice (1985) and Cory-Slechta et al. that lead administered to primates and rodents under controlled conditions is associated with later behavioral deficits.

12. The limitations of IQ measurements

Kaufman returns to this theme repeatedly, noting that they measure a limited aspect of human functioning and are not to be used as the sole criterion for drawing inferences about educational or societal implications. He closes his paper with a discussion of Sternberg's triarchic model of intelligence. He recommends, in particular, Sternberg's unpublished group-administered test (STAT). While this advice is useful, it ignores some of the practical realities involved in setting exposure standards. Decisions have to be made on the basis of the available database, which will never be complete, perfect, or monolithic in terms of findings. Once more, Kaufman wants to have it two ways: to suggest the need for different types of data; and to simultaneously ignore the reports on different types of data: school failure, attention, aggression, and delinquency.

13. Conclusion

Causal inference is a complex endeavor. Hume asserted that cause is a relationship between experiences rather than one between facts. Because causality is not subject to empirical proof, the chief task of scientists is to make causal inferences. The construction of trustworthy causal

inferences about widely distributed toxic products requires an enormous amount of work, discipline, and modesty. It demands a serious as well as careful look at all of the investigations, without a priori selections of the data to be considered, and with a bit of humility.

The clearest example of this is, of course, the issue of tobacco and health. Out of the struggle to reach sound conclusions from the wealth of epidemiological and experimental laboratory studies on the health consequences of smoking, the Causal Canons of Sir Austin Bradford Hill listed below were developed:

1. Temporality. Cause precedes effect.
2. Experiment. Has the agent been systematically manipulated?
3. Biological gradient. The dose–response effect.
4. Plausibility, coherence, analogy. Is the putative cause biologically plausible? Is it coherent — that is, is it not in conflict with known facts? Is it nonspurious, that is, not due to some confounding factor? Is it analogous to known relationships between other agents and disease?
5. Strength. Is the relationship strong?
6. Consistency. Is the effect consistent across different populations? Is it consistent with experimental animal data?
7. Specificity. Is the effect specific to the cause?

Applying these canons to the case of low-level lead exposure, we find that the evidence shows that

1. Exposure to lead precedes behavioral deficit in many experimental situations and in epidemiological studies of infants from pregnancy onward.
2. The animal data are strongly supportive of the epidemiological findings. The natural experiment in the Boston study, in which the direction of the association between prenatal lead exposure and sociodemographic indicators was the opposite of what would be expected, was cited above.
3. The dose–response relationship between lead burden and degree of deficit has been established in many studies.
4. Lead at low dose is a plausible biological cause of brain dysfunction and consequent behavioral deficit. At high doses, it kills brain cells and causes cerebral hemorrhage and edema. Administration to animals produces effects similar to those found in humans at similar doses. A large number of possible biochemical mechanisms of lead's neurotoxicity have been demonstrated.
5. The relationship is consistent and robust across populations. The measured effect size is modest, but the effect projected across populations is substantial.
6. Consistency has been established in studies from the US, the UK, Italy, Greece, Scandinavia, New Zealand, Australia, China, Mexico, and Germany, among others.
7. The effect is not specific.

We believe that Kaufman's effort is not a fair test of the causal inferences made and accepted by investigators around the world. He recognizes the limitations of IQ measurement

but restricts his area of investigation to it, and thus avoids examination of some of the most relevant evidence of lead's effects on the CNS. We have shown that within the narrow area of studies that he did select, he ignored data that conflicted with his principal assertions. By shrinking from examination of the entire body of information, he forfeits the right to comment on the issues of temporality, natural experiment, biological gradient, plausibility, or coherence. By slighting the animal literature, he avoids confronting the elegant work in primates and rodents showing behavioral deficits and cognate neurochemical changes, all at low exposures, that provide support and explanatory power for the conclusions drawn from epidemiological studies.

Applying Kaufman's criteria for IQ assessment, a sufficient database does not exist for regulating lead or, for that matter, for any other environmental neurotoxicant. His standard of evidence may be appropriate for a purely academic debate but is of little use in deciding a public health question in which regulatory decisions or nondecisions carry important human costs.

The final sentence of his paper calls for *conclusive* evidence of a negative impact on children's functioning. This is disingenuous; Kaufman ignores the provisional, contingent basis of all scientific conclusions. Conclusions about lead are always open to revision even if the probability of no effect is vanishingly small, i.e., in the range of 2.8×10^{-13}.

The process of risk assessment balances the costs of Types I (enacting exposure standards more stringent than they need to be) and II errors (failing to enact exposure standards that are sufficiently protective of the most vulnerable subgroup of the population). The Type I error has been called the "Producer's Error," and the Type II error the "Consumer's Error."[1] This neatly captures the practical results of each error. If we commit a Type I error, profits from the sale of lead products are reduced; if we commit a Type II error, IQ and CNS integrity are reduced. In general, we, as public health figures, would prefer a Type I rather than a Type II error. This is the position also taken by many of the US regulatory agencies that Kaufman takes to task. His critique, however, fails under critical scrutiny. Nor does it merit the adjective "careful" that he has bestowed on it.

References

Atkinson, S. E., Crocker, T. D., & Needleman, H. L. (1987). The importance of specification uncertainty and intolerance to measurement error in a study of the impact of dentine lead on children's IQs. *International Journal of Environmental Studies*, *29*, 127–138.

Averill, D., & Needleman, H. L. (1980). Neonatal lead exposure retards cortical synaptogenesis in the rat. In: H. L. Needleman (Ed.), *Low level lead exposure: the clinical implications of current research* (pp. 201–210). New York: Raven Press.

Bellinger, D., Leviton, A., Allred, E., & Rabinowitz, M. (1994). Pre- and postnatal lead exposure and behavior problems in school age children. *Environmental Research*, *66*, 12–30.

Bellinger, D., Leviton, A., & Waternaux, C. (1989). Lead, IQ, and social class. *International Journal of Epidemiology*, *18*, 180–185.

Bellinger, D., Leviton, A., Waternaux, C., & Allred, E. (1985). Methodological issues in modeling the relationship

[1] In this regard, we note that Kaufman acknowledges the support of Ethyl Corporation of America, the producers of tetraethyl lead. A similar paper was read by Kaufman before the EPA's Science Advisory Board under the auspices of The Lead Industry Association.

between low-level lead exposure and infant development: examples from the Boston lead study. *Environmental Research*, *38*, 119–129.

Bellinger, D., Leviton, A., Waternaux, C., Needleman, H., & Rabinowitz, M. (1987). Longitudinal analyses of prenatal and postnatal lead exposure and early cognitive development. *New England Journal of Medicine*, *316*, 1037–1043.

Bellinger, D., Leviton, A., Waternaux, C., Needleman, H., & Rabinowitz, M. (1989). Low-level lead exposure and early development in socioeconomically advantaged urban infants. In: M. Smith, L. Grant, & A. Sors (Eds.), *Lead exposure and child development: an international assessment* (pp. 345–356). Boston: Kluwer Academic Publishing.

Bellinger, D., Needleman, H., Bromfield, R., & Mintz, M. (1984). A follow-up study of the academic attainment and classroom behavior of children with elevated dentine lead levels. *Biological Trace Element Research*, *6*, 207–223.

Bellinger, D., Stiles, K. M., & Needleman, H. L. (1992). Low-level lead exposure, intelligence and academic achievement: a long-term follow up study. *Pediatrics*, *90*, 855–861.

Brandt, J., & van Gorp, W. (1999). American Academy of Clinical Neuropsychology Policy on the use of non-doctoral level personnel in conducting clinical neuropsychological evaluations. *Journal of Clinical and Experimental Neuropsychology*, *21*, 1.

Byers, R. K., & Lord, E. E. (1943). Late effects of lead poisoning on mental development follow-up reassessment. *Pediatrics*, *67*, 911–919.

Cohn, J., Cox, C., & Cory-Slechta, D. (1933). The effects of lead exposure on learning in a multiple repeated acquisition performance schedule. *Neurotoxicology*, *14*, 329–346.

Cory-Slechta, D., Pokora, M., & Widzowski, D. (1992). Postnatal lead exposure induces supersensitivity to the stimulus properties of D2-D3 agonist. *Brain Reserach*, *598*, 162–172.

Ernhart, C., Landa, B., & Schell, N. (1981). Subclinical levels of lead and developmental deficit—A multivariate follow-up reassessment. *Pediatrics*, *67*, 911–919.

Fergusson, D., Horwood, J., & Lynskey, M. (1993). Early dentine lead levels and subsequent cognitive and behavioral development. *Journal of Child Psychology and Psychiatry*, *34*, 215–227.

Fergusson, D., Horwood, J., & Lynskey, M. (1997). Early dentine lead levels and educational outcomes at 18 years. *Journal of Child Psychology and Psychiatry*, *38*, 471–478.

Greene, T., & Ernhart, C. B. (1993). Dentine lead and intelligence prior to school entry: a statistical sensitivity analysis. *Journal of Clinical Epidemiology*, *46*, 323–339.

Hatzakis, A., Kokkevi, A., Katsouyanni, K., Maravelias, K., Salaminios, F., Kalandidi, A., Koutselinis, A., Stefanis, K., & Trichopculous, D. (1987). Psychometric intelligence and attentional performance deficits in lead-exposed children. In: *International conference: heavy metals in the environment* (pp. 204–209). Edinburgh: CEP Consultants.

Lasley, S. M., Greenland, R. D., Minnema, D. J., & Michaelson, I. A. (1984). Influence of chronic inorganic lead exposure on regional dopamine and 5-hydroxytryptamine turnover in rat brain. *Neurochemical Research*, *90*, 1675–1688.

Leamer, E. (1978). *Specification searches: ad hoc inference with nonexperimental data*. New York, NY: Wiley.

Markovac, J., & Goldstein, G. W. (1988). Picomolar concentrations of lead stimulate brain protein kinase C. *Nature*, *334*, 71–73.

Meredith, P. A., Moore, M. R., Campbell, B. C., Thompson, G. G., & Goldberg, A. (1978). Delta-aminolevulinic acid metabolism in normal and lead-exposed humans. *Toxicology*, *9*, 1–9.

National Academy of Sciences/National Research Council. (1993). *Measuring lead exposure in infants, children, and other sensitive populations*. Washington, DC: National Academy of Sciences Press.

Needleman, H. L., & Gatsonis, C. (1990). Low level lead exposure and the IQ of children. *Journal of the American Medical Association*, *263*, 673–678.

Needleman, H. L., Gunnoe, C., Leviton, A., Peresie, H., Maher, C., & Barret, P. (1979). Deficits in psychological and classroom performance of children with elevated dentine lead levels. *New England Journal of Medicine*, *300*, 689–695.

Needleman, H. L., Leviton, A., & Bellinger, D. (1982). Lead associated intellectual deficit. *New England Journal of Medicine*, *306*, 367.

Needleman, H. L., Schell, A., Bellinger, D., Leviton, A., & Allred, E. N. (1990). The long-term effects of exposure to low doses of lead in childhood: an 11-year follow-up report. *New England Journal of Medicine*, *322*, 83–88.

Perino, J., & Ernhart, C. (1974). The relation of subclinical lead level to cognitive and sensorimotor impairment in Black preschoolers. *Journal of Learning Disabilities*, *7*, 26–30.

Poole, C. (1991). Multiple comparisons? No problem!. *Epidemiology*, *2*, 241–243.

Rabinowitz, M., Allred, E., Bellinger, D., Leviton, A., & Needleman, H. (1990). Lead and childhood propensity to infectious and allergic disorders: is there an association? *Bulletin of Environmental Contamination and Toxicology*, *44*, 657–660.

Rice, D. C. (1985). Chronic low lead exposure from birth produces deficits in discrimination reversal in monkeys. *Toxicology and Applied Pharmacology*, *75*, 201–210.

Rothman, K. (1990). No adjustments are needed for multiple comparisons. *Epidemiology*, *1*, 43–46.

Rothman, K., & Poole, C. (1988). A programme for strengthening weak associations. *International Journal of Epidemiology*, *319*, 955–959.

Schlesselman, J. J. (1978). Assessing the effects of confounding variables. *American Journal of Epidemiology*, *108*, 3–8.

Schwartz, J. (1993). Beyond LOEL's, *P* values and vote counting; methods for looking at the shapes and strengths of associations. *Neurotoxicology*, *14*, 237–246.

Sciarillo, W. (1992). Lead exposure and child behavior. *American Journal of Public Health*, *82*, 1356–360.

Silva, P. A., Hughes, P., Williams, S., & Faid, J. M. (1988). Blood lead, intelligence, reading attainment and behaviour in eleven year old children in Dunedin, New Zealand. *Journal of Child Psychology and Psychiatry*, *29*, 43–52.

Smith, H. (1964). Pediatric lead poisoning. *Archives of Environmental Health*, *8*, 256–261.

Taylor, D., Nathanson, J., Hoffer, B., Olson, L., & Seiger, A. (1978). Lead blockade of norepinephrine-induced inhibition of cerebellar Purkinje neurons. *Journal of Pharmacology and Experimental Therapeutics*, *206*, 371–381.

Thomas, D., Stram, D., & Dwyer, J. (1993). Exposure measurement error: influence of exposure–disease relationships and methods of correction. *Annual Review of Public Health*, *14*, 69–94.

Thomson, G. O. B., Raab, G., Hepburn, W. S., Hunter, R., Fulton, M., & Laxen, D. P. H. (1989). Blood lead levels and children's behaviour — results from the Edinburgh lead study. *Journal of Child Psychology and Psychiatry*, *30*, 515–528.

Waternaux, C., Laird, N., & Ware, J. (1989). Methods for analysis of longitudinal data: blood lead concentrations and cognitive development. *Journal of the American Statistical Association*, *84*, 33–41.

Widmer, H. R., Butikofer, E. E., Schlumf, M., & Lichtenseiger, W. (1991). Pre- and postnatal lead exposure affects the serotonergic system in the immature rat brain. *Experientia*, *47*, 463–466.

Yule, W., & Lansdown, R. (1981). Blood lead concentrations and school performance. *British Medical Journal*, *283*, 1336.

Yule, W., Lansdown, R., Urbanowicz, M., & Miller, I. (1981). The relationship between blood lead concentrations. Intelligence and attainment in a school population: a pilot study. *Developmental Medicine and Child Neurology*, *23*, 567–576.