Pergamon

Archives of Clinical Neuropsychology
16 (2001) 403–431

Archives
of
CLINICAL
NEUROPSYCHOLOGY

How dangerous are low (not moderate or high) doses of lead for children's intellectual development?

Alan S. Kaufman

Child Study Center, School of Medicine, Yale University, 230 South Frontage Road, PO Box 207900, New Haven, CT 06520-7900, USA

Accepted 28 August 2000

Abstract

This paper addresses the points raised by five groups of scientists who were invited to respond to my article on the relationship of low blood lead to IQ loss. I dealt with these comments as a scientist who believes that the case is not closed on this topic, as some respondents believe, but that debate is healthy and can move the field to the next level. The criticisms about the measurement of parents' IQ, multiple comparisons, the linearity of the lead–IQ relationship, and the societal consequences of a few points of IQ loss appear weak in the face of an array of evidence that bears on these topics. However, criticisms about my emphasis on the need to control for a wide variety of potential confounders has validity. Ultimately, however, the case for the relationship of low blood lead to IQ loss seems to rest tenuously on data obtained from samples that included numerous subjects with moderate to severe levels of blood lead. © 2001 National Academy of Neuropsychology. Published by Elsevier Science Ltd.

Keywords: Blood lead; Intelligence; IQ; Neuropsychological functioning; Neurotoxins; Lead research

> So it's up to me is it? Well, I'll tell y' — in all my years I never heard, seen, nor smelled an issue that was so dangerous it couldn't be talked about. Hell yes, I'm for debatin' anything — Rhode Island says Yea!
>
> — Stephen Hopkins, 1776 (Peter Stone)

I believe that lead is a neurotoxin. I do not believe that ingesting or inhaling lead is a smart idea and I would not let my children or grandchildren munch on paint chips. However, the deleterious effect of lead is not the issue under debate. The issue is the possible harm that low levels of lead can cause regarding children's intellectual development. More specifically, the issue is whether blood lead levels (BLLs) within the 10–20-

0887-6177/01/$ – see front matter © 2001 National Academy of Neuropsychology.
PII: S0887-6177(00)00090-1

μg/dl range (“silent doses,” according to Needleman & Bellinger, 2001) cause children to lose a few points of IQ.

To Needleman and Bellinger (2001), there is no issue or debate; the “hypothesis that lead damages children’s brains at silent doses” is “widely accepted.” To Nation and Gleaves (2001), likewise, the case is closed: “the data on the topic are so compelling that the only sensible conclusion is that ‘safe’ blood lead levels must be adjusted downward .... There is indeed a ‘trout in the milk’ and we believe the bulk of the evidence is very strong in showing that lead threatens early childhood development.” Nation and Gleaves rely on a concrete visual image, as well as an intricate web of statistical and methodological “proofs” to demonstrate the strength of their position and the weakness of mine. In contrast, Needleman and Bellinger prefer to demean with phrases like “one-eyed astronomer,” “one more recital of a set of criticisms raised primarily by spokespersons of the lead industry,” and “Kaufman’s comments on this issue are once again colored by his failure to understand. ...”

However, neither an excess of emotional investment in the lead–IQ research nor knee-jerk acceptance of hypotheses as trout-in-the-milk fact should squelch scientific debate on the topic; and despite the vitriolic comments of a few of the respondents to my article, this debate is, indeed, a scientific one. Many of the issues are not yet resolved, and, as Hebben (2001) makes abundantly clear in her supportive response to my article, it does not sound, look, or smell like something that is so dangerous it cannot be talked about.

1. Limiting the scope of the literature review

Let us start with my choice to limit the topic to the impact of low BLLs on IQ loss and to confine myself primarily to the 26 articles considered of sufficient quality to be included in at least one of the three meta-analyses on the topic (i.e., Needleman & Gatsonis, 1990; Pocock, Smith, & Baghurst, 1994; Schwartz, 1994a, 1994b). Needleman and Bellinger (2001) strongly criticized my restriction of the field “to those studies that were the subject of two (sic) meta-analyses,” believing me to be “oblivious to the human literature on lead and attention, lead and school failure, lead and aggression, ... the cognate studies of behavior in lead-exposed primates and rodents, and the vast experimental literature on the neurochemistry of lead.”

I do not believe that I need to defend my choice to confine myself to IQ loss as the variable of interest any more than the authors of the three meta-analyses had to defend their decisions: “The major outcome of interest is full-scale IQ” (Needleman & Gatsonis, 1990, p. 673); “*Main outcome measures* — For each study, the regression coefficient of IQ on lead” (Pocock et al., 1994). Yet, Schwartz (1994a) did defend his decision: “Full-scale IQ in school-age children was chosen as the outcome for several reasons. IQ is again an outcome that has received the most public policy attention — it is the outcome measure of interest. Full-scale IQ is chosen because it is always reported and because its use avoids issues of which subscale is more relevant for lead exposure” (p. 45). Those reasons are still valid, and account for my selecting IQ as the variable of choice. In addition, IQ is my specialty, as test developer, text writer, and researcher. I am not guilty of “selection bias” (Needleman & Bellinger, 2001), but of selecting a reasonable and clearly defined body of literature as the

area of debate. Three highly respected meta-analyses limited their primary focus to low BLLs and IQ. I limited my primary focus to these three meta-analyses.

This decision to restrict the playing field does not render me "oblivious" of other pertinent literature on lead (Needleman & Bellinger, 2001) any more than Needleman and Bellinger's recitation of the studies in support of their claims of an association between lead level and attention/behavior implies that they are oblivious to the body of literature that is nonsupportive or minimally supportive of the alleged relationship (e.g., Harvey et al., 1988; Lansdown, Yule, Urbanowicz, & Hunter, 1986; Smith, Delves, Lansdown, Clayton, & Graham, 1983; Wasserman, Staghezza-Jaramillo, Shrout, Popovac, & Graziano, 1998) or of the scientists who are frankly critical of the methodology and conclusions of Needleman's and other studies that strongly support the lead/behavior relationship (Epstein, 1998; Ernhart, 1996; Sachs, 1996; Sayre, 1996; Wasserman et al., 1998). In fact, I made my reasons to exclude the attention/behavior literature from my critique quite clear in the third paragraph of my literature review (Kaufman, 2001), and demonstrated awareness of pertinent, illustrative literature on both sides of the issue.

2. Lead level and educational outcomes

Needleman and Bellinger (2001) chastised me for excluding from my paper "one widely recognized outcome [they] published in the *New England Journal of Medicine* in 1990. Higher exposures to lead in early childhood were associated with a 7-fold increase in high school failure, and a six-fold increase in reading disabilities." In that study (Needleman, Schell, Bellinger, Leviton, & Allred, 1990), their "high-lead" group had tooth lead levels of >20 ppm (parts per million, which is the same as microgram per gram).

Though dentine and BLLs are not directly comparable and sometimes correlate poorly (e.g., Rabinowitz, Leviton, & Bellinger, 1993), several studies show correlations of about .50 (e.g., Smith et al., 1983). When studies have reported both tooth lead levels and BLLs for the same children, the numerical values for blood lead in micrograms per deciliter are invariably higher than the values for tooth lead in micrograms per gram. Winneke (1979) computed a factor of 2.5, such that 2 μg/g dentine lead would be equivalent to about 5 μg/dl blood lead. In Bergomi et al.'s (1989) investigation, the microgram per deciliter values for blood lead were about five points higher than the micrograms per gram values for dentine lead. The differences were a little smaller (2.2 points) in one study (Fulton, Paterson, Raab, Thomson, & Laxen, 1989) and a bit larger (about four to eight points) in another (McMichael et al., 1994) and are known to vary both with the age at which the blood sample is taken (Rabinowitz et al., 1993) and with the type of tooth that is analyzed (Smith et al., 1983).

Whether one uses the factor of 2.5 or adds a constant, one thing is clear: Needleman et al.'s "high-lead" group, with dentine lead values >20 μg/g, is truly high in their body burden of lead, or maybe moderate, but certainly not low. Indeed, even their "low-lead" group is conceivably not low. Subsamples of their "high" and "low" groups that were ultimately followed up 11 years later in Needleman et al.'s (1990) study were, in fact, assessed for BLL about 3 to 4 years prior to the evaluation of their teeth. The subsample of "high-lead" children had a mean of 35.5 μg/dl (S.D. = 10.1) with a maximum value of 54

and the "low-lead" group averaged 23.8 μg/dl (S.D. = 6.0). Neither of these values is within the 10–20-μg/dl range that defines low blood lead. Needleman et al. have shown a relationship between moderate to high levels of lead and educational outcome, but not to low levels of lead. Needleman et al. measured the current (adolescent) lead level of the first 48 subjects, but discontinued the blood tests "because none had a lead level exceeding . . . 7 μg/dl" (p. 84). However, the finding of low lead in adolescence does not mean a reclassification of the subjects as having a low lead level because, "All students of lead exposure know that 2 years of life is precisely when lead levels peak" (Needleman & Bellinger, 2001). The title of their outcome study, which includes "low doses of lead" in its title, is clearly misleading.

In addition, even the validity of the significant relationships in the outcome study is open to challenge. Epstein (1998) makes one powerful and insightful criticism of the Needleman et al. (1990) educational outcome study — the failure to control for, or even consider, the potentially confounding relationship between city of residence and school performance. The two school districts from which Needleman et al. selected their sample (the neighboring cities of Chelsea and Somerville in Massachusetts) differed considerably. For example, the high school drop-out rate between 1988 and 1990 averaged about 18% for Chelsea vs. 6% for Somerville, and "the Chelsea public school system suffered a widely publicized total collapse in the late 1980s" (Epstein, 1998, p. 130). It is quite conceivable that Needleman et al.'s significant results were a spurious finding owing to their failure to consider a key potential confounder, one they should have been aware of in view of the publicity surrounding the collapse of the Chelsea school district.

Needleman and Bellinger (2001) claimed that the results of a similar outcome study in New Zealand by Fergusson, Horwood, and Lynskey (1997) gave strong support for their 1990 findings at even lower lead levels. The latter study did indeed investigate children with lower dentine lead levels than in Needleman et al.'s (1990) follow-up, and they did find statistically significant results with similar educational outcomes. Fergusson et al. (1997) did not, however, control for or consider city of residence as a potential confounder. Perhaps in New Zealand, that would have been a meaningful factor, perhaps not. Of greater concern is the meaningfulness of their findings in general. They had a large sample ($N = 881$), which is a positive aspect of their longitudinal study. However, as Kerlinger (1979) observed more than 20 years ago, "With a large number of subjects . . . tests of significance show statistical significance even when a difference between means is quite small, perhaps trivial, or a correlation coefficient is very small and trivial" (p. 318). The reporting of statistical significance needs to be augmented routinely with measures of effect size to help determine the meaningfulness of a significant result (e.g., Daniel, 1998; Nix & Barnette, 1998). Fergusson et al. reported significant correlations between children's lead level and later educational outcomes, correlations that remained significant even after control of numerous covariates. The authors report the magnitude of the coefficients before adjustment of covariates, but not after the adjustment. Before adjustment, these values ranged from absolute values of 0.14 to 0.18. Though effect sizes are not specifically reported by the authors they are easily computed by squaring these coefficients, a process that indicates shared variance between predictor and outcome variables ranging from 2.0% to 3.2%. In view of the fact that nine different covariates (measures of family background) also

correlated significantly with dentine lead level, it is reasonable to assume that the effect sizes of the adjusted coefficients are even more trivial than the unadjusted values. Statistically significant? Yes. Meaningful? No.

In fact, Fergusson et al. (1997) offered as one of their main supports for their results the agreement of their findings with those of Needleman et al. (1990). Although Fergusson et al. concede that "the effects of lead on cognitive development appear relatively small" (p. 477), they apparently feel buoyed by the replicability of the findings: "This convergence of findings clearly suggests that the associations between dentine lead levels and longer-term outcomes are not specific to this sample" (p. 476). Ultimately, both Needleman and Bellinger (2001) and Fergusson et al. have planted their flags firmly into a quagmire.

3. The animal-lead literature

Needleman and Bellinger (2001) stated: "By slighting the animal literature, [Kaufman] avoids confronting the elegant work in primates and rodents showing behavioral deficits and cognate neurochemical changes, *all at low exposures*" (italics mine). First, I had no reason to enter the experimental psychology world of the rat–monkey literature or the neurochemistry literature because nowhere in my paper did I challenge the fact that lead is a neurotoxin that affects the nervous system adversely when lead enters the bloodstream in sufficiently high doses. The issue is the effect of low levels of lead on intelligence. Needleman and Bellinger (2001) emphasized the animal research work of Cory-Slechta, Rice, and others as definitive proof of their contentions about the hazardous effects of lead at low doses, subsuming this research under the heading "Experimental Studies of Lead at Low Dose" and including the phrase "all at low doses" in the previous quote. But is it low dose? In the studies specifically cited by Needleman and Bellinger, Rice (1985) studied monkeys with BLLs of 15 or 25 μg/dl during infancy and steady-state levels of 11 or 13 μg/dl; she observed deficits in "reversal" tasks, but not in the acquisition of the discrimination tasks. Cohn, Cox, and Cory-Slechta (1993) observed learning deficits in rats having blood leads as low as 20–25 μg/dl. Neither the Rice BLLs during infancy nor the Cory-Slechta BLLs are comfortably within the low dose range of 10–20 μg/dl.? In addition, Rice (1996) interpreted the reversal deficits as probable evidence of distractibility in the monkeys and Cory-Slechta (1997) cited perseverance as the key element in the rats' deficits. Both distractibility and perseverance are in the behavioral, not the intellectual, domain.

Furthermore, the number of animal studies that found significant results with BLLs within or near the 10–20-μg/dl "low-dose" range are clear outliers in the body of literature that has accumulated on this topic. From reviews of the animal-lead research (Banks, Ferretti, & Shucard, 1997; Rice, 1993), it is evident that most of the studies cited cannot be considered low lead level by today's standards. In one study, cynomolgus monkeys were dosed with lead from birth through the first year; BLLs peaked at 50 μg/dl and averaged about 30 μg/dl. In another study, newborn and infant monkey BLLs ranged from 30 to 35 μg/dl, and in several others, rhesus monkeys and rats were dosed in the 25–50-μg/dl range. Rice (1993) also reviewed studies in which lead levels that peaked at 300 μg/dl, and leveled out at 90 μg/dl during the first year of life, caused disruption in rhesus monkeys' performance on a test of

spatial memory; and in which "[b]lood levels of approximately 50 or 90 μg/dl were associated with impairment early in life on a series of tasks including spatial, color, and size discrimination reversal tasks" (p. 169). The issue is not whether moderate or high doses of lead causes cognitive impairment. The issue is low lead levels, and causality at low levels has not been established with rats or monkeys.

In addition, even well-designed animal studies that are above reproach on psychometric or research-design grounds cannot unilaterally address the issue of causality in children. Animals differ from children in fundamental and crucial ways in the size and functioning of the cerebral cortex; what is considered a low level of lead in humans may not truly be low in primates or other animals. For example, evidence suggests that rats and monkeys may tolerate higher lead levels than humans before toxic effects are evident (Davis, Otto, Weil, & Grant, 1990). Brown (2001), an experimental psychologist, makes a cogent, pertinent point in his response: "Appropriateness of animal models is not always easily determined. Not only are *equivalence of dosage and measures difficult to determine*, but also apparently minor methodological differences between human and nonhuman studies may be quite important" (italics mine).

4. How much do we know about low lead levels and IQ?

After realizing the degree to which Needleman et al.'s (1979) initial sample and follow-up sample (Needleman, Riess, Tobin, Biesecker, & Greenhouse, 1996) did not fit into the low BLL category by any stretch of the definition, I decided to examine the lead levels of each of the 26 samples included in the meta-analyses. Table 1 presents the results of this systematic examination.

5. Lead levels in the 26 studies

In Table 1, I have indicated in study-by-study fashion the lead levels of each separate sample, using whatever data were provided by the investigators. When lead data were presented for multiple age levels, I reported only the values for age 2 years; when means and S.D.s were provided, I gave the mean along with the values that corresponded to +1S.D. and +2S.D. I have also included a column that answers the question, "Does the sample qualify as 'low lead' level?" Of the 26 studies, I was able to answer 'yes' to six of the samples. Of the other 20 studies whose samples were not legitimately classifiable as low BLL, I used the answer 'not even close' for nine. To illustrate these classifications: (a) 'yes' — the sample for Lansdown et al. (1986) had BLLs that ranged from 7 to 24 μg/dl (mean = 12.8, +2S.D. = 18.9); (b) 'no' — the sample for Baghurst et al. (1992) included a 'low group' with a mean of 11.6 μg/dl and a 'high group' with a mean of 27.1 μg/dl; and (c) 'not even close' — the sample for Hatzakis et al. (1987, 1989) had a BLL range of 7.4–63.9 μg/dl (mean = 23.7, +2S.D. = 42.1).

The claims for the effect of low BLLs on IQ have been based on a number of studies that have included portions of the sample, sometimes substantial portions, that have been well above the low BLL range of 10–20 μg/dl. How can we be sure that the significant effects

Table 1
Description of lead levels in 26 studies and classification of the lead levels as "low" or "not low"

Reference	Type study	Does the sample qualify as "low lead" level?	Blood lead levels (μg/dl)
Baghurst et al., 1992	Prospective	No	Age 2 Low group = 11.6 High group = 27.1
Dietrich et al., 1993	Prospective	Not even close	Age 2 Mean +1S.D. +2S.D. 17.1 25.5 33.9 (35% of sample had at least one Pb ≥ 25 in the first 5 years)
Ernhart et al., 1989	Prospective	Not even close	Age 2 Mean +1S.D. +2S.D. 16.7 23.2 29.6 (Range = 5.4–41.8)
Cooney, Bell, & Stavron, 1991	Prospective	Not even close	Age 2 Mean Maximum 15.8 40
Bellinger et al., 1992	Prospective	Yes	Age 2 Mean +1S.D. +2S.D. 7.7 12.6 17.5
Hatzakis et al., 1987; also, Hatzakis et al., 1989	Cross-sectional	Not even close	Mean +1S.D. +2S.D. 23.7 32.9 42.1 (Range = 7.4–63.9)
Fulton et al., 1987	Cross-sectional	No	Mean = 11.5 (Range = 3.3–34); 2% had Pb levels > 25
Winneke, Brockhaus, Ewers, Kramer, & Neuf, 1990	Cross-sectional	No	Includes eight samples. Overall range is from < 5 to about 60. Only one sample produced significant WISC results (mean Pb = 22.0, S.D. = 1.4)
Silva, Hughes, Williams, & Faed, 1988	Cross-sectional	No	Mean +1S.D. +2S.D. 11.1 16.0 20.9 (Range = 4–50; 2 *S*'s > 30; distribution showed small hump at 20–31)
Yule, Lansdown, Millar, & Urbanowicz, 1981	Cross-sectional	No	Mean +1S.D. +2S.D. 13.5 17.6 21.8 (Range = 7–32)
Lansdown et al., 1986	Cross-sectional	Yes	Mean +1S.D. +2S.D. 12.8 15.8 18.9 (Range = 7–24)
Harvey et al., 1988	Cross-sectional	No	Mean +1S.D. +2S.D. 13.0 17.2 21.3 (Range = 4–29)

(continued on next page)

Table 1 (*continued*)

Reference	Type study	Does the sample qualify as "low lead" level?	Blood lead levels (μg/dl)
Wang, Xu, Thang, & Wang, 1989	Cross-sectional	Not even close	Mean +1S.D. +2S.D. 21.1 31.2 41.3 (Range = 4.5 – 52.8)
Ernhart, Landa, & Wolf, 1985	Cross-sectional	Not even close	Range = 10 – 70 Low group ≤ 30 High group ≥ 40
Schroeder et al., 1985	Cross-sectional	Not even close	Initial study: Range = 6 – 59 (2/3 or more >20); 5-year follow-up: all <30. Only initial study found significance.
Hawk et al., 1986	Cross-sectional	Not even close	High blood level Mean +1S.D. +2S.D. 26.7 40.8 54.9 (Range = 6.2 – 56.0) Mean blood level Mean +1S.D. +2S.D. 21.8 31.6 41.4 (Range = 4.5 – 52.8)
Winneke et al., 1985	Cross-sectional	Yes	Mean +1S.D. +2S.D. 8.2 9.6 11.0 (Range = 3.9 – 22.8)
Fergusson, Fergusson, Horwood, & Kinzett, 1988	Cross-sectional	Yes	Tooth lead only Mean = just over 6 μg/g Low group = 0 – 2 High group = 12 +
Smith et al., 1983	Cross-sectional	No	Tooth and blood lead Mean tooth level (μg/g), adjusted for type of tooth, was 2.8 for the lowest lead group, 6.0 for the medium group, and 11.0 for the high group. Blood lead level (μg/dl) for a subsample: Mean +1S.D. +2S.D. 13.1 18.2 23.3 (Range = 6 – 43; next highest = 27; 6.5% above 20)
McMichael et al., 1994	Cross-sectional	No	Tooth and blood Pb Mean tooth level (μg/g) is followed by mean lifetime blood level (μg/dl); lower group = 4.2/11.6; medium group = 9.0/16.7; higher group = 16.7/21.4
Fulton et al., 1989	Cross-sectional	Yes	Tooth and blood Pb Mean tooth level (μg/g) is followed

(continued on next page)

Table 1 (*continued*)

Reference	Type study	Does the sample qualify as "low lead" level?	Blood lead levels (μg/dl)
			by mean blood level (μg/dl) for "tooth givers" = 9.3/11.5
Needleman et al., 1979	Cross-sectional	Not even close	Tooth and blood Pb Mean tooth level (μg/g) for lowest group, < 5.1; mean for highest group, > 27.0. Blood lead level (μg/dl), 4–5 years before shedding teeth, given for two subsamples, one low and one high on dentine lead. Low lead level Mean +1S.D. +2S.D. 23.8 29.8 35.8 High lead level Mean +1S.D. +2S.D. 35.5 45.6 55.7 (Maximum = 54.0)
Winneke et al., 1983	Cross-sectional	No	Tooth and blood Pb Mean tooth level (μg/g) for sample = 6.2 (Range = 1.9–38.5); mean blood Pb level (μg/dl) provided for 72% of sample = 14.3 (Range = 6.8–33.8)
Bergomi et al., 1989	Cross-sectional	Yes	Tooth and blood Pb Tooth level (μg/g) is followed by blood level (μg/dl); Mean 6.0/11.0; 95th percentile 12.7/17.9
Pocock, Ashby, & Smith, 1987	Cross-sectional	No	Tooth lead only Tooth level (μg/g): Low group < 1.5 High group > 9.6 Maximum = 34
Hansen, Trillingsgaard, Beese, Lyngbye, & Grandjean, 1989	Cross-sectional	No (tooth Pb is high despite low blood Pb)	Tooth lead and blood lead Tooth level (μg/g): Total mean = 10.7 Mean low group = 3.2 Mean high group = 26.8 Range = 0.4–168.5 Blood level (μg/dl) mean = 5.1

(when the effects were significant) were not due to the inclusion of numerous children with moderate to high BLLs? Indeed, that seems to be the case for at least two studies. In Dietrich, Berger, Succop, Hammond, and Bornschein's (1993) investigation, three groups of children classified by BLL (all between 0 and 20 μg/dl) earned mean WISC-R Performance IQs of

about 90 ± 2. The only group that differed markedly (earning a mean Performance IQ of about 85) was the group with the highest lead level (> 20 μg/dl). Similarly, Hatzakis et al. (1989) reported nearly identical mean adjusted WISC-R Full Scale IQs (90 ± 1) for their two subsamples with the lowest BLLs (≤ 14.9 and 15.0–24.9 μg/dl), although the three samples with the highest mean values (25.0–34.9, 35.0–44.9, and ≥ 45.0) μg/dl) scored lower, each earning an average IQ of about 85 ± 2. How many other studies among the 26 may not have found a significant relationship between BLL and IQ loss if the analyses were truly limited to children with lead levels of about 20 μg/dl or below?

Nevertheless, am I justified in examining the 26 studies in the three meta-analyses one at a time? Nation and Gleaves (2001) believe that my "basic strategy of reexamining the original 26 studies in terms of 'positive or negative outcomes' is misguided ... [and] arguably is moving science backward rather than forward." They rely on Rosenthal's (1994) description of meta-analysis as an "ethical imperative" to support their accusation that I was wrong to reexamine each component study. Perhaps, as McLean (1995) suggested to me (J.E. McLean, personal communication, May 19, 2000), they ought to have consulted one of the classic texts on meta-analysis (Glass, McGaw, & Smith, 1981) before implying that one can enter garbage into a meta-analysis and wind up with gold. Glass et al. (1981) stated: "An important part of every meta-analysis with which we have been associated has been the *recording of methodological weaknesses in the original studies* and the examination of their relationship to study findings. Thus, the influence of study quality on findings has been regarded as an *empirical a posteriori question*, not an a priori matter of opinion or judgment used to exclude large numbers of studies from consideration" (p. 22, italics mine).

Personally, I believe that the blind acceptance of the results of meta-analyses, and the simple conclusion that "the converging clinical and preclinical data indicate that even low level lead exposure can have a profound effect on public health" (Nation & Gleaves, 2001) — without even verifying whether the component studies in the meta-analyses used samples of children with uniformly low BLLs — moves science in the wrong direction.

6. Lead levels in more recent studies

The 26 studies included in the three meta-analyses were published between 1979 and 1994, ranging from Needleman et al. (1979) to McMichael et al. (1994). Needleman and Bellinger (2001) criticized me because my "survey ends with papers published in 1993 (sic)." Are there any well-controlled research investigations conducted after the meta-analyses were published that provide clear-cut support of a relationship between low lead level and IQ loss? I went through the more recent studies to address this question. I have already mentioned one newer study (Needleman et al., 1996), which not only included numerous children who had moderate to high lead levels when they were younger, but also identified a significant association between lead level and IQ gain. In the Port Pirie study, Tong, Baghurst, McMichael, Sawyer, and Mudge (1996) related lifetime lead exposure to WISC-R IQs at ages 11–13 years. They found a loss of 3 IQ points to be associated with an increase in BLL from 10 to 20 μg/dl. However, their sample does not qualify as "low lead." The

lifetime average blood lead at age 7 was 17.8 μg/dl. Corresponding values for +1S.D. and +2S.D. were 23.6 and 29.4, respectively. Similarly, other Port Pirie studies not included in the meta-analyses (e.g., Tong, Baghurst, Sawyer, Burns, & McMichael, 1998) are not based on "low-lead" samples.

Wasserman et al. also conducted several recent studies, most of which focused on cognitive rather than behavioral variables. Wasserman et al. (1994) found an estimated loss in the McCarthy Scales (McCarthy, 1972) General Cognitive Index (GCI) of 3.8 points associated with an increase of blood lead from 10 to 25 μg/dl. However, her sample of 4-year-olds was a mixture of two subsamples, one with a mean BLL of 9.6 μg/dl (unexposed group) and the other with a mean of 39.9 μg/dl (exposed by a smelter). Overall, their sample cannot nearly be classified as having low blood level, although her graphs of GCI loss for different groups classified by BLL does suggest a relationship between the two variables within the 10–20-μg/dl range. Similarly, other studies by Wasserman, Factor-Litvak, et al., which are based on the same sample followed longitudinally on intelligence and other variables, are not investigations of low lead level (e.g., Factor-Litvak, Wasserman, Kline, & Graziano, 1999; Wasserman et al., 1997).

The study by Mendelsohn et al. (1999) cited by Brown (2001) as evidence of linearity between BLLs and IQ did involve low lead levels (mean = 10.3 μg/dl, S.D. = 5.9, range = 0–24.9) and did find a significantly lower Bayley-II (Bayley, 1993) Mental Developmental Index (MDI) for a subsample of infants and toddlers (mean age = 22.9 months, S.D. = 6.6) with a mean BLL of 14.5 vs. a subsample with a mean BLL of 4.8, even after adjustment for covariates. However, subsamples were small (37 and 31, respectively) and infant development, as measured by tests such as the Bayley-II, is not the same construct as childhood intelligence; such studies of lead and cognitive ability were excluded from the meta-analyses. "The conclusion that emerges from longitudinal studies is that preschool tests (especially when administered after the age of two years) have moderate validity in predicting subsequent intelligence test performance, but that infant tests have virtually none" (Anastasi & Urbina, 1997, p. 328). According to Black and Matula (2000), "The predictability of infant test scores to later IQ or academic functioning is low" (p. 81). Therefore, though the findings from the Mendelsohn et al. (1999) investigation are provocative regarding maturational development (not intelligence) and are worthy of follow-up both with larger samples and longitudinally with childhood IQ, the results do not bear on the questions addressed here. Neither do the studies of Ruff et al. of 2-year-olds with moderate levels of BLL between 25 and 55 μg/dl (e.g., Ruff, Markowitz, Bijur, & Rosen, 1996).

I do not see how any of the studies published subsequent to the three meta-analyses add appreciably to the main issue of whether low levels of lead have an adverse effect on children's IQs. Certainly, they would not have altered any of my main arguments against the premature conclusions drawn by some lead researchers about the lethality of low BLLs.

7. Linearity

In view of the study-by-study review of blood levels summarized in Table 1, and a review of the blood levels of the children included in pertinent studies conducted more recently, there

seem to be a relative dearth of investigations devoted exclusively or nearly exclusively to children with BLLs ≤ 20 μg/dl. The arguments in favor of a significant relationship between low BLLs and IQ loss must rest, therefore, on the case for the linearity of the relationship. If the relationship is linear, then one can legitimately infer a BLL–IQ association from data that span a wide range of BLLs (e.g., <10, 10–20, 21–30, and 31–40 μg/dl).

Needleman and Bellinger (2001) deal with the crucial issue of linearity in a few brief sentences, basing their entire argument on Schwartz' (1993, 1994a) nonparametric analyses of Needleman's own data. This technique involves a general approach called "kernel smoothing" and led Schwartz (1993) to reach the astonishing conclusion that the dose–response relationship of blood lead to McCarthy scores, based on Bellinger–Needleman data, "has no threshold down to blood levels of 1 μg/dl" (p. 237). It is nice that Schwartz and Needleman–Bellinger have so much confidence in the nonparametric kernel smoothing technique that somehow extrapolates down to the most minute amount of blood lead. However, despite the statistical facade of using the LOWESS approach that "fits a linear regression within each window rather than just taking an average" (p. 239), these researchers are making estimations from smoothed data sets and are seeing what they want to see. Schwartz (1994a, Fig. 3) presented a plot of IQ loss vs. mean blood lead from eight studies and concluded that, "A trend toward higher slopes at lower mean BLLs is evident" (p. 50). To whom? Not to me, but certainly to those who are already convinced of a linear dose–response relationship before they even analyze the data. There is nothing compelling about their statistical techniques that rely so heavily on estimation and smoothing, or on Schwartz' (1993, 1994a) use of concordance in effect sizes to buttress his arguments. In fact, meaningful effect sizes can be achieved by chance. Barnette and McLean (1999) found that the proportion of random effect sizes in a large Monte Carlo experiment equaling or exceeding Cohen's (1988) criteria for small, medium, and large effect sizes were 0.804, 0.245, and 0.084, respectively. That is to say that, you can achieve a small effect size merely by chance 80% of the time and a medium effect size 24.5% of the time.

Wasserman, Factor-Litvak, et al. have published data that are suggestive of a linear relationship between lead level and IQ at low doses, yet, they are appropriately cautious about this complex issue. Wasserman and Factor-Litvak (2001) state: "Kaufman makes an excellent point and we agree that there are no data to support a linear dose–response relationship at every step on the exposure/IQ curve .... [F]ew studies have sufficient children with very low or very high blood lead concentrations. Any estimate of association for lead levels at very high or very low concentrations of BPb would therefore be imprecise." There may be a linear dose–response relationship at low BLLs, but it is preposterous to rely so heavily on a nonparametric technique that can make claims at the 1–2-μg/dl range (Schwartz, 1993), much less the 10–20-μg/dl range. This key question of linearity, on which so much public policy decisions have rested, remains an open scientific question for further investigation. It is not a done deal, easily answered by statistical manipulations, as Nation and Gleaves (2001), Needleman and Bellinger (2001), or Schwartz (1993, 1994a) would have us believe.

Indeed, statistics can be used to provide "incontrovertible" proof on both sides of the coin. Schwartz (1993) makes an apparently compelling statistical argument for linearity, but

he is no more compelling than Marais and Wecker (1998), who argue that the significant relationships between lead level and IQ loss is probably an artifact of biased coefficient estimates, stemming from omitted variables, as well as unreliable variables that are included in the regression model. Based on a statistical technique referred to as "regression estimation with auxiliary information" and an array of derived formulas, Marais and Wecker (1998) reanalyzed data from four often-cited lead–IQ studies (e.g., Fulton et al., 1987; Needleman et al., 1979). They state, "When mother intelligence is measured with error and father intelligence is omitted ... [t]he result is a spurious regression relation between lead level and child intelligence" (Marais & Wecker 1998, p. 495). Based on their reanalyses of data from the four lead–IQ studies, Marais and Wecker (1998) concluded that, "we demonstrated, using published values of auxiliary parameters, that bias-corrected estimates of the effect of lead on IQ are reduced in size and are not significantly different from 0" (p. 500).

However, I am no more impressed with Marais and Wecker's (1998) arguments based on statistical manipulations and estimations than I am by Schwartz's (1993) statistical manipulations and smoothing techniques. I have not, therefore, used the Marais–Wecker statistical arguments to support my contentions about the relatively unreliable measures of mother's IQ that have been used in many lead–IQ studies or the need to test fathers in these studies (see Kaufman, 2001, and later discussion in this paper); and I do not accept Needleman and Bellinger's (2001) reliance on Schwartz's statistical machinations to defend their crucial linearity argument. These decisions rest on logic, accumulated research data, and on improved future studies with large numbers of children with low BLLs, not on formulas and figures produced by statisticians and economists.

8. Multiple comparisons

I criticized some of the lead–IQ studies for not taking into account multiple comparisons when reporting the results of their studies. The problem with this approach, as I explained, is that investigators take advantage of the chance errors that are likely to occur when many comparisons are made at once. Nation and Gleaves (2001) took exception to this point, chiding me for treating the topic as if there was only one perspective, and spewing forth an array of statistical arguments. While they are correct that I did not allow for the possibility of alternative viewpoints, I believe that they are missing the boat by focusing so much on the statistical side of the argument. There is also the conceptual side, an aspect of the problem that I did address in my initial paper (Kaufman, 2001): "It is simply bad science to conduct 'multiple analyses' and then interpret only the ones that give the answers the researchers were seeking. The type of research that involves conducting many analyses at once, and then picking and choosing the analyses the experimenters like best, is known informally as a 'shotgun approach'."

It is this conceptual problem, the "seek and ye shall find" aspect of research, that is common among some teams of lead–IQ investigators, including the team that believes that my concerns about making multiple comparisons "are once again colored by [my] failure to understand the context from which the 26 studies [I review] were drawn" (Needleman &

Bellinger, 2001). These researchers believe that it is fine to focus only on age 2 because of the crucial role played by this age group from a developmental perspective. No, I understand the developmental psychology and neuropsychology quite well. I also knew that the Needleman–Bellinger team had identified lead level at age 2 years as a significant correlate of IQ loss in a previous study. That does not give them the right to focus only on the one age that produced significance and ignore the six other ages as if they did not exist. Suppose that age 2 did not produce significant results, but striking evidence of IQ loss was observed at every other age, from the mother's blood cord level to the level at age 10 years. Or suppose that the only significant lead–IQ relationships occurred at ages 18 months and 57 months. Does anyone believe for a minute (or even a nanosecond) that Bellinger, Stiles, and Needleman (1992) would have continued to focus on the results at age 2 and to have declared their results nonsupportive of their contentions about the evils of low doses of lead?

In fact, how did this team of researchers deal with unexpected findings in some of their other research reports? In Needleman's highly publicized study about lead level and delinquent behavior (Needleman et al., 1996), he and his colleagues conducted an unusually large number of multiple, simultaneous comparisons. Many of them did not produce significant results, such as the four factors measured by an attention battery, the seven scores yielded by the Neurobehavioral Evaluation System, and children's behavioral problems reported by parents and teachers at age 7. When adjusted for covariates, the subjects' self-reported delinquency (SRD) did not discriminate significantly between the high-lead and low-lead group. This negative finding was reported in the following positive light: "Subjects' SRD at 11 years (Table 5) was significantly related to bone lead without covariate adjustment ($P = .04$). This finding was slightly altered by entering covariates ($P = .07$)" (Needleman et al., 1996, p. 366). The authors did not even mention, in their discussion of their Table 5, that no significant difference emerged, with or without adjustment, for self-reported antisocial behavior at age 7.

Furthermore, Needleman et al. (1996) found one significant finding, after adjusting for covariates, in the IQ data, and it was opposite the prediction of IQ loss: The high-lead group earned adjusted mean Verbal IQs on the Wechsler Intelligence Scale for Children-Revised (WISC-R; Wechsler, 1974) that were 4.5 points higher than the means for the low-lead group. How did this research team treat the many nonsignificant findings in the study or the one, very intriguing, IQ gain by the low-lead subjects? By reconsidering their long-held positions about the evils of low BLLs? Of course not. By trying to bury the results as much as possible. Naturally. Though many cognitive and neuropsychological variables were included in the investigation, the title of the article is the eye-grabbing, "Bone Lead Levels and Delinquent Behavior." The Abstract makes no mention of the significant IQ gain that was opposite to prediction or of any of the myriad nonsignificant comparisons that were reported in the text of the article. This approach — focusing on the findings that support a researcher's contention and damn the contradictory results — is the essence of why my multiple comparisons criticism is valid in a strongly conceptual way even if the statistical arguments are not clear-cut.

Nevertheless, the aforementioned scientific travesty of focusing on significant findings and burying nonsignificant and disconfirming results is at least discernible from the published work, permitting readers of the study to detect the subjectivity and selection bias.

That has not always been the case, as Ernhart (1993) indicates in her Letter to the Editor concerning Needleman's misconduct investigation. Quoting the Inquiry Panel at the University of Pittsburgh, Ernhart (1993) stated that they concluded: "The issue of concern, however, is not the choice of confounders to consider for a model .... The problem lies in the fact the six different models that were fit to the data [led] to different results and that these alternate results were not discussed in the 1979 *NEJM* paper" (p. 172). Conducting multiple analyses and choosing to publish the one that yields the most satisfying results represents an analogous type of multiple comparisons error, but one that required a formal hearing to detect.

Other research teams are also guilty of variants of this poor methodology. Phelps (1999), in a well-reasoned critique of the lead–IQ literature, makes the following salient observation about the research team that has conducted the Cincinnati Lead Study (e.g., Dietrich et al., 1990, 1993):

> Another essential consideration is the correction for Type I errors (false positives) when a single prospective data set is analyzed multiple times. Running analyses year after year on the same children is a common practice in toxicology explorations; yet, few scientists attempt to correct for false positives .... One research team, for example, completed a total of 286 regression analyses on 297 children followed from birth to 6.5 years of age. — (p. 480)

In their response to my point about multiple comparisons, Wasserman and Factor-Litvak (2001) state, "The sequential measurements of both blood lead concentrations and IQ are strengths of the prospective studies." I agree completely with this insightful observation. The multiple measurements do promote good science, and have the potential to permit objective evaluation and reporting of the data, as evidenced by the topnotch prospective studies conducted and written up by Wasserman, Factor-Litvak, et al. (e.g., Wasserman et al., 1994, 1997). However, multiple measurements, multiple analyses, and multiple studies also have the propensity for subjective interpretation and abuse of data. Such abuse — whether deliberate or unwitting — has been prevalent within the lead literature.

9. Control of parental IQ and other confounders

Several issues were raised by respondents regarding the general issue of the control of confounders, especially those that are poorly measured or not measured at all. First, I discuss parents' IQ, which I considered to be poorly measured, and then I deal with uncontrolled or poorly controlled confounders other than parents' IQ.

10. Parental IQ

Parents' IQ is arguably the most important confound to control in the lead–IQ studies, but many investigators have done a poor job of measuring this key variable. Needleman and Bellinger (2001) admit that "the instruments used to measure parental intelligence in many lead studies are neither the most sensitive nor accurate." Despite this rare, almost

unprecedented, admission of imperfection in the body of lead literature, they proceed to argue that because the IQ distortion might be unsystematic, the use of a poor measure of parents' IQ "will tend to *underestimate* the effect of lead" (italics theirs). What absolute rubbish! As Wasserman and Factor-Litvak (2001) correctly state: "Poorly measured parental intelligence results in poor control for this potentially confounding variable; such poor control *may spuriously increase the association between lead and IQ (since parental IQ may be inversely associated with lead exposure and positively associated with childhood IQ)*" (italics mine). Maternal IQ indeed qualifies as a true confound in many lead–IQ studies, with the tests used to measure this construct invariably correlating positively with children's IQ and negatively with lead burden (e.g., Ernhart, Morrow-Tlucak, Wolf, Super, & Drotar, 1989; Fulton et al., 1987; Schroeder, Hawk, Otto, Mushak, & Hicks, 1985). The better the measurement of this key confound, the greater the amount of confounding variance (NOT error variance) will be pulled out of the equation in most lead–IQ studies. The result will usually be to decrease, not increase, the number of IQ points attributed to lead.

Although Wasserman and Factor-Litvak (2001) agree with my contentions about the potential impact of poorly measured maternal IQ, they dispute my belief that the possible poor measurement represents a problem in the lead–IQ literature by noting that all of the prospective studies used some measure of maternal IQ and "most find rather similar associations between lead exposure and childhood IQ." First, whereas Needleman and Bellinger (2001) believe that I was wrong to limit myself to the 26 studies included in the meta-analyses, Wasserman and Factor-Litvak (2001) implicitly suggest that I included too many studies in my pool of research; as in their defense of using a variety of measures of maternal IQ that may differ in quality, many of their points in their response to my article focus on the prospective studies, as if the cross-sectional studies are not worthy of discussion. Yet, the conclusions drawn from the three meta-analyses are based on both types of studies. Perhaps the prospective studies (including the ones by Wasserman et al., conducted after the meta-analyses were published) all measured maternal IQ, but that was not true for the 26 studies of prime interest here. In fact, of the 26 investigations, parents' IQ was measured inadequately (i.e., by a picture vocabulary test or a group-administered test) in nine studies, or not at all (eight studies). Therefore, in 17 of the 26 investigations, maternal IQ measurement was a methodological problem.

Was the problem inconsequential, as Wasserman and Factor-Litvak (2001) suggest? What is their evidence that most of the studies find similar associations between maternal IQ (no matter how measured) and both lead exposure and children's IQ? Different investigators provide different data in their articles, sometimes making comparisons across studies difficult. From my examination of the studies, picture vocabulary tests tend to correlate in the mid-.30s with children's IQ (Ernhart et al., 1989; Hawk et al., 1986; Schroeder et al., 1985), whereas the more reliable oral vocabulary tests or other conventional measures of verbal intelligence correlated in the mid-.40s or higher (Fulton et al., 1987; Schroeder et al., 1985). Although McMichael et al. (1994) did not report a correlation coefficient between the complete WAIS-R that they administered to the mothers and their children's WISC-R IQs, they did present mean children's Full Scale IQs for three categories of mother's IQs. Mothers with WAIS-R Full Scale IQs less than 85 had children

whose mean Full Scale IQ was 98.7, which is more than 1S.D. below the mean of 116.3 for children whose parents had IQs greater than 100. That degree of relationship is too extreme to reflect a mid-.30s correlation, but is more consistent with coefficients in the .40s or .50s.

One of the big criticisms of the Peabody Picture Vocabulary Test-Revised (PPVT-R; Dunn & Dunn, 1981) and other picture vocabulary tests is the chance element that accompanies the selection of one out of four responses. That heavy dose of nonsystematic error variance is, indeed, null-biasing, to use a favorite phrase of Needleman and Bellinger (2001). In contrast, higher correlations between parents' and children's IQs are far more likely when using more reliable measures of parents' intelligence, such as short forms of Wechsler's scales, oral vocabulary–Raven's matrices combinations, or brief intelligence tests that have been specifically normed on representative populations, such as the Kaufman Brief Intelligence Test (K-BIT; Kaufman & Kaufman, 1990) or the Wechsler Abbreviated Scale of Intelligence (WASI; The Psychological Corporation, 1999).

The methodological problem concerning the measurement of maternal IQ is so easily solvable in future research studies, and such an apparent flaw in many prior studies, that it surprises me to find such resistance or defensiveness on the part of some researchers. As Brown (2001) states: "Maternal PPVT score as the only measure of parental intelligence, as is often the case in developmental research, is simply inexcusable …. Sadly, to the uninformed, use of the PPVT may provide research with an unjustified façade of control. More sadly, it appears to provide the same façade to presumably informed reviewers and editors."

Even the authors of the PPVT-R stated unambiguously in the test manual, "The PPVT-R is designed primarily to measure a subject's receptive (hearing) vocabulary …. It is not, however, a comprehensive test of general intelligence" (Dunn & Dunn, 1981, p. 2); and no one has offered a good explanation for the systematic elimination of fathers from the parents' IQ equation, except for a single investigation (Lansdown et al., 1986) that tested both parents, even though the combination of father's and mother's IQs correlates substantially higher with children's IQs than does either one alone (Kaufman, 1990). Needleman and Bellinger (2001), as well as Nation and Gleaves (2001), have offered an array of statistical arguments to defend their position, once again completely missing the forest for the trees.

11. Uncontrolled confounds other than parents' IQ

Much was made by the respondents of my concern about unmeasured or poorly measured confounds (apart from parents' IQ), such as otitis media. Wasserman and Factor-Litvak (2001) took me to task for my criticisms of the HOME (Caldwell & Bradley 1984), and I am convinced by their data-based arguments; I was unreasonably harsh in my anti-HOME comments. Whereas I did not consider any of the studies that used the HOME, or any interview technique that assessed parenting and home-environment variables, to have a shortcoming in their measurement of SES and related confounds, I do stand corrected for my unfair remarks about the HOME.

My use of the otitis media research was intended to be illustrative of the kinds of nonlead variables that have sometimes shown significant relationships to IQ and that might also plausibly relate to children's lead levels. In general, children from low SES backgrounds are more susceptible to higher BLLs, and there is some evidence that low SES children may be more susceptible to chronic secretory otitis media (CSOM) than high SES children and that the low SES children receive less medical treatment for their disease (Webster, Bamford, Thyer, & Ayles, 1989): "Examining children at age 7 years, the National Child Development Study ... reported 1 in 12 children with signs of past or present middle ear disease, with children from social class V more than twice as likely to have purulent, discharging ears as those from social class I .... In contrast to the figures for prevalence, children in social class I are twice as likely to receive surgical treatment for CSOM than are children in lower social groups" (p. 533). Though neither conclusive nor necessarily generalizable, this finding raises the possibility that otitis media may be a true confound in some lead–IQ studies such that its lack of control would reduce — not increase — the effects of lead.

Still, otitis media is merely an illustration of other variables whose control might not be null-biasing but might truly remove variance in IQ that is attributed to lead level. I was not holding up the otitis media research as paragons of methodology, and noted that the "topic has produced numerous research investigations that vary in quality and, like the lead–IQ studies, have produced conflicting results" (Kaufman, 2001); Wasserman and Factor-Litvak (2001) are indeed correct to point out that no otitis media study "controlled for the important social determinants of IQ, including parent skill and parental intelligence." However, medical variables, both prevalence and treatment, are conceivably related both to IQ and social class (and, therefore, potentially to lead level), as are subtle parenting variables, prenatal care, and prenatal abuse. Their control in some of the lead–IQ studies would, consequently, reduce the number of points of IQ loss attributed to lead.

Furthermore, the respondents are correct to state that the inclusion of any number of additional confounds in lead–IQ studies, overall, will have an unpredictable effect on the magnitude of the IQ loss attributed to lead. Nation and Gleaves (2001) reflect the beliefs of most respondents when they state, "although the actual effects of lead on IQ may be lower than those obtained because of uncontrolled variables, the actual effects also may be higher." The issues involved are far more complex, both statistically and methodologically, than I made them seem in my article (see, especially, Nation & Gleaves, 2001; Needleman & Bellinger, 2001). Whereas I believe that it is incumbent on future lead–IQ researchers to try to identify, and control, many more potential confounds than has been done in the past, I concede that the outcome of such confound control regarding the IQ loss attributed to lead is unpredictable. Parenthetically, however, note that 12 studies were considered to have a shortcoming in the control of confounds (Kaufman, 2001, Table 1), and all of these studies used a global measure of SES such as parents' educational attainment rather than more specific measures such as the HOME. I believe that the failure of those 12 teams of investigators to control for more specific aspects of SES are likely to have exaggerated the effect of lead level on IQ loss — a belief that is borne out by the frequent reduction in IQ loss, when adjusted for covariates, in those better-controlled studies that did measure specific aspects of a child's sociocultural environment

and parent–child interactions (e.g., Baghurst et al., 1992; Bellinger et al., 1992; Ernhart et al., 1989).

Furthermore, Hebben (2001) makes the key point that, "Studies have shown that while covariates such as heritability, parenting, and social factors account for over 50% of the variance in child cognitive ability, lead accounts for only 1–2%." Wasserman and Factor-Litvak (2001), though convinced that the accumulated research findings "suggest that exposure to lead does have adverse consequences for childhood development," do concur, however, "that the deficits are likely to be small in comparison to the contribution of measured social factors."

12. Issues concerning IQ measurement

Several of the issues dealt with by the respondents dealt with the IQ construct or its measurement. These topics — which include quality control of the measurement of children's IQ, interpretation of a fraction of an IQ point, and societal impact of the loss of a few IQ points — are treated in the sections that follow.

13. Quality control

I have trained clinical and school psychologists in the administration, scoring, and interpretation of IQ tests for a generation; I supervised the nationwide standardizations of the McCarthy Scales and WISC-R a generation ago; I have collaborated in research studies with licensed psychologists for three decades; and I have coauthored with my wife, Nadeen, several IQ tests for children, adolescents, and adults. Plain and simple, even the best examiners are prone to clerical errors, make administrative decisions that can compromise the validity of the obtained scores, and must religiously and systematically check their record forms for accuracy to ensure valid scores. Training in administration requires much hands-on, one-on-one, supervised observation, and even then some otherwise good clinicians make technical errors of administration, careless scoring errors, or judgment mistakes in the establishment and maintenance of rapport.

When research investigations rely on IQ as a key variable, they should ensure the validity of the data by using highly qualified and trained examiners; they should employ a check-scoring technique to reach consensus on the scoring of ambiguous responses and to eliminate careless mistakes; and they should report all of this pertinent information in their research publications. Am I guilty, as Brown (2001) states, of turning an error of omission into an error of commission? When these examiners report the precise qualifications of the nurses who drew the blood for the assessment of lead level, their omissions about the qualifications of the IQ examiners and the procedures followed to ensure valid IQ data speak volumes to me about their lack of appreciation for the nuances of individually administered IQ tests. Without such appreciation, errors invariably follow; and these types of errors are completely avoidable.

I really do not care if research studies in areas other than lead–IQ suffered from even more quality control issues than the lead studies (Brown, 2001) or if the quality-control

errors are likely to be random and, therefore, to underestimate the influence of lead (Nation & Gleaves, 2001; Needleman & Bellinger, 2001). IQ is the main outcome variable in the 26 studies in the meta-analyses and the findings pertaining to possible IQ loss have been instrumental in setting public policy. If you are going to measure IQ, and if it is so important as an outcome variable, then get it right. Perhaps my criteria for proper administration and scoring of IQ tests is more stringent than the criteria that Needleman and Bellinger (2001) cite as being proposed by the American Academy of Clinical Neuropsychology. I can acquiesce to the latter criteria, even though I am strongly opposed to bachelor's level IQ examiners. Yet, a number of the 26 studies fell short of providing the specific training and supervision of their examiners by licensed doctoral-level psychologists (Kaufman, 2001). In addition, despite the protestations of Needleman and Bellinger (2001), the statistical arguments of Nation and Gleaves (2001), and the prove-it-to-me stance of Brown (2001), I am not alone in my contentions about quality control being a potential problem. Wasserman and Factor-Litvak (2001), who perform prospective investigations in the lead-research trenches (in contrast to my IQ-trench experiences) state: "Kaufman correctly points out that many of the studies do not report quality control assessments for key measured variables. This is unfortunately common and inexcusable in clinical epidemiology."

Needleman and Bellinger (2001) consider my requirements for proper IQ assessments to be "a formidable criterion that most researchers would find difficult to summon the budget and personnel to meet." I believe that if some lead researchers had as much passion for the accurate measurement of their outcome variables as they have for promoting the translation of the outcomes of their studies to public policy, then both the money and the personnel would be readily available from the funding agencies. These agencies do, in fact, provide multimillions of dollars to the investigators of lead-level studies and would undoubtedly find a reserve supply of dollars to address any issue that the researchers deemed important.

14. Interpreting fractions of an IQ point

Brown (2001) states, "Since virtually all authors present group IQ means at least to the first and frequently the second decimal point, I may have missed Kaufman's point." My concern about the fractionation of an IQ point does not come from the reporting of group data and does not dispute the treatment of IQ as an interval measure. I dispute the interpretations that are sometimes given to a fraction of 1 IQ point, some of which border on the surreal and absurd. Here are some illustrations.

- "Applying the mean changes to the cohort of 4.73 million children in California below age 7 . . ., the current ambient concentration of 0.06 $\mu g/m^3$ relates to an average loss of 0.08 IQ points" (Ostro, Mann, Collins, Vance, & Alexeef, 1996, p. 5–10).
- "Using this slope coefficient [of 0.245 IQ points per $\mu g/dl$], we can estimate that a permanent reduction in blood lead concentrations of 1 $\mu g/dl$ will produce a net present value benefit of $1300 per child for the cohort turning 6 years of age each year, for a total benefit of $5.06 billion per year" (Schwartz, 1994b, p. 114).

- "... with IQ, average losses of 0.257 per 1 μg/dl may indeed have enormous societal consequences" (Nation & Gleaves, 2001).

The standard error of measurement of the best IQ tests is about 3 points. Two conscientious examiners will obtain different scores for the children they test, sometimes by several points, merely because they differ in their subjective interpretations of which Verbal responses to query and when to click the stopwatch to denote that a child has completed a Performance item. Also, scientists are interpreting fractions of IQ points as low as 0.08 of a point in terms of billions of dollars and societal consequences? Give me a break!

15. Societal consequences of a few points of IQ loss

I do not believe that the loss of a few IQ points — if the allegation concerning the loss of IQ due to low levels of BLL is true — can have meaningful consequences for society. My arguments in defense of this point (Kaufman, 2001) elicited some misinterpretations of my position. Brown (2001) interprets my statements that IQ measures a limited aspect of functioning and is a narrow concept to mean that "Kaufman claims that [IQ] is not particularly useful," and he refers to "Kaufman's sweeping condemnation of the importance of IQ." Needleman and Bellinger (2001) claim that "[Kaufman] recommends, in particular, Sternberg's unpublished group-administered test (STAT)."

According to these respondents, I believe that IQ is not useful and should be condemned as unimportant and that clinical IQ tests (implicitly, including my own) should be replaced by an unpublished group-administered test by a psychologist (Sternberg, 1984) who has been among the most critical of my Kaufman Assessment Battery for Children (K-ABC; Kaufman & Kaufman, 1983). Right! I also believe that Needleman is going to urge the federal government to raise the "safe level" of BLL to 30 μg/dl!

IQ tests, both the ones I have developed with Nadeen (e.g., Kaufman & Kaufman, 1983, 1990, 1993) and the ones I have helped clinicians interpret (e.g., Kaufman, 1979) have formed an important part of my life's work — and continue to do so (Kaufman & Lichtenberger, 1999, 2000; Lichtenberger, Broadbooks, & Kaufman, 2000). Nevertheless, throughout my years of writing, research, test development, and teaching, I have never lost sight of their limitations. IQ tests are valuable for predicting school achievement, for helping to identify individuals with problems such as specific learning disabilities (Kaufman & Kaufman, in press), for helping teachers identify children's preferred teaching and learning strategies (Kaufman & Kaufman, 1983, chapter 7), for identifying cognitive strengths and weaknesses (Kaufman, 1994), for facilitating neuropsychological assessment (Kaufman, 1990), and so forth. However, I have always argued that the global IQ is the least important result of an IQ test administration, that one must be respectful of the errors of measurement that are necessarily built into it, and that the results of an IQ assessment are only meaningful if they are buttressed by the results of other instruments and interpreted within the context of the person's specific background and behaviors observed during the evaluation. All of these aspects of what I have called "intelligent testing" (Kaufman, 1979, 1994) are what make the IQ test a potentially valuable tool for clinical evaluation, especially when trained examiners are able to interpret the profile of scores

from the perspective of diverse theories. Indeed, the clinician who works in tandem with the IQ test is no less valuable than the test itself. I embrace the value of IQ tests. I do not condemn them or relegate them to subservience to unpublished group-administered tests. Nevertheless, I also embrace their limitations.

IQ tests cannot predict life success. They are limited in what they measure. That is a fact of life. The loss of a few IQ points — an amount of points that is well within a reasonable band of error around the observed score — is inconsequential in terms of meaningful life outcomes. In addition, the global IQ itself, standing naked as a score without the benefit of context or expert interpretation or theory, is the most inconsequential of all. Sternberg's (1985) theory is merely illustrative of the kind of more comprehensive approach to intelligence (relative to the conventional IQ approach) that needs to be employed before any reasonable scientist dares to substantiate the claim that the loss of a few IQ points has societal impact. Indeed, the usual IQ loss attributed to low BLL is quite similar in magnitude to the 2- to 3-point mean gender differences, favoring males, in Verbal IQ and Full Scale IQ that have characterized Wechsler's scales for a half-century (Kaufman, 1990, chapter 6; Seashore, Wesman, & Doppelt, 1950). Yet, Wechsler considered this discrepancy trivial in magnitude and never gave any thought to offering separate IQ norms for males and females.

What IQ tests do, they do well. Their value is greatly enhanced when interpreted by intelligent testers in behavioral and theoretical contexts. However, without measurement of other aspects of intellectual functioning such as creativity, social intelligence, practical intelligence, adaptive behavior, mechanical ability, and the like, the IQ tests offer a limited range of one's overall cognitive functioning. When reduced to a single number plus or minus a few points, it is not sensible to speak of societal impact.

How did all of us who are writing articles for this special issue ever make it successfully into the 21st century in view of the mean BLLs in the US that were found to be 58 μg/dl in 1935 (Kehoe, Thamann, & Cholak, 1935), around 30 μg/dl from the late 1930s to the mid-1950s (Kaplan & McDonald, 1942; Kehoe, Cholak, & Story, 1940; Survey of Lead, 1965) and about 20 μg/dl during the decade of the 1960s (Goldwater & Hoover, 1967; Survey of Lead, 1965; Tepper & Levin, 1972)? By way of contrast, the mean BLL in the US for the population ages 1 year and older in 1991–1994, based on NHANES III data, was 2.3 μg/dl, with a mean of 2.7 μg/dl obtained for children ages 1–5 years. (Pirkle et al., 1998).

16. Conclusions

The respondents have made some important comments and criticisms regarding my paper on lead and IQ. Some of the shortcomings of the studies that I listed, even if improved in future investigations, may actually increase the magnitude of IQ loss attributed to lead. Brown (2001) is undoubtedly correct in pointing out that the lead–IQ studies are generally far superior methodologically to much of the research on other toxicological/teratological agents. He concludes, for example, that "most of the studies Kaufman evaluated are models of design rectitude when compared with frequently cited research claiming to show adverse effects of maternal smoking on children's development." He also chides me for seeking

perfection in human research when such perfection is impossible: "Strictly speaking, [Kaufman] is right, but the position is Humean skepticism to the point of nihilism" (Brown, 2001).

I agree that the lead–IQ research, as a whole, is among the best of its type. I also agree that I overstepped reasonable bounds by implying, unwittingly, that no study can ever meet my standards. The obvious solution would be to stop the research altogether if the standards cannot be met in the real world. Yet, that was not my intention, so I must backtrack to a more tenable position. At the same time, I cannot accept the equally unreasonable conclusions reached by both Needleman–Bellinger and Nation–Gleaves that low BLLs cause IQ loss — case closed — so let us stop wasting time with a scientific debate that has only one side (the side of the angels, their side).

There is room for scientific debate. The issue has not been decided by the bulk of research studies that have been done. It is not even close. The causality issue, though blindly accepted by Needleman–Bellinger and Nation–Gleaves without question, is more complex than they are able to admit. This point is made convincingly by Brown (2001) who cites an excellent recent article on the topic by Reynolds (in press) that I have also read; I agree with Brown (2001) that Reynolds' article "should be must reading for researchers and policy makers," as should a related paper by Reynolds (1999) on the causality between smoking and low birth weight.

Hebben (2001) agrees with me that the issues about the relationship between lead level and IQ loss remain a topic for scientific study and debate, stating that "Kaufman's review presents a number of important and convincing points that should induce readers to maintain a skeptical eye toward the evidence for the behavioral toxicological impact of low lead levels on IQ." Wasserman and Factor-Litvak (2001) also see merit in scientific debate, noting that, "To the degree that [Kaufman's] paper calls attention to inaccuracies and guides future research, it will make a contribution to the field."

That is my goal — to have future researchers be more aware of the limitations of the existing literature and make conscientious efforts to measure parents' IQ and socio-economic confounders with state-of-the-art instruments and techniques, to focus on methodological issues such as the making of multiple comparisons to ensure that they do not simply find the answers that they seek, and to identify samples that truly are low in BLL such that conclusions about low BLL will be based on appropriate samples. I was delighted when G.A. Wasserman (personal communication, February 15, 2000) telephoned me to ask my suggestion for an instrument to measure parents' IQ in a new study she was designing; I was pleased to suggest that she use either the 2-subtest or 4-subtest version of the new WASI (The Psychological Corporation, 1999), a test she had not heard of at that time.

Hebben (2001) makes some salient points about the way research on groups of children has been applied by the public to individual children: "I have seen low lead level "causally" linked to mental retardation (even in the face of serious birth injury), autism, specific arithmetic disability, and meaningful large differences between VIQ and PIQ (even in the face of bilingualism." Ruff (1999) who, like Wasserman and Factor-Litvak, conducts lead research on children (e.g., Ruff et al., 1996), has a perspective similar to Hebben's regarding the application of the group data to individuals and to the real-life

effects of these applications. Ruff (1999) objects to the way in which research results are sometimes presented to the public, especially regarding the need to counsel the parents of children with slight lead elevations who have been scared by newspaper reports of the irreversible dangers of low blood lead. She also has concerns about the way lawsuits infer specific problems in individuals based on the results for populations, inferences that sometimes border on the absurd, as Hebben (2001) recounts from her own forensic experience. Ruff (1999) asks "why do parents, lawyers, and other concerned parties usually think of research results as relevant to individuals?" (p. 43). The answer is readily clear from the responses of Nation and Gleaves (2001) and Needleman and Bellinger (2001), who uncritically accept that the research findings demonstrate clearly that low levels of lead cause IQ loss, attentional problems, behavioral problems, and so forth; who interpret the findings for populations in terms of their effects on specific individuals; who use the findings to shape public policy, always striving to lower the safe limit as far as the government officials will allow; and who foster the scare tactics by emphasizing the dire societal consequences of a little bit of lead in the blood.

The studies to date have shortcomings, some of which are of more concern than others. The problem with the measurement of parents' IQ is serious and compromises the results of a number of existing studies; it is, however, a problem that is easily remedied for future investigations. The problem of multiple comparisons is likewise substantial, but not in terms of the Type I vs. Type II arguments that really have no one right answer. The multiple comparisons problem is more of the "selection bias" issue Needleman and Bellinger (2001) accuse me of in my paper (Kaufman, 2001), but Needleman is extremely guilty in his own reporting of research methods and results (Ernhart, 1993; Needleman et al., 1996; Phelps, 1999). Emphasizing the findings that fit one's theory and burying the ones that do not, or conducting as many regression analyses as one has subjects in the course of numerous publications, represents a serious and ethical abuse of scientific objectivity.

In addition, perhaps the biggest problem of all in the inference of IQ loss at low BLLs is the fact that many of the studies conducted, on both animals and children, are based on samples with moderate BLLs or on samples that include children with both low and moderate BLLs (see Table 1). To make inferences about the IQ loss as one goes from 10 to 20 μg/dl requires assurance that the relationship between lead level and IQ is linear, and that evidence is simply lacking.

There are still unknowns in the lead–IQ equation, there is more that we can learn by improving the methodology within the current state of the art (and without living in an ideal world), and there is still room for scientific debate on the topic. It may be that Nation and Gleaves (2001) see a clear visual image of a trout in the milk. However, on closer inspection, the trout is probably suffering from moderate to severe lead poisoning, not from low doses of lead in its system.

Acknowledgments

I am extremely grateful to the following psychologists for their assistance with this article: James C. Kaufman, Nadeen L. Kaufman, James E. McLean, and LeAdelle Phelps.

References

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice-Hall.

Baghurst, P. A., McMichael, A. J., Wigg, N. R., Vimpani, G. V., Robertson, E. F., Roberts, R. J., & Tong, S.-L. (1992). Environmental exposure to lead and children's intelligence at the age of seven years: the Port Pirie cohort study. *New England Journal of Medicine*, *327*, 1279–1284.

Banks, E. C., Ferretti, L. E., & Shucard, D. W. (1997). Effects of low level exposure on cognitive function in children: a review of behavioral, neuropsychological and biological evidence. *Neurotoxicology*, *18*, 237–282.

Barnette, J. J., & McLean, J. E. (1999, November). Empirically Based Criteria for Determining Meaningful Effect Size. Paper presented at the annual meeting of the Mid-South Educational Research Association, Point Clear, AL.

Bayley, N. (1993). *Manual for the Bayley Scales of Infant Development* (2nd ed.). San Antonio, TX: The Psychological Corporation.

Bellinger, D. C., 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.

Bergomi, M., Borella, P., Fantuzzi, G., Vivoli, G., Sturloni, N., Cavazzuti, G., Tampieri, A., & Tartoni, P. L. (1989). Relationship between lead exposure indicators and neuropsychological performance in children. *Developmental Medicine and Child Neurology*, *31*, 181–190.

Black, M. M., & Matula, K. (2000). *Essentials of Bayley Scales of Infant Development-II assessment*. New York: Wiley.

Brown, R. T. (2001). Behavioral teratology/toxicology: how do we know what we know? *Archives of Clinical Neuropsychology*, *16*, 389–402.

Caldwell, B., & Bradley, R. (1984). *Administration manual: home observation for measurement of the environment* (Rev. ed.). Little Rock, AR: University of Arkansas at Little Rock.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohn, J., Cox, C., & Cory-Slechta, D. A. (1993). The effects of lead exposure on learning in a multiple schedule of repeated acquisition and performance. *Neurotoxicology*, *14*, 329–346.

Cooney, G., Bell, A., & Stavron, C. (1991). Low level exposures to lead and neurobehavioural development: the Sydney study at seven years. *Heavy metals in the environment* (pp. 16–19). Edinburgh: CEP Consultants.

Cory-Slechta, D. A. (1997). Relationships between Pb-induced changes in neurotransmitter system function and behavioral toxicity. *Neurotoxicology*, *18*, 673–688.

Daniel, L. G. (1998). Statistical significance testing: a historical overview of misuse and misinterpretation with implications for the editorial policies of educational journals. *Research in the Schools*, *5*, 23–32.

Davis, J. M., Otto, D. A., Weil, D. E., & Grant, L. D. (1990). The comparative developmental neurotoxicity of lead in humans and animals. *Neurotoxicology and Teratology*, *12*, 215–229.

Dietrich, K. N., Berger, O. G., Succop, P. A., Hammond, P. B., & Bornschein, R. L. (1993). The developmental consequences of low to moderate prenatal and postnatal lead exposure: intellectual attainment in the Cincinnati lead study cohort following school entry. *Neurotoxicology and Teratology*, *15*, 37–44.

Dietrich, K. N., Succop, P. A., Bornschein, R. L., Krafft, K. M., Berger, O. G., Hammond, P. B., & Buncher, C. R. (1990). Lead exposure and neurobehavioral development in later infancy. *Environmental Health Perspectives*, *89*, 13–19.

Dunn, L., & Dunn, L. (1981). *Manual for the Peabody Picture Vocabulary Test-Revised (PPVT-R)*. Circle Pines, MN: American Guidance Service.

Epstein, R. J. (1998). Lead paint, toxic torts, and the housing stock: a case study in risk assessment. *Journal of Real Estate Finance and Economics*, *17*, 127–138.

Ernhart, C. B. (1993). Letters to the editor — deliberate misrepresentations. *Pediatrics*, *1*, 171–172.

Ernhart, C. B. (1996). Bone lead levels and delinquent behavior — to the editor. *Journal American Medical Association*, *275*, 1726.

Ernhart, C. B., Landa, B., & Wolf, A. W. (1985). Subclinical lead level and developmental deficit: reanalysis of data. *Journal of Learning Disabilities*, *18*, 475–479.

Ernhart, C. B., Morrow-Tlucak, M., Wolf, A. W., Super, D., & Drotar, D. (1989). Low level exposure in the

prenatal and early preschool periods: intelligence prior to school entry. *Neurotoxicology and Teratology*, *11*, 161–170.

Factor-Litvak, P., Wasserman, G. A., Kline, J. K., & Graziano, J. (1999). The Yugoslavia prospective study of environmental lead exposure. *Environmental Health Perspectives*, *107*, 9–15.

Fergusson, D. M., Fergusson, J. E., Horwood, L. J., & Kinzett, N. G. (1988). A longitudinal study of dentine lead levels, intelligence, school performance and behaviour: Part II. Dentine lead and cognitive ability. *Journal of Child Psychology and Psychiatry*, *29*, 793–809.

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

Fulton, M., Paterson, L., Raab, G., Thomson, G., & Laxen, D. (1989). Blood lead, tooth lead and child development in Edinburgh. In: J. P. Vernet (Ed.), *Heavy metals in the environment* (vol. 2, pp. 68–71). Edinburgh: CEP Consultants.

Fulton, M., Raab, G., Thomson, G., Laxen, D., Hunter, R., & Hepburn, W. (1987). Influence of blood lead on the ability and attainment of children in Edinburgh. *Lancet*, *1*, 1221–1226 (Saturday 30 May).

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage Publications.

Goldwater, L. J., & Hoover, W. A. (1967). An international study of ''normal'' levels of lead in blood and urine. *Archives of Environmental Health*, *15*, 60–63.

Hansen, O. N., Trillingsgaard, A., Beese, I., Lyngbye, T., & Grandjean, P. (1989). A neuropsychological study of children with elevated dentine lead level: assessment of the effect of lead in different socio-economic groups. *Neurotoxicology and Teratology*, *11*, 205–213.

Harvey, P. G., Hamlin, M. W., Kumar, R., Morgan, J., Spurgeon, A., & Delves, H. T. (1988). Relationships between blood lead, behaviour, psychometric and neuropsychological test performance in young children. *British Journal of Developmental Psychology*, *6*, 145–156.

Hatzakis, A., Kokkevi, A., Katsouyanni, K., Maravelias, K., Salaminios, F., Kalandidi, A., Koutselinis, A., Stefanis, K., & Trichopouylos, D. (1987). Psychometric intelligence and attentional performance deficits in lead-exposed children. *Heavy Metals in the Environment*, *1*, 204–209.

Hatzakis, A., Kokkevi, A., Maravelias, C., Katsouyanni, K., Salaminios, F., Kalandidi, A., Koutselinis, A., Stefanis, C., & Trichopolous, D. (1989). Psychometric intelligence deficits in lead-exposed children. In: M. A. Grant, L. D. Grant, & A. L. Sora (Eds.), *Lead exposure and child development* (pp. 211–223). London: Kluwer.

Hawk, B. A., Schroeder, S. R., Robinson, G., Otto, D., Mushak, P., Kleinbaum, D., & Dawson, G. (1986). Relation of lead and social factors to IQ of low-SES children: a partial replication. *American Journal of Mental Deficiency*, *91*, 178–183.

Hebben, N. (2001). Low lead levels and neuropsychological assessment: let's not be mislead. *Archives of Clinical Neuropsychology*, *16*, 353–357.

Kaplan, E., & McDonald, J. M. (1942). Blood lead determinations as a health department laboratory service. *American Journal of Public Health*, *32*, 481–486.

Kaufman, A. S. (1979). *Intelligent testing with the WISC-R*. New York: Wiley.

Kaufman, A. S. (1990). *Assessing adolescent and adult intelligence*. Boston: Allyn and Bacon.

Kaufman, A. S. (1994). *Intelligent testing with the WISC-III*. New York: Wiley..

Kaufman, A. S. (2001). Do low levels of lead produce IQ loss in chidren? A careful examination of the literature. *Archives of Clinical Neuropsychology*, *16*, 303–341.

Kaufman, A. S., & Kaufman, N. L. (1983). *Manual for the Kaufman Assessment Battery for Children (K-ABC)*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (1990). *Manual for the Kaufman Brief Test of Intelligence (K-BIT)*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (1993). *Manual for the Kaufman Adolescent and Adult Intelligence Test (KAIT)*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (Eds.) (in press). *Specific learning disabilities: psychological assessment and evaluation*. (Included in the Series Cambridge Monographs in Child and Adolescent Psychiatry.) Cambridge, England: Cambridge Univ. Press.

Kaufman, A. S., & Lichtenberger, E. O. (1999). *Essentials of WAIS-III assessment*. New York: Wiley.

Kaufman, A. S., & Lichtenberger, E. O. (2000). *Essentials of WISC-III and WPPSI-R assessment*. New York: Wiley.

Kehoe, R. A., Cholak, J., & Story, R. V. (1940). A spectrochemical study of the normal ranges of concentration of certain trace metals in biological materials. *Journal of Nutrition*, *19*, 579–592.

Kehoe, R. A., Thamann, F., & Cholak, J. (1935). Normal absorption and excretion of lead. *Journal of American Medical Association*, *104*, 90–92.

Kerlinger, F. N. (1979). *Behavioral research: a conceptual approach*. New York: Holt, Rinehart and Winston.

Lansdown, R., Yule, W., Urbanowicz, M. A., & Hunter, J. (1986). The relationship between blood-lead concentrations, intelligence, attainment and behaviour in a school population: the second London study. *International Archives Occupational Environmental Health*, *57*, 225–235.

Lichtenberger, E. O., Broadbooks, D. A., & Kaufman, A. S. (2000). *Essentials of cognitive assessment with the KAIT and other Kaufman measures*. New York: Wiley.

Marais, M. L., & Wecker, W. E. (1998). Correcting for omitted-variables and measurement-error bias in regression with an application to the effect of lead on IQ. *Journal of the American Statistical Association*, *93*, 494–517.

McCarthy, D. (1972). *Manual for the McCarthy Scales of Children's Abilities*. San Antonio, TX: The Psychological Corporation.

McLean, J. E. (1995). *Improving education through action research*. Thousand Oaks, CA: Corwin Press.

McMichael, A. J., Baghurst, P. A., Vimpani, G. V., Wigg, N. R., Robertson, E. F., & Tong, S. (1994). Tooth lead levels and IQ in school-age children: the Port Pirie cohort study. *American Journal of Epidemiology*, *140*, 489–499.

Mendelsohn, A. L., Dreyer, B. P., Fierman, A. H., Rosen, C. M., Legano, L. A., Kruger, H. A., Lim, S. W., Barasch, S., Au, L., & Courtlandt, C. D. (1999). Low-level lead exposure and cognitive development in early childhood. *Developmental and Behavioral Pediatrics*, *20*, 425–431.

Nation, J. R., & Gleaves, D. H. (2001). Low level lead exposure and intelligence in children. *Archives of Clinical Neuropsychology*, *16*, 375–388.

Needleman, H. L., & Bellinger, D. (2001). Studies of lead exposure and the developing central nervous system: a reply to Kaufman. *Archives of Clinical Neuropsychology*, *16*, 359–374.

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

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

Needleman, H. L., Riess, J. A., Tobin, M. J., Biesecker, G. E., & Greenhouse, J. B. (1996). Bone lead levels and delinquent behavior. *Journal American Medical Association*, *275*, 363–369.

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. *New England Journal of Medicine*, *322*, 83–88.

Nix, T. W., & Barnette, J. J. (1998). The data analysis dilemma: ban or abandon. A review of null hypothesis significance testing. *Research in the Schools*, *5*, 3–14.

Ostro, B. D., Mann, J. K., Collins, J. F., Vance, W. A., & Alexeef, G. V. (1996). *Health effects of airborne inorganic lead*. Air Toxicology and Epidemiology Section, Office of Environmental Health Hazard Assessment, California Environmental Protection Agency, Sacramento (Scientific Review Panel Draft for Public Comment).

Phelps, L. (1999). Low-level lead exposure: implications for research and practice. *School Psychology Review*, *28*, 477–492.

Pirkle, J. L., Kaufmann, R. B., Brody, D. J., Hickman, T., Gunter, E. W., & Paschal, D. C. (1998). Exposure of the U.S. population to lead, 1991–1994. *Environmental Health Perspectives*, *106*, 745–750.

Pocock, S. J., Ashby, D., & Smith, M. (1987). Lead exposure and children's intellectual perfomance. *International Journal of Epidemiology*, *16*, 57–67.

Pocock, S. J., Smith, M., & Baghurst, P. (1994). Environmental lead and children's intelligence: a systematic review of the epidemiological evidence. *British Medical Journal*, *309*, 1189–1197.

The Psychological Corporation. (1999). *Wechsler Abbreviated Scale of Intelligence (WASI) manual*. San Antonio, TX: The Psychological Corporation.

Rabinowitz, M. B., Leviton, A., & Bellinger, D. (1993). Relationships between serial blood levels and exfoliated teeth dentin levels: models of tooth lead kinetics. *Calcification Tissue International*, *53*, 338–341.

Reynolds, C. R. (1999). The inference of causality between smoking and low birth weight: good science or good politics. *Journal of Forensic Neuropsychology*, *1*, 55–87.

Reynolds, C. R. (in press). Inferring causality from relational data and designs: historical and contemporary lessons for research and clinical practice. *The Clinical Neuropsychologist*.

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

Rice, D. C. (1993). Lead induced changes in learning: evidence for behavioral mechanisms from experimental animal studies. *Neurotoxicology*, *14*, 167–178.

Rice, D. C. (1996). Behavioral effects of lead: commonalities between experimental and epidemiologic data. *Environmental Health Perspectives*, *104*, 337–351.

Rosenthal, R. (1994). Science and ethics in conducting, analyzing, and reporting psychological research. *Psychological Science*, *5*, 127–134.

Ruff, H. A. (1999). Population-based data and the development of individual children: the case of low to moderate lead levels and intelligence. *Developmental and Behavioral Pediatrics*, *20*, 42–49.

Ruff, H. A., Markowitz, M. E., Bijur, P. E., & Rosen, J. F. (1996). Relationships among blood lead levels, iron deficiency, and cognitive development in two-year-old children. *Environmental Health Perspectives*, *104*, 180–185.

Sachs, H. K. (1996). Bone lead levels and delinquent behavior — to the editor. *Journal American Medical Association*, *275*, 1725–1726.

Sayre, J. W. (1996). Bone lead levels and delinquent behavior — to the editor. *Journal American Medical Association*, *275*, 1726.

Schroeder, S. R., Hawk, B., Otto, D. A., Mushak, P., & Hicks, R. E. (1985). Separating the effects of lead and social factors on IQ. *Environmental Research*, *38*, 144–154.

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.

Schwartz, J. (1994a). Low level lead exposure and children's IQ: a meta-analysis and search for a threshold. *Environmental Research*, *65*, 42–55.

Schwartz, J. (1994b). Societal benefits of reducing lead exposure. *Environmental Research*, *66*, 105–124.

Seashore, H. G., Wesman, A. G., & Doppelt, J. E. (1950). The standardization of the Wechsler Intelligence Scale for Children. *Journal of Consulting Psychology*, *14*, 99–110.

Silva, P. A., Hughes, P., Williams, S., & Faed, 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, M., Delves, T., Lansdown, R., Clayton, B., & Graham, P. (1983). The effects of lead exposure on urban children: the Institute of Child Health/Southampton study. *Developmental Medicine and Child Neurology*, *25* (Suppl. 47), 1–54.

Sternberg, R. J. (1984). The Kaufman assessment battery for children: an information-processing analysis and critique. *Journal of Special Education*, *18*, 269–279.

Sternberg, R. J. (1985). *Beyond IQ: a triarchic theory of human intelligence*. New York, NY: Cambridge Univ. Press.

Survey of lead in the atmosphere of three urban communities. (1965, January). US Public Health Service Publication, No. 999-AP-12, Environmental Health Series, Air Pollution.

Tepper, L. B., & Levin, L. S. (1972). A survey of air and population lead levels in selected American communities. Department of Environmental Health, Kettering Laboratory, University of Cincinnati Final Report.

Tong, S., Baghurst, P. A., McMichael, A. J., Sawyer, M., & Mudge, J. (1996). Lifetime exposure to environmental lead and children's intelligence at 11–13 years: the Port Pirie cohort study. *British Medical Journal*, *312*, 1569–1575.

Tong, S., Baghurst, P. A., Sawyer, M. G., Burns, J., & McMichael, A. J. (1998). Declining blood lead levels and changes in cognitive function during childhood. *JAMA, the Journal of the American Medical Association*, *280*, 1915–1919.

Wang, L., Xu, S., Thang, G., & Wang, W. (1989). Study of lead absorption and its effect on children's development. *Biomedical and Environmental Sciences*, *2*, 325–330.

Wasserman, G. A., & Factor-Litvak, P. (2001). Methodology, inference and causation: environmental lead exposure and childhood intelligence. *Archives of Clinical Neuropsychology*, *16*, 343–351.

Wasserman, G. A., Graziano, J. H., Factor-Litvak, P., Popovac, D., Morina, N., Musabegovic, A., Vrenezi, N., Capuni-Paracka, S., Lekic, V., Preteni-Redjepi, E., Hadzialjevic, S., Slavkovich, V., Kline, J., Shrout, P., & Stein, Z. (1994). Consequences of lead exposure and iron supplementation on childhood development at age 4 years. *Neurotoxicology and Teratology*, *16*, 233–240.

Wasserman, G. A., Liu, X., Lolacono, N. J., Factor-Litvak, P., Kline, J., Popovac, D., Morina, N., Musabegovic, A., Vrenezi, N., Capuni-Paracka, S., Lekic, V., Preteni-Redjepi, E., Hadzialjevic, S., Slavkovich, V., & Graziano, J. H. (1997). Lead exposure and intelligence in 7-year-old children. *Environmental Health Perspectives*, *105*, 956–962.

Wasserman, G. A., Staghezza-Jaramillo, B., Shrout, P., Popovac, D., & Graziano, J. (1998). The effect of lead exposure on behavior problems in preschool children. *American Journal of Public Health*, *88*, 481–486.

Webster, A., Bamford, J. M., Thyer, N. J., & Ayles, R. (1989). The psychological, educational and auditory sequelae of early, persistent secretory otitis media. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *30*, 529–546.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children-Revised (WISC-R)*. San Antonio, TX: The Psychological Corporation.

Winneke, G. (1979). Neuropsychological studies in children with elevated tooth lead levels. Paper presented at the Conservation Society Symposium on Toxic Effects of Environmental Lead.

Winneke, G., Beginn, U., Ewert, T., Havestadt, C., Kraemer, U., Krause, C., Thron, H. L., & Wagner, H. M. (1985). Comparing the effects of perinatal and later childhood lead exposure on neuropsychological outcome. *Environmental Research*, *38*, 155–167.

Winneke, G., Brockhaus, A., Ewers, U., Kramer, U., & Neuf, M. (1990). Results from the European multicenter study on lead neurotoxicity in children: implications for risk assessment. *Neurotoxicology and Teratology*, *12*, 553–559.

Winneke, G., Kramer, U., Brockhaus, A., Ewers, U., Kujanek, G., Lechner, H., & Janke, W. (1983). Neuropsychological studies in children with elevated tooth-lead concentrations: II. Extended study. *International Archives Occupational Environmental Health*, *51*, 231–252.

Yule, W., Lansdown, R., Millar, I. B., & Urbanowicz, M. A. (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.