Pergamon

Archives of Clinical Neuropsychology
16 (2001) 375–388

Archives of CLINICAL NEUROPSYCHOLOGY

Low-level lead exposure and intelligence in children

Jack R. Nation*, David H. Gleaves

Psychology Department, Texas A&M University, College Station, TX 77843-4235 USA

Accepted 28 August 2000

Abstract

Numerous prospective and cross-sectional studies of the relation between low-level lead exposure and cognitive functioning in children have suggested that intellectual and academic performance declines as lead burdens increase. Kaufman [Arch. Clin. Neuropsychol. (2001)] raises questions regarding interpretive issues along these lines, and therein challenges the wisdom of using the available lead/IQ data complex as an essential element of the decision-making process that leads to policy statements. In this article, we address some of the concerns expressed by Kaufman, and conclude that each of his five points are logically or statistically flawed, as is his overall strategy of critiquing individual studies after methodologically sound meta-analyses have been performed. Kaufman is perhaps correct that the findings from correlational research on low lead levels and IQ loss should be interpreted with caution, but the caution extends equally if not more greatly in the direction of previous research having underestimated the relationship between the two variables in question.
© 2001 National Academy of Neuropsychology. Published by Elsevier Science Ltd.

Keywords: Children; Cognition; Lead; Intelligence

Some circumstantial evidence is very strong, as when you find a trout in the milk.
— Henry David Thoreau (November 11, 1850)

This entry into the journal of the American naturalist was aimed at the common practice among 19th century milk suppliers who, in an effort to deceive customers and thereby increase their profits, would deliver containers with water added to the milk. The point being, of course, that on occasion, related events take on greater significance when practical considerations limit alternative explanations of the phenomenon. This would seem to be the position taken by many members of the scientific community concerned with characterizing

* Corresponding author.

0887-6177/00/$ – see front matter © 2000 National Academy of Neuropsychology.
PII: S0887-6177(00)00088-3

Downloaded from https://academic.oup.com/acn/article/16/4/375/2046 by guest on 24 April 2024

the relation between chronic exposure to low levels of lead in children and declining scores on conventional intelligence tests. The argument is that, on the whole, the data on the topic are so compelling that the only sensible conclusion is that "safe" blood lead levels must be adjusted downward. In addition, this is precisely what happened in 1991 when the Centers for Disease Control (CDC) established the current blood lead guideline of 10 μg/dl.

Sensitivity to the issue is understandable, given the enduring nature of the problem. Despite the fact that child blood lead levels were reduced significantly between 1976 and 1991 (Pirkle et al., 1994), lead contamination persists from exposure to soil (Mielke, 1999), painted surfaces (Shannon, 1996), and other environmental vectors. Indeed, in 1994 Brody et al. estimated that 1.7 million or almost 9% of children in the U.S. under the age of 6 years had blood lead levels that were equal to or exceeded 10 μg/dl (Brody et al., 1994).

As relates to matters of intellectual function, Kaufman (2000) provides a welcome opportunity to debate the issue of lead/intelligence interactions within a scholarly forum. Few systematic challenges to the reference literature in this area have been attempted, and considering the enormous financial, health, and societal impact of policy decisions regarding acceptable blood lead levels, careful and cautious review of relevant findings is in order. It is within this context, that the following comments are offered and the strength of the "circumstantial evidence" assessed.

In his examination of the literature on developmental lead exposure and loss of intelligence, Kaufman adopts the approach of evaluating 26 of the "better" studies in the area that were included in one or more of three meta-analyses (Needleman & Gatsonis, 1990; Pocock, Smith, & Baghurst, 1994; Schwartz, 1994). Each investigation is examined vis-à-vis five recognized shortcomings. We have three general concerns with Kaufman's argument, as well as more explicit concerns with each of his five points. Our first general concern is that his basic strategy of reexamining the original 26 studies in terms of "positive or negative outcomes" is misguided. The purpose of meta-analysis is to go beyond examining significant or nonsignificant results at the individual study level, and calculate effect sizes across numerous studies. In his recent article "Science and ethics in conducting, analyzing, and reporting psychological research," Rosenthal (1994) describes meta-analysis as "an ethical imperative" and describes in detail why this should be so:

> Meta-analytic procedures use more of the information in the data, thereby yielding (a) more accurate estimates of the overall magnitude of the effect or relationship being investigated, (b) more accurate estimates of the overall level of significance of the entire research domain, and (c) more useful information about the variables moderating the magnitude of the effect or relationship being investigated. (p. 130).

Along similar lines, Schmidt (1996) offers a numeric example where a series of studies are completed on a treatment with, *in reality*, a medium effect size $d = 0.5$. Because of the common problem of low power (which would be much worse if researchers had controlled for multiple comparisons as Kaufman suggests — more on this issue below), in this hypothetical example, this real effect would be found only in 37% of the studies. However, meta-analysis would correctly determine a medium effect size. This hypothetical example sounds rather like Kaufman's discussion of the literature on blood level and IQ (e.g., p. 320). Furthermore, his use of the term "negative results" is inappropriate here. This term would

Downloaded from https://academic.oup.com/acn/article/16/4/375/2046 by guest on 24 April 2024

imply IQ gains. All that can be said here is "not statistically significant" but, as noted above, the meta-analyses take the science beyond this practice and synthesize research to better estimate effect sizes. Thus, Kaufman's overall strategy arguably is moving science backward rather that forward.

A second general concern is that, throughout the paper, Kaufman makes statements as if they are factual when they could be more accurately described as conjectural. For example on page 322, he writes:

> . . .the results [of a multiple regression analysis] for one sample are not readily generalizable to samples that differ in meaningful ways from the original sample. Therefore, the results of the Bellinger et al. (1992) study only generalize to similar samples of high SES white children, not to minorities or to whites from lower social classes.

A more correct statement is that the results may or may not be generalizable to other samples. One should not assume that the results are generalizable, but it is also erroneous to assume (as does Kaufman) that they are not generalizable. Another example of Kaufman stating something as if it were a fact is on page 323 where he writes, "only the adjusted values [from an analysis of covariance] are interpretable." As we will discuss in greater detail below, this may or may not be the case. In some circumstances (e.g., with high multicollinearity), the adjusted values from such analyses are the *least* interpretable and meaningful.

A final example of Kaufman stating with certainty what in actuality is opinion, perspective, or possibility is that:

> The *correct* [italics added] procedure is to apply a statistical correction to control for the error that occurs when many analyses are done at once. Several of the lead researchers have used this shotgun approach, but none have controlled for the known errors that accompany this approach. (p. 330)

This statement is erroneous for two reasons. First, as we will discuss later, there is not widespread agreement that the "correct" procedure when performing multiple comparisons is to adjust the alpha level. Whether or not one does this is really based on the nature of the research question being asked, and whether one wants to minimize Type I or Type II errors (see Stevens, 1996). Using a correction may minimize Type I errors, but may drastically reduce power and increase the rate of making Type II errors. It is also erroneous to refer to the "known errors" that accompany conducting multiple tests. Such a statement is based on a misunderstanding of the probability of making Type I errors (see Pollard & Richardson, 1987). Type I errors may or may not occur when multiple tests are performed; it all depends on whether or not the null is true. If the null is false (in this case that lead does cause IQ losses in children) then it is impossible to make a Type I error (i.e., to conclude that lead is harmful when it is not) no matter how many analyses one does. One could only make a Type II error, and the lower one's alpha level, the greater the probability of making a Type II error.

Our final general concern is how Kaufman treats and refers to the issue of causation. On page 313, he makes the statement that "resolving the causality issue is beyond the scope of this paper." We find this statement particularly curious when one examines the overall context of the paper. Kaufman's general thesis is about causation, and throughout the paper,

Downloaded from https://academic.oup.com/acn/article/16/4/375/2046 by guest on 24 April 2024

he makes numerous statements strongly implying causality. As an example, on page 311, he writes "whatever environmental variables are responsible for the steady gain are likely to lead to loss of intellectual functioning when they are diminished or withdrawn." Thus, his argument is clearly that environmental factors other than lead are "causing" IQ losses.

Although we believe that these overall, general concerns are sufficient to call into question Kaufman's conclusions, we have additional concerns with each of the five shortcomings that Kaufman points to in the original studies. Below, we address each of the five points.

Downloaded from https://academic.oup.com/acn/article/16/4/375/2046 by guest on 24 April 2024

1. Critique of the best-designed studies relating low BLLs to IQ loss

1.1. Shortcoming #1 — uncontrolled variables cloud conclusions drawn from even the best studies

The initial methodological concern expressed in the Kaufman paper relates to sources of uncontrolled variation. Included in the list of uncontrolled variables, is parenting skill/style. Because parental communication and participation is correlated with IQ (e.g., Fulton et al., 1987), and because IQ tends to be higher and lead levels lower when both parents are in the home (Baghurst et al., 1992; McMichael et al., 1994), the suggestion is that parental attention, involvement, etc., may represent concomitant variables that contribute substantially to observed IQ losses in lead-exposed children. Such an argument should be considered, of course, but it is unconvincing for a number of reasons, not the least of which is the fact that, as Kaufman notes on page 306, it is impossible to document the veracity of reports from parent interviews of the sort that were used in investigations such as Fulton et al. (1987). Even if the results of such parent studies could be believed, lead is not excused as a primary determinant of IQ loss, inasmuch as the toxicity profile may include disturbances in caretaking that may play a modulatory role in intellectual development. This line of reasoning obviously offers no support for the notion that lead contamination causes a drop in IQ, but neither can the pattern of results be used to minimize concerns of lead/IQ interactions.

A major problem with Kaufman's argument is how he continually writes about the need for and use of statistical methods to control for these potentially confounding variables. He writes that "technically, a confound in a lead–IQ study correlates significantly with both the dependent variable of children's IQ and with the independent variable of BLL" (p. 310). He then describes the data analytic strategy:

> ... investigators have uniformly used appropriate, sophisticated regression analysis procedures. They have identified various sets of potentially confounding or contaminating variables and have entered these variables into the regression equation in an attempt to control as much of the unwanted variability and possible. Then, lead level is added to the equation to see if it contributes a significant amount of variance to the prediction of IQ, over and above the prediction that is obtainable from the confounds alone. (p. 312)

Kaufman's definition of a confound is accurate. However, the scenario he describes is just the type of situation when statistical control is, at best, problematic, or more likely totally

inappropriate. It is the essential problem of multicollinearity, which makes interpretation of a multiple regression analysis difficult. For example, as noted by Wampold and Freund (1987):

> The high correlation among some or all of the independent variables is called multicollinearity. ... multicollinearity is undesirable for several reasons. First and most importantly, if the independent variables are highly correlated, none of them will demonstrate a substantial unique contribution to the prediction of the dependent variable... Nevertheless, it would be incorrect to conclude that [the predictors] are unimportant... A second problem of multicollinearity is that the estimates of population partial regression coefficients are highly unstable, resulting in decreased probabilities of obtaining statistically findings. (p. 378)

When used properly, statistical control techniques such as analysis of covariance or hierarchical regression can be powerful tools for the behavioral scientist. They allow the researcher to increase the precision of research by partitioning out error variance (Hinkle, Wiersma, & Jurs, 1988; Pedhazur, 1982). However, their use is most appropriate in randomized designs (e.g., Kirk, 1982). Lord (1969) noted that when subjects are not randomly assigned to treatment groups, there is no logical or statistical procedure that can properly allow for uncontrolled, preexisting differences between groups. More importantly, for a covariate (control variable) to be effective in controlling for an extraneous variable, the covariate must be linearly related to the dependent variable and *unaffected by the treatment or independent variables* (Hinkle et al., 1988; Porter & Raudenbush, 1987), or precisely opposite the situation described by Kaufman. If a covariate is significantly correlated with both the dependent variable(s) and the independent variables, then the results of the analysis may be highly misleading and, at times, meaningless (Pedhazur, 1982). The analysis will simply partial out not just variance attributable to the covariate, but also variance shared by the covariate and the independent variable.

Thus, when you have this situation (high correlations between lead levels and the various confounds), while you are ''controlling'' for [partialling out] the effects of the confounding variables, you are also partialling out [removing] the effects for lead. The more highly correlated are lead and the confounds, the more of the variance directly attributable to lead is removed. Overall, you get an underestimate of the actual effect of lead and the results are less likely to be statistically significant. Thus, Kaufman's statement that ''A more correct statement is that any significant increase in prediction is due not only to lead level, but also to all other potentially important variables, known or unknown, that were not controlled in the study'' (p.312) is not justifiable. A better statement is that 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* because of the overlap in predictor variables and statistical techniques having partialled out the effects of lead.

In conclusion, it is likely that there are uncontrolled sources of variation in human lead studies of IQ loss. However, it is just as likely that the use of statistical control has controlled for not only additional sources of variance but also for the effects of lead, thus leading to underestimates of the actual effect sizes. Furthermore, when a converging body of data points to a common underlying variable as a responsible agent in cognitive and intellectual dysfunction, caution must be exercised in minimizing such a causal variable. Diverting

Downloaded from https://academic.oup.com/acn/article/16/4/375/2046 by guest on 24 April 2024

attention to other possible sources of negative influence in no way alters the need to attend to a primary risk factor, especially when that risk factor may impact the other possible sources (another situation in which statistical control is inappropriate, cf. Briere & Elliot, 1993). A case in point involves lead and ethanol, a drug that when consumed excessively by pregnant women decreases intellectual functioning in offspring (Olson, Feldman, Streissguth, Sampson, & Bookstein, 1998). An expanding clinical (Shaper et al., 1982) and preclinical (Nation, Baker, Fantasia, Ruscher, & Clark, 1987; Nation, Baker, Taylor, & Clark, 1986) literature shows that increasing lead burdens increase the volitional intake of alcohol. In such instances what is accomplished by focusing exclusively on the effects of alcohol on IQ and ignoring metal/drug interactions? Lead is no less a risk factor when it acts indirectly to compromise the intellectual apparatus.

1.2. Shortcoming #2 — parental IQ is typically measured poorly or not at all

According to Kaufman, a second shortcoming of the 26 previously published studies is that, while attempting to control for "confounds," parental IQ is typically poorly measured or not measured at all. Kaufman's arguments here are contradictory and illogical. He first criticizes several studies for having used "short forms" of the Wechsler Adult Intelligence Scale but then wrote that "The substitution of a Wechsler short form for the full battery . . . is still an acceptable procedure for use in multiple regression analysis and will not adversely affect the outcomes of the study" (p. 313). If the latter is true, then why raise the issue as a concern? Kaufman similarly criticizes the use of the Peabody Picture Vocabulary Test — Revised before pointing out that one could argue that it correlates "highly enough with conventional IQ tests [e.g., $r = .70$ to .80 range] to be an adequate substitute for an IQ test" (p. 314). Kaufman tries to dismiss this argument by noting that the average correlation between IQ the PPVT is somewhat lower [medium $r = .51$] in disadvantaged samples. However, given that in some disadvantaged samples, the correlations were as high as $r = .87$, this is not a convincing argument and it is clear that the PPVT explains a large proportion of the variance in IQ. Thus, when used as a covariate, it partials out a substantial proportion of the variability due to parental IQ and, as noted above, also a substantial portion of the variability due to lead.

1.3. Shortcoming #3 — failure to control for multiple comparisons

Kaufman's third shortcoming was that numerous researchers had performed several analyses without statistically controlling for the number of tests being performed. There are three fundamental problems associated with this argument: (a) it is based on a misunderstanding of the probability of making Type I errors, (b) it assumes that Type I errors are more problematic than Type II, and (c) it ignores the fact that meta-analysis gets around this issue entirely by not calculating significance tests from single studies and, instead, calculating effect sizes across numerous studies. This third problem with the argument may be most critical but, because we discussed it in detail above, we will not discuss it here.

Regarding the first problem, Kaufman seems to have a common misconception about the probability of making Type I errors. It would be beyond the scope of this paper to discuss

Downloaded from https://academic.oup.com/acn/article/16/4/375/2046 by guest on 24 April 2024

this issue in complete detail (see Pollard & Richardson, 1987 for thorough discussion); however, one point is critical to address. Conducting more statistical tests does not automatically lead to more Type I errors, as Kaufman explicitly states. As noted above, one can only make a Type I error when the null hypothesis is true, which in this case means lead does result in IQ loss. The reader may be familiar with the concept of familywise Type I error rate, which can be calculated as $1-(1-\alpha)^k$, where k = the number of statistical tests performed. However, this is a conditional probability, an extrapolation of the conditional prior probability (Pollard & Richardson, 1987), and can be defined as the probability of making at least one Type I error, *when the null hypothesis is true*. Thus, if the null is true and one sets an α at .05, there is a 5% chance of making a Type I error in one analysis. If one performs 10 analyses, and the null is true in every case, then there is approximately a 40% chance of making *one* error out of the 10 statistical tests. However, this probability inflation only occurs when the null is true. If it is false, one cannot make a Type I error.

The second problem with Kaufman's argument is that one should not arbitrarily assume that Type I errors are more problematic than Type II errors (cf. Stevens, 1996), The decision of which to make more important (where to set the alpha level), should be based on the question being addressed. When the issue is health risks in young children, we would argue that minimizing Type II errors would be more important, and using a Bonferroni correction (as Kaufman suggests one *must* do) would radically increase the rate of Type II errors (i.e., concluding that lead is not harmful when actually it is).

1.4. Shortcoming #4 — comparisons of the IQs of the two extreme "lead-level" groups

Kaufman writes that "It is inappropriate to compare the IQs of the two extreme "lead-level" groups when several additional groups are included in the study"(p. 330), claiming that this "violates the rules for interpreting multiple regression analysis" (p. 330). It is our position, however, that Kaufman's assertions are unjustified. As expressed by Lord (1953) years ago, "the numbers do not know where they come from" (p. 751), and as long as the statistical assumptions of multiple regression are met, there is no reason why one cannot perform the analyses that Kaufman criticizes. What has been done is simply an analysis of variance comparing top and bottom groups. Thus, there is no statistical rationale for concluding that IQ losses attributed to lead are "bogus" when comparing extreme groups. Kaufman's contention that regression is only valid for whole samples is simply not true. In actuality, using a dichotomous variable rather than a continuous variable may lead to a lower r^2 value because one has decreased the variability in the predictor variable and it is impossible to predict more variability from less variability[1]. Thus, such a data analytic strategy would lead to an underestimation of the relationship between lead levels and IQ

[1] The one exception would be when there is a strong nonlinear relationship between a predictor and criterion variable. Analysis of the entire sample might obscure the relationship that would only be detected by comparing specific groups. Kaufman speculates that the relationship between lead and IQ loss is nonlinear, but he presents no empirical support for this argument. Furthermore, if he is correct, then multiple regression analyses that did not examine specific groups would have systematically underestimated the relationship between lead and IQ loss.

Downloaded from https://academic.oup.com/acn/article/16/4/375/2046 by guest on 24 April 2024

loss. Perhaps the researcher who only compares extreme groups is making a methodological error, but it most likely will be an underestimate of the actual relationship between blood lead levels and IQ loss.

1.5. Shortcoming #5 — lack of quality control in measuring children's IQ

Kaufman's fifth shortcoming was what he referred to as a "lack of quality control in measuring children's IQ" (p. 28). Here, he is apparently not referring to the quality of the IQ *tests* because he had earlier referred to the "rigorous attempts that most experimenters made to use state of the art IQ tests for *children* [original italics] in the studies" (p. 313). Rather, he seems to be referring to how the tests are administered. However, his argument that the data are questionable because the tests might have been administered by advanced graduate students is not compelling when one considers that extensive training in the administration of standardized intellectual tests is an integral part of graduate training in psychology. Furthermore, even if there were sources of error associated with the administration and scoring of the IQ tests (which certainly is possible), there is no reason to expect that it would be anything other than random error. That is, there is no evidence that there would be more error associated with higher or lower lead level or higher or lower IQs. Thus, as with several arguments discussed above, any unreliability in the dependent variable IQ would simply diminish the effect size for the relationship between lead and IQ loss. Kaufman has once again pointed to a possible way in which the relationship between lead and IQ loss has been *underestimated* by the previous research.

2. Kaufman's overview of the five points

Finally, in an overview that completes Kaufman's coverage of the five principal shortcomings that limit the value of findings in the area of developmental lead toxicity and IQ loss, he returns to the all-important issue of controlling for differences in socioeconomic status (SES). It is within this context that the report by Bellinger, Stiles, and Needleman (1992), is acknowledged to have "... controlled more important SES and family-related variables than any other lead–IQ investigation" (p. 321). Following approbation for exercising excellent control of potentially confounding variables ostensibly overlooked by most other investigators, Kaufman proceeds to challenge the ultimate worth of the Bellinger et al. findings because of shifting norms, statistical violations, and potentially compromising sample characteristics.

It is the Bellinger et al. (1992) sample that invites the greatest criticism in Kaufman's paper. Kaufman accurately notes that virtually the entire sample was comprised of middle-class and upper-middle-class white children who fell into the "high average" range on the WISC-R. Although it is not explicitly stated in the Kaufman paper, the Bellinger et al. findings are tacitly dismissed as unimportant. The argument is that because "... the results of the Bellinger et al. study only generalize to similar samples of high SES white children, not to minorities or to whites from lower social classes" (p. 322), the finding that increased lead levels are associated with an IQ decrement in this study is not a major worry. As noted above,

Downloaded from https://academic.oup.com/acn/article/16/4/375/2046 by guest on 24 April 2024

the results from one study may or may not be generalizable to another sample[2]. However, even if Kaufman is correct about the generalizability issue (and he may or may not be correct), this position, of course, obscures the fact that "upper-middle class white children in the U.S." constitutes a sizable population that merits the attention of public health officials. To be sure, a parent of a child falling into this category would likely want to believe that prophylactic measures would be considered by environmental specialists, and that attempts at improving their child's quality of life were being made. Moreover, the same can be said about Kaufman's expressed concerns over cross-cultural comparisons and generalizability. Restricting statements of generalization to subpopulations defined by millions of children whom may be at risk due to chronic low-level lead exposure can hardly be fashioned as a "... limitation of empirical research."

3. Questionable interpretations of the IQ loss attributed to low-level lead

In addition to suggesting that the developmental lead/IQ literature is fraught with numerous methodological flaws, Kaufman questions the appropriateness of some of the conclusions that have been drawn in this area. That is, even if lead does cause IQ loss, are some of the assertions that have been made justified?

3.1. There is no documented linear relationship between lead level and IQ

Kaufman claims that there is lack of support for a linear relationship between lead level and IQ. Assuming that he is correct, once again he has pointed out how previous research has *underestimated* the relationship between the two variables. That is, multiple regression is based upon the assumption that the relationship between all predictors (which would include all "confounds") and criterion is linear. If the relationship is nonlinear (e.g. quadratic, cubic or higher) a multiple regression equation would *underestimate* the overall relationship between BLL and IQ. Furthermore, Kaufman's statement that "to determine whether the IQ–lead level relationship is linear, the authors of the various studies would have needed to present adjusted IQs for children at each portion of the lead level-continuum" (p. 323) is incorrect. They would only have needed to perform a trend analysis or test for a quadratic or higher order regression effect.

[2] As noted above, Kaufman overstates the problem of generalizability in multiple regression. Often, what does not generalize across samples is the exact variable weights of each predictor, and the more highly correlated the predictors are, the more unstable the weights will be. However, this is different from concluding that what was found to be an important predictor in one sample *cannot* be assumed to be a predictor in another, and research often demonstrates that the findings are indeed generalizable. For example, if one performed a study predicting percent body fat from caloric intake, activity level, and parental percent body fat and all three were found to be significant predictors, would the fact that one used only a sample of high SES adults preclude generalizability to other societal groups? The weights of the individual variables might vary, but all three would still be found to be meaningful predictors.

Downloaded from https://academic.oup.com/acn/article/16/4/375/2046 by guest on 24 April 2024

Kaufman's basic argument here seems to be that there is a threshold effect. However, his only empirical support for this notion was the "eyeball test." That is, he looked at the graphs and concluded that most seemed nonlinear. Apart from the fact that he is not an unbiased rater, his statements regarding his conclusions are ambiguous (e.g., "there was *decidedly* no linearity for children with lead levels below 15 μg/dl"[p. 323] — decided by whom? Or, "The values seem virtually identical" (p. 324) — seems to whom?). Overall, Kaufman's arguments here are not convincing, and we would challenge the reader to perform the same eyeball test of the graphs in, for example, Bellinger et al. (1992).

3.2. Interpreting fractions of an IQ point has no scientific meaning

In this section of Kaufman's discussion, he argues that it is inappropriate to interpret fractions of an IQ point. There are several problems with his argument. First, but less important, his examples are exaggerations. Comparing 1/2 an IQ point to 1/100 of a pound distorts the issue; a better comparison would be to 1/2 a pound. More importantly, the fractions of an IQ point (e.g. .257, p. 304) are per 1 μg/dl; thus with larger increases in lead levels, the issue is not fractions of an IQ point but rather somewhere around 6 IQ points. Most importantly, the reported changes are averages across multiple subjects. Kaufman's example of how meaningful a fraction of a pound would be to one person is misleading because it is based on the individual. Perhaps half of a pound would have little meaning to the individual, but an average of 1/2 pound or, more comparably, six or more pounds *averaged across thousands of people* might have enormous meaning in terms of preventing heart disease, type II diabetes, etc. Kaufman's long discussion of sources of error is interesting, but irrelevant because it is random error that he is describing. It would only be meaningful it he could demonstrate that there was more error associated with high or low IQs or high or low blood lead levels, which he does not do. The bottom line here is that we may be dealing with small effect sizes but as Kaufman himself noted (p. 305), small effect sizes can have practical significance. As noted by Rosenthal (1994) "effect sizes can be of major importance even when they are so small as to have $r^2 = 0.00$." It depends decidedly on the variables being studied; with IQ, average losses of .257 per one μg/dl may indeed have enormous societal consequences.

3.3. Societal consequences of lead level

As indicated by Kaufman, Herbert Needleman and his colleagues have suggested that there will be social and economic backlash consequent to a lead induced displacement to the left in the normal curve that has long characterized the distribution of IQ scores (Bellinger, Stiles, & Needleman, 1992). The idea is that pooling at the lower end of the distribution, coupled with a drain at the upper end, will increase educational costs while depleting the talent base necessary to meet the financial demands of educating greater numbers of impaired children. Kaufman counters with the rationale that Needleman's thinking is flawed because IQ is a relative, not absolute, concept. The abiding theme is that norms change as scores shift within the overall distribution, thus at the extremes the percentages under the curve will remain constant.

Downloaded from https://academic.oup.com/acn/article/16/4/375/2046 by guest on 24 April 2024

To illustrate this point, Kaufman used the example of "an artificial neuronal transmitter that would increase everyone's IQ by 20 points" (p. 326). As with any mathematically operation where a constant is added to every score in the distribution, there would be a shift in the mean but not the variance, consequently the new distribution of IQ scores would have a mean value of 120 in contrast to the previous mean of 100. Yet, because the properties of the normal curve remain essentially intact, the percentages under the curve would remain fundamentally the same. That is, after the restandardization process there still would be 2% at the extremes, etc. Relatively speaking, nothing would change at least as comparisons are made one IQ score to another.

Although Kaufman's argument that IQ is a relative concept is sound from a purely measurement perspective, the illustration is directionally opposite what would be expected regarding developmental lead/IQ interactions. What if there was an event that preferentially altered neurotransmitter function in such a manner that IQ was decreased in the population by a set amount? What would be the consequences in an absolute as well as a relative sense?

In point of fact, there is a burgeoning neurochemical literature that shows lead exposure does alter a neurotransmitter system integral to cognitive and intellectual functioning. In vivo as well as in vitro studies have reliably demonstrated that inorganic lead targets the *N*-methyl-D-aspartate (NMDA) receptor complex, which is a glutamate receptor subtype (Guilarte, 1997; Lasley, Green, & Gilbert, 1999). NMDA receptor activation is known to be essential for biochemical/neurochemical changes that form the basis for learning and memory operations (e.g., Madison, Malenka, & Nicoll, 1991). For instance, hippocampal long-term potentiation (LTP) represents a form of neural plasticity that is thought to be a cellular substrate for learning and conditioning in the mammalian brain (Akhondzadeh, 1999). Inasmuch as lead has been shown to be a potent inhibitor of NMDA-activated LTP (cf. Lasley et al., 1999), it is not surprising that there is a rich preclinical literature that shows a causal connection between developmental lead exposure and impairments in learning and performance (Cory-Slechta, 1997).

Necessarily, understanding the mechanism of action of lead on the developing nervous system has required the use of animal models. But as Rice (1996) points out in her careful review of the commonalities between the available experimental and epidemiological data, there is a remarkable congruence between these two literatures with respect to identifying those processes that determine behavioral deficits. As noted, on p. 313 Kaufman steps away from the experimental literature by stating that "... resolving the causality issue is beyond the scope of this paper." Although this is a reasonable statement, it does not lessen the relevance of emerging parallels between the animal and human data.

In any case, should lead-based disturbances in glutamatergic function, or some other underlying physiologic process related to cognitive ability and performance, actually decrease IQ by a set value, what does this mean in a practical sense? Kaufman is correct in his assertion that in relative terms nothing changes. There are neither fewer nor greater percentages of people who fall at the low end of the curve. The advantaged/disadvantaged ratios will remain the same, so as a population we are no worse off than before. But we are worse off than before. When the ability of an entire population to reason or to respond to changing environmental conditions is compromised, society assumes a definite risk. Within

Downloaded from https://academic.oup.com/acn/article/16/4/375/2046 by guest on 24 April 2024

this context, relative comparisons are meaningless. It hardly matters that a child with a low IQ is no farther behind others in the population than she/he was prior to the introduction of environmental lead. The significant issue is that societal growth is retarded in an absolute sense. The bottom line is that *measurement* of IQ may be in relative terms, but the construct is not. In his hypothetical example of the artificial neurotransmitter that raised IQ points by 20 points, it would be obvious that this had somehow helped the people, even if renormed tests showed their IQs to be the same. Similarly, if lead is lowering IQs, children (and adults) are being harmed. It would seem, then, that we can take little solace from knowing that IQ is a relative concept.

4. Conclusion

In conclusion, Kaufman's statement that "The preceding five points, taken together, represent a rationale for interpreting all of the lead–IQ research, both the positive and the negative findings regarding the impact of low BLLs on IQ, with caution"(p. 320) is perhaps true. However, the caution should be extended at least as strongly if not more strongly in the other direction, i.e., the obtained effect sizes may be underestimates rather than overestimates. Virtually every limitation of the research described by Kaufman points to ways in which the relationship between blood lead level and IQ may have been underestimated. In the final analysis, the converging clinical and preclinical data indicate that even low-level lead exposure can have a profound effect on public health. Human and animal findings consistently point to lead-induced learning impairment (Bellinger et al., 1992; Cory-Slechta, 1997; Rice, 1996), social delinquency and aggression (Delville, 1999; Needleman, Reiss, Tobin, Biesecker, & Greenhouse, 1996), and numerous other disturbances. 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.

References

Akhondzadeh, S. (1999). Hippocampal synaptic plasticity and cognition. *Journal of Clinical Pharmacy and Therapeutics*, *24*, 241–248.

Baghurst, P. A., McMichael, A. J., Wigg, N. R., Vimpani, G. V., Robertson, E. F., Roberts, R. J., & Tong, S. (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.

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.

Briere, J., & Elliot, D. M. (1993). Sexual abuse, family environment, and psychological symptoms: on the validity of statistical control. *Journal of Consulting and Clinical Psychology*, *61*, 284–288.

Brody, D. J., Pirkle, J. L., Kramer, R. A., Flegal, K. M., Matte, T. D., Gunter, E. D., & Paschal, D. C. (1994). Blood lead levels in the US population: phase 1 of the Third National Health and Nutrition Examination Survey (NHANES III, 1988–1991). *Journal of the American Medical Association*, *272*, 277–283.

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

Downloaded from https://academic.oup.com/acn/article/16/4/375/2046 by guest on 24 April 2024

Delville, Y. (1999). Exposure to lead during development alters aggressive behavior in golden hamsters. *Neurotoxicology and Teratology*, *21*, 445–449.

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*, *I*, 1221–1226 (Saturday 30 May).

Guilarte, T. R. (1997). Pb inhibits NMDA function at high and low affinity sites: developmental and regional brain expression. *Neurotoxicology*, *18*, 43–52.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1988). *Applied statistics for the behavioral sciences*. Boston: Houghton Mifflin.

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

Kirk, R. E. (1982). *Experimental design: procedures for the behavioral sciences*. Belmont, CA: Brooks/Cole.

Lasley, S. M., Green, M. C., & Gilbert, M. E. (1999). Influence of exposure period on in vivo hippocampal glutamate and GABA release in rats chronically exposed to lead. *Neurotoxicology*, *20*, 619–629.

Lord, F. (1953). On the statistical treatment of football numbers. *American Psychologist*, *8*, 750–751.

Lord, F. (1969). Statistical adjustments when comparing pre-existing groups. *Psychological Bulletin*, *70*, 162–179.

Madison, D. V., Malenka, R. C., & Nicoll, R. A. (1991). Mechanisms underlying long-term potentiation of synaptic transmission. *Annual Review of Neuroscience*, *14*, 379–397.

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.

Mielke, H. W. (1999). Lead in the inner cities. *American Scientist*, *87*, 62–73.

Nation, J. R., Baker, D. M., Fantasia, M. A., Ruscher, A. E., & Clark, D. E. (1987). Ethanol consumption and free operant avoidance performance following exposure to dietary lead. *Neurotoxicology*, *8*, 561–569.

Nation, J. R., Baker, D. M., Taylor, B., & Clark, D. E. (1986). Dietary lead increases ethanol consumption in the rat. *Behavioral Neuroscience*, *100*, 525–530.

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

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

Olson, H. C., Feldman, J. J., Streissguth, A. P., Sampson, P. D., & Bookstein, F. L. (1998). Neuropsychological deficits in adolescents with fetal alcohol syndrome: clinical findings. *Alcoholism: Clinical and Experimental Research*, *22*, 1998–2012.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research*. New York: Holt.

Pirkle, J. L., Brody, D. J., Gunter, E. W., Kramer, R. A., Paschal, D. C., Flegal, K. M., & Matte, T. D. (1994). The decline in blood levels in the United States: the national health and nutrition examination surveys (NHANES). *Journal of American Medical Association*, *272*, 284–291.

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.

Pollard, P., & Richardson, J. T. E. (1987). On the probability of making type I errors. *Psychological Bulletin*, *102*, 159–163.

Porter, A. C., & Raudenbush, S. W. (1987). Analysis of covariance: its model and use in psychological research. *Journal of Counseling Psychology*, *34*, 383–392.

Rice, D. C. (1996). Behavioral effects of lead: commonalities between experimental and epidemiological 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.

Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: implications for training of researchers. *Psychological Methods*, *1*, 115–129.

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

Shannon, M. (1996). Epidemiology of childhood lead poisoning. In: S. M. Puschel, J. G. Linekas, & A. Anderson (Eds.), *Lead poisoning in childhood* (pp. 37–57). Baltimore: Paullbrook.

Downloaded from https://academic.oup.com/acn/article/16/4/375/2046 by guest on 24 April 2024

Shaper, G., Pocock, S. J., Walker, M., Wale, C., Clayton, B., Delves, H. T., & Hinks, H. (1982). Effects of alcohol and smoking on blood lead in middle-aged British men. *British Medical Journal*, *284*, 299–302.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Thoreau, H. D. (1850/1999). In: R. Sattelmeyer, M. R. Patterson, & W. Rossi (Eds.), *Journal* (vol. 3, pp. 1848–1851). Princeton: Princeton University Press.

Wampold, B. E., & Freund, R. D. (1987). Use of multiple regression in counseling psychology research: a flexible data-analytic strategy. *Journal of Counseling Psychology*, *34*, 372–382.

Downloaded from https://academic.oup.com/acn/article/16/4/375/2046 by guest on 24 April 2024