

Introduction: Biomarkers in Neurodevelopmental Toxicology

by Herbert L. Needleman*

The search for markers of toxicant exposure and effect upon the development of organisms presents a set of challenges that differ in many ways from those encountered in the study of markers in reproduction or pregnancy. These latter two fields specify a relatively narrow set of organs or biological systems. The term "development," on the other hand, can apply to any organ system, or to any set of phenomena that changes in an ordered way over time. For this reason the papers presented in the session on development were chosen to narrow the focus to neurodevelopmental markers, as such markers may be altered by neurotoxic exposure.

In attempting to meet this somewhat daunting task, we have been able to select a group of investigators who work at the leading edges of their respective fields of developmental neuroanatomy, neurotoxicology, neuroendocrinology, neuropsychology, and infant development. In introducing this topic, I offer a few summary comments on the utility of some behavioral measures, particularly measures of attention, as markers; on the meaning of "adverse health effect"; on the importance of prior information about the sensitivity and specificity of any candidate marker; and on certain epistemic issues encountered in drawing causal inferences.

The notion that toxicants could affect behavior certainly is not new. Recent knowledge that behavioral aberrations can occur at exposures below those which produce organic changes, and that behavioral observation might provide early markers of effect has given rise to two new fields: behavioral toxicology and behavioral teratology.

Teachers as Behavioral Toxicologists

The application of behavioral markers in evaluating neurotoxins does not require behavioral toxicologists; it can be done, for example, by teachers. Figure 1 shows the responses of all the first and second grade teachers in Somerville and Chelsea, MA, when asked to complete

an 11-item forced choice questionnaire grading the classroom performance of over 2000 students whose dentine lead levels were known (1). As dentine lead level increases, the proportion of bad reports for each item increases in regular monotonic fashion. The same instrument, in the hands of British and Greek teachers applied to schoolchildren classified by blood lead levels, produced strikingly similar results (2,3).

Attention as a Marker

One of the behavioral functions that appears to be most responsive to neurotoxins is attention. A number of the contributors to this symposium address this competence. Figure 2 displays two experiments in which reaction time at two intervals of delay, a measure of vigilance, are plotted in subjects classified by lead burden (4). It can be seen that blood lead level and reaction time at longer intervals of delay are closely correlated. This nested series of curves compiled from schoolchildren in two separate countries describes a dose-response relationship.

Values in Defining Adverse Health Effects

Discriminating between biological markers and health markers requires specifying what is meant by an adverse health effect. Is any change of state in and of itself an adverse health effect? Consider for example, free erythrocyte protoporphyrin (FEP), a heme precursor pigment consistently elevated at levels of lead exposure that are not necessarily attended by brain edema, anemia, or kidney failure. Is an elevated FEP evidence of disease? Figure 3, taken from a paper by Hernberg (5), clarifies the question with regard to lead exposure, and can be applied to any neurotoxin. Internal lead exposure is plotted on the abscissa, and frequency of measured effect is plotted on the ordinate. A number of measured outcomes of varied health consequence are then plotted. At the far right is the most serious health consequence, death. D_0 represents the dose at which the first death

*Department of Psychiatry, Children's Hospital of Pittsburgh, Pittsburgh, PA 15213.

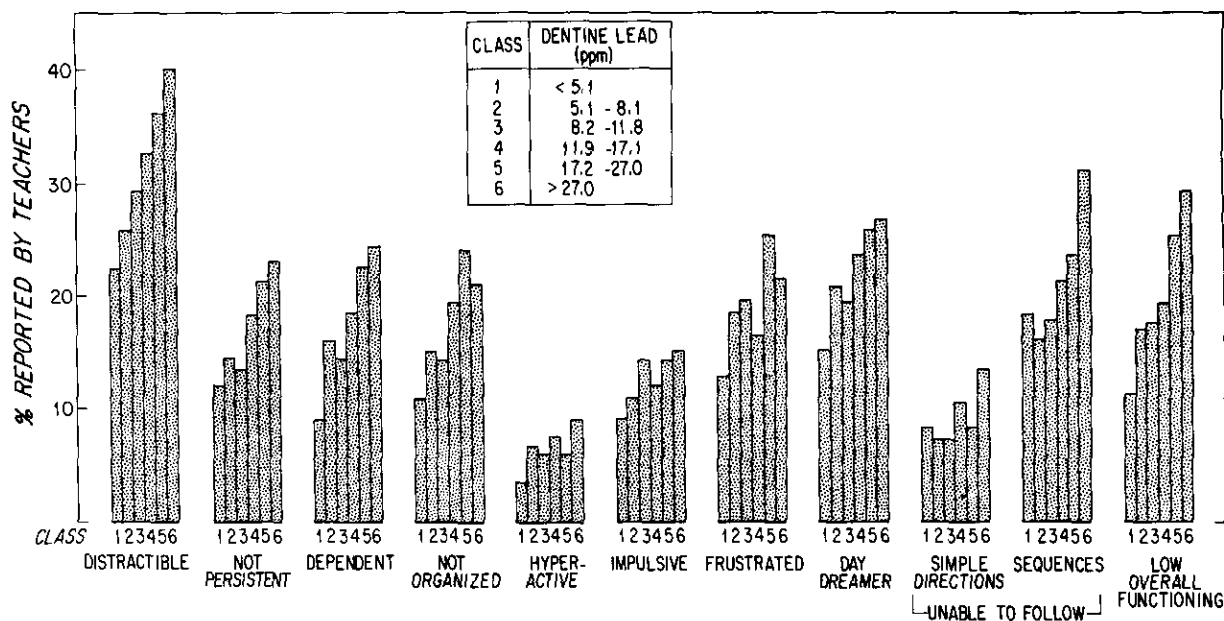


FIGURE 1. Teachers' ratings of classroom behavior in relation to dentine lead level. Teachers were blind to lead levels and had known students for at least 2 months, $n = 2146$. From Needleman et al. (1).

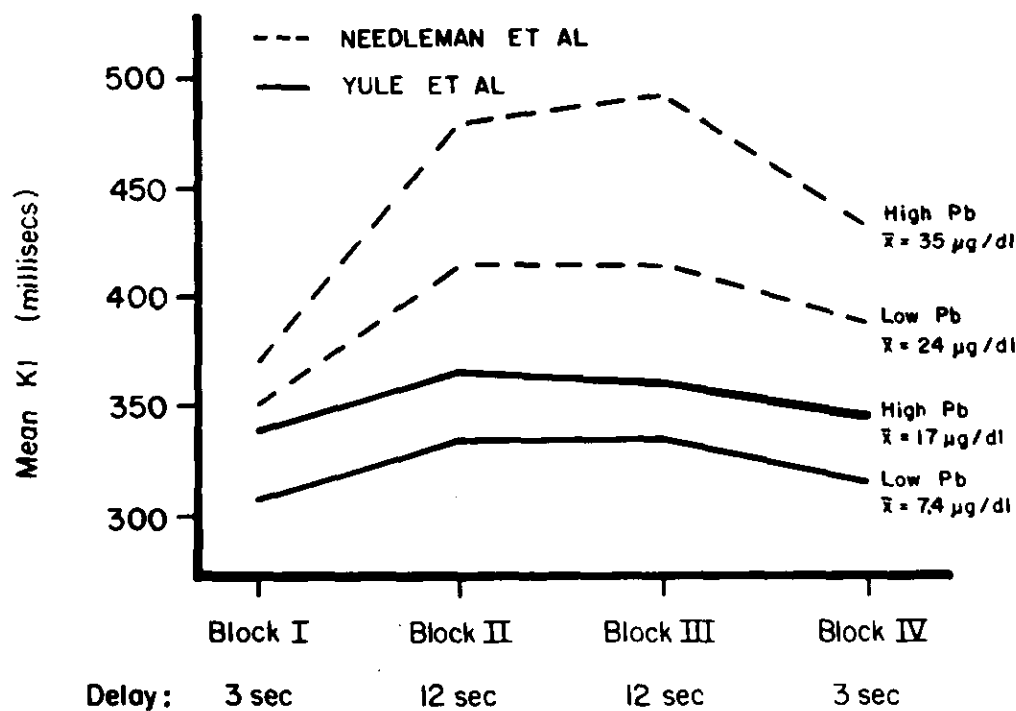


FIGURE 2. Comparison of reaction times at varying intervals of delay in subjects at four different exposure levels. Two separate studies are plotted here: Needleman et al. (1) and Hunter et al. (4). The data are ordered by mean blood lead level for each group.

will occur. D_{100} represents the uniformly lethal dose. At the far left of the graph is plotted a very sensitive response that begins at the lowest measurable exposure. One can visualize a family of curves, each representing a different outcome, each having its own threshold and frequency distribution. For some outcomes,

using lead as a paradigm, there will be no quarrel as to the health significance, e.g., death, brain hemorrhage, or renal tubular disease.

For other outcomes, the definition of adverse health effect is more ambiguous, and values play a significant role. Figure 4 displays intensity of any given effect

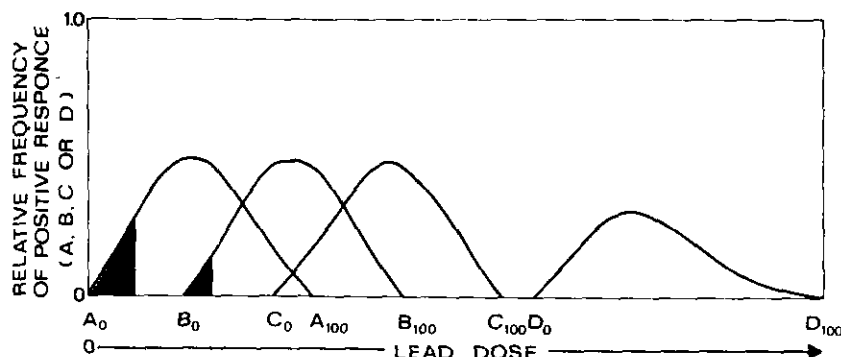


FIGURE 3. Frequency response of various outcomes in relation to internal dose of lead. From Hernberg (5).

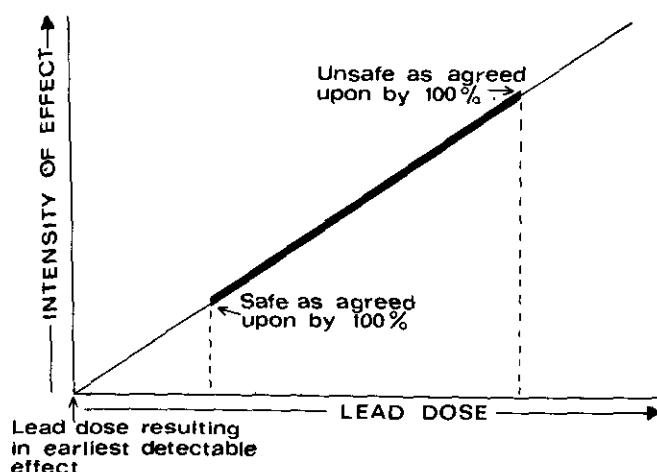


FIGURE 4. Intensity of any single effect in relation to lead dose. Influence of judgment on definition of adverse health effect.

against the dose of the agent. It shows that some outcomes (noncritical, nonrate-limiting) are very small degrees of effect that are of little health consequence to the host, and other outcomes are of undoubted health relevance. It is between these boundaries that the argument of health significance exists, and sometimes rages. It can be seen that in the positioning of these boundaries, questions of value inevitably insert themselves.

Sensitivity and Specificity of Markers

The use of markers in screening enterprises for diseases of low prevalence can produce surprising results if one is unaware of the prior probabilities and the sensitivity and specificity of the marker. Sensitivity and specificity are defined in Table 1.

Let us imagine a marker of quite high sensitivity (0.9), and excellent specificity (0.95) applied to a relatively rare disease, with a rate of 10/100,000. Table 2 shows that if a population of 100,000 were completely sampled, 5000 false positive diagnoses would be made

Table 1. Definitions of sensitivity^a and specificity.^b

Marker	Disease present	Disease absent
Present	a	b
Absent	c	d

^aSensitivity = $a/(a + c)$.

^bSpecificity = $d/(d + b)$.

Table 2. Use of a highly sensitive marker for a rare disease.^a

Marker	Disease present	Disease absent
Present	9	5000
Absent	1	949,900

^aDisease rate = 10/100,000.

Table 3. Use of a highly sensitive marker for a common disease.^a

Marker	Disease present	Disease absent
Present	900	4950
Absent	100	94,050

^aDisease rate = 1000/100,000.

in order to find 9 true cases of the disease. Table 3 applies the same analysis to a relatively common disease (rate = 1%), and shows that even with a test of this high quality, the false positive rate is 5.5 times the true positive rate. The rational use of markers to diagnose disease requires information on specificity, sensitivity, and the prior knowledge of the disease rate. This is a requirement too often ignored.

Type I and Type II Errors in Causal Inferences

The establishment of causal relationships is a central issue in the validation of disease representation. If a given outcome or marker is posited to be an effect of a neurotoxin, the establishment of a causal nexus is entailed in validating the status of the outcome as a marker. This is no simple task, particularly in obser-

vational studies with multiple predictor variables. There are two types of errors to be made. Scientists have long been justifiably concerned with avoiding type I or alpha errors: accepting spurious errors as real. This is felt to be defending scientific rigor; it avoids superstitious behavior, and limits the number of erroneous papers in the literature and the need for replications to invalidate spurious relationships.

Less attention has been paid to type II or beta errors: treating real relationships as spurious. Listed below are a few methodologic or epistemic solecisms frequently encountered which increase the chance of type II errors.

Making $p = 0.05$ Sacrosanct

Many authors interpret studies with p values of 0.06 or 0.1 as not significant and infer that no relationship exists in nature. R. A. Fisher, creator of the significance test, treated the value $p = 0.05$ differently in his book *The Design of Experiments*:

It is convenient to take this point [$p = 0.05$] as a limit in judging whether a deviation is to be considered significant or not. Deviations exceeding twice the standard deviation are thus formally regarded as significant.

It is relevant to note the use of the term "convenient." There is nothing sacred about this value. Jerome Cornfeld's comments on this point are worth noting:

[T]he prespecification of a significance level, e.g., .05 or .01 has no sound logical basis and remains unjustified.

The Building of Nonveridical Models (Overcontrol)

Variates that are measured in a study may be independent variables which affect the outcome under examination, or they may themselves be affected by lead. They may in fact occupy both positions in the causal chain. To control for such variates may be to subtract out variance which properly belongs to the main effect. Investigators are required at the least to report the results with and without controlling for these variates.

Making No-Effect Inferences from Samples of Inadequate Power

Many studies have reported finding no effect when the sample size chosen has inadequate power to find an effect if it were present.

Demanding Causal Proof

Two hundred years ago, David Hume stated that causality is a concept not susceptible to empirical dem-

onstration. Epidemiologists, and bench scientists as well, accept more modest goals for themselves: the accretion of incremental bits of data that assemble themselves into a coherent picture from which lawfulness can be inferred. They should not be burdened with a philosophically unreachable goal.

In the section to follow, the authors examine the neurobiologic and behavioral substrates that will provide the material from which to extract valid and efficacious markers of toxicity. Joseph Altman examines the effect of precisely timed doses of X-irradiation upon micro-neuronal migration and consequent behaviors. He has shown that precise measures of cellular migration can be correlated with altered behaviors that resemble attention deficit. Barry Hoffer discusses the model he has developed for studying neurogenesis and those factors which impede it. Bruce McEwen's studies of the effects of hormones and pseudohormones on brain anatomy have enriched our understanding of gender differences in structure and behavior, and imply that many behaviors may be markers for early CNS-hormone interactions. Edward Tronick's studies of early infant competence are among those which have sharpened our abilities to discover risk factors earlier in the course of a child's life and to follow the developmental elaboration of such risks. Alan Mirsky discusses his model of attentional function, drawn from his long and intensive studies in primate and seizure states in humans.

REFERENCES

1. Needleman, H., Gunnoe, C., Leviton, A., Reed, R. R., Peresie, H., Maher, C., and Barrett, P. Deficits in psychologic and classroom performance in children with elevated dentine lead levels. *N. Engl. J. Med.* 300: 584-695 (1979).
2. Lansdown, R., Yule, W., Urbanowicz, M. A., and Miller, I. Blood lead intelligence, attainment and behavior in school children: overview of a pilot study. In: *Lead versus Health* (M. Rutter and R. R. Jones, Eds.), John Wiley and Sons, New York, 1983, pp. 267-296.
3. Hatzakis, A., Salaminios, F., Kokevi, A., et al. Blood lead and classroom behavior of children in two communities with different degrees of lead exposure: evidence of a dose-related effect? *International Conference on Heavy Metals in the Environment*, Athens. CEP Consultants, Edinburgh, 1985, p. 47.
4. Hunter, J., Urbanowicz, M. A., Yule, W., and Lansdown, R. Automated testing of reaction time and its association with lead in children. *Int. Arch. Occup. Environ. Health* 57: 27-34 (1985).
5. Hernburg, S. Biological effects of low lead doses. *International Symposium on Environmental Health Aspects of Lead* (D. Barth, A. Berlin, T. Engel, R. Recht, and J. Smeets, Eds.), Commission of the European Communities Center for Information and Documentation, U.S. EPA, Luxembourg.
6. Cornfeld, J. Recent methodological contributions to clinical trials. *Am. J. Epidemiol.* 104: 408-421 (1976).