

# A Benchmark Dose Analysis of Prenatal Exposure to Polychlorinated Biphenyls

Joseph L. Jacobson,<sup>1</sup> James Janisse,<sup>1</sup> Mousumi Banerjee,<sup>2</sup> Jennifer Jester,<sup>1</sup> Sandra W. Jacobson,<sup>3</sup> and Joel W. Ager<sup>2</sup>

<sup>1</sup>Department of Psychology, Wayne State University, Detroit, Michigan, USA; <sup>2</sup>Center for Health Care Effectiveness Research and <sup>3</sup>Department of Psychiatry and Behavioral Neurosciences, Wayne State University School of Medicine, Detroit, Michigan, USA

Benchmark dose (BMD) analysis is used to determine levels of exposure to environmental contaminants associated with increased public health risk. In this study we used a benchmark approach to evaluate the risks associated with prenatal exposure to polychlorinated biphenyls (PCBs). We evaluated for intellectual impairment a cohort of children whose prenatal PCB exposure had been assessed from biologic specimens. We calculated BMDs and lower-bound confidence limits (BMDLs) for four end points using four sets of risk criteria. BMDLs were estimated using three different statistical methodologies. The BMDs and BMDLs were remarkably consistent across the four end points for each set of risk criteria, but differed substantially for the different risk criteria. The proportion of the sample considered at risk ranged from 9.8% for the least protective criteria to 74.1% for the most protective. Two methodologies, likelihood ratio and bootstrapping, generated generally similar BMDLs. BMD analysis provides a straightforward, reliable method for evaluating levels of exposure associated with increased public health risk. In the analyses performed in this study, the number of individuals considered at risk depended more on the risk criterion selected than on the outcome assessed. **Key words:** benchmark dose analysis, dose–response, environmental contaminants, *in utero* exposure, organochlorine contaminants, polychlorinated biphenyls, risk analysis. *Environ Health Perspect* 110:393–398 (2002). [Online 8 March 2002] <http://ehpnet1.niehs.nih.gov/docs/2002/110p393-398jacobson/abstract.html>

Polychlorinated biphenyls (PCBs) are a family of synthetic hydrocarbon compounds used extensively between the 1930s and mid-1970s for a variety of industrial purposes such as insulating materials and dielectric fluids in electric transformers and capacitors. Although banned in most industrial nations, they continue to be among the most ubiquitous environmental contaminants. Four prospective longitudinal studies have linked background levels of prenatal PCB exposure to poorer behavioral and intellectual functioning in infancy and childhood: the North Carolina study (1), the Michigan study (2), and studies in the Netherlands (3), and Oswego, New York (4,5). Prenatal PCB exposure is associated with less optimal newborn behavioral function (4,6), poorer infant recognition memory (5,7), lower levels of general intellectual competence during the preschool period [(3,8,9), but see Gladen et al. (10)], and poorer performance on standardized tests of verbal IQ and reading comprehension at age 11 years (2). These effects have been linked specifically to PCB exposure during the prenatal period; virtually no adverse effects have been observed in relation to postnatal exposure from breast-feeding.

Until recently, most risk assessments conducted for environmental regulation have attempted to identify a no-observed-adverse-effect level (NOAEL)—that is, an exposure level below which adverse effects are not observed. The NOAEL methodology

was designed for data from laboratory animal experiments in which groups of animals are administered discrete doses of a toxic agent. The NOAEL is defined as the highest dose at which no adverse effect is observed. The NOAEL approach is not well suited for data from studies relating human exposure to adverse effects, where the distribution of exposure levels is usually continuous. In this context it requires the imposition of arbitrary cut points on the continuous data to create discrete exposure groups, a procedure that can generate spurious thresholds in the data and can reduce the power to detect adverse effects.

Some human studies using linear multiple regression analysis to evaluate teratogenic effects have performed supplementary analysis in which the children are divided into discrete groups according to exposure level. Visual inspection of bar graphs displaying the group means generated by these analyses often suggests a threshold of exposure below which adverse effects are not observed. In Rogan et al. (6), for example, poorer newborn behavioral function was evident only in the 49 infants most heavily exposed to PCBs, the top 5.7% of the sample. In Jacobson and Jacobson (2), adverse effects of prenatal PCB exposure on verbal IQ were observed only in the top 16.8% of the sample, although poorer reading comprehension was observed at somewhat lower exposure levels. Adverse effects on intellectual competence were evident at even lower levels in the

Netherlands study (3), and poorer infant recognition memory was observed at lower exposure levels in Oswego (5) than in the Michigan study (7). One problem in assessing thresholds in human studies is that threshold values can vary considerably depending on the reliability of the measure of developmental outcome, the sensitivity of the end point at the age at which it is assessed, and the exposure group cut points selected by the researcher (11). Improvements in the reliability and sensitivity of biologic assays for PCBs are likely responsible for the detection of adverse effects at lower exposure levels in the Netherlands and Oswego than in the North Carolina and Michigan studies.

Given the high degree of between-study variability in threshold values and the fact that for some end points no thresholds are seen, it is often very difficult to identify definitive threshold values in human data. One alternative that has been used in risk assessments in recent years is benchmark dose (BMD) analysis (12). BMD analysis assumes that the relation between exposure and adverse outcome is continuous, thus avoiding the identification of specific threshold values, and uses statistical and policy criteria to establish cutoffs for risk assessment. The BMD approach begins by identifying a criterion for adverse effect. For some end points, such as IQ score, the criterion might be based on clinical criteria (e.g., an IQ < 70 is considered to constitute mental retardation). In most analyses, however, the criterion for adverse effect has been the bottom 5th percentile in the distribution of the test scores in a nonexposed population; the latter criterion is referred to as a  $p_0$  of 0.05. The BMD is defined as the level of exposure that

Address correspondence to J.L. Jacobson, Department of Psychology, Wayne State University, 71 W. Warren, Detroit, MI 48202 USA. Telephone: (313) 993-5454. Fax: (313) 993-3427. E-mail: jjacobso@sun.science.wayne.edu

The Michigan study was supported by grants R01-ES03256 and R01-ES05843 from the National Institute of Environmental Health Sciences/National Institutes of Health and grant CR80852010 from the U.S. Environmental Protection Agency. The study protocol was approved by the Wayne State University Human Investigation Committee, and informed consent was obtained from all participants.

Received 5 March 2001; accepted 10 October 2001.

will increase the risk of performance below the designated cutoff score by a prespecified amount (e.g., from 5% to 10% or from 5% to 15%). This increase is referred to as the benchmark response (BMR). Given a  $p_0$  of 0.05, a BMR of 0.05 represents a doubling of risk (i.e., from 5% to 10%), whereas a BMR of 0.10 represents a tripling (5% to 15%).

It can be argued that the  $p_0 = 0.05$  criterion used in most benchmark analyses to date is not sufficiently protective because it focuses exclusively on the increased risk of a deficit in the moderate-to-severe range. In our research on prenatal PCB exposure in a predominantly white, middle-class cohort in western Michigan, we found no evidence of moderate-to-severe deficit. All but one of the children performed in the normal range, and the one child who was mentally retarded was excluded on the grounds that she was an outlier (2). Nevertheless, our analyses showed that prenatal PCB exposure at or above 1.25  $\mu\text{g/g}$  was associated with a tripling of the incidence of poor performance in the low-normal range, defined as more than 1 standard deviation below the mean. Although the children performing in that range were not more likely to require special education services, it can be assumed that, given their low IQ and reading scores, they had to struggle to keep up in a normal classroom. In light of these findings, it seems reasonable also to use benchmark analysis to determine the level of exposure associated with an increase in relatively subtle deficits. If subtle deficit is defined as a score more than 1 standard deviation below the sample mean (i.e., in the bottom 16th percentile), the cutoff would be termed a  $p_0$  of 0.16 in benchmark methodology (13). Given a  $p_0$  of 0.16, a BMR of 0.05 would represent a 31.3% increase in risk (i.e., from 16% to 21%); a BMR of 0.10 would represent a 62.5% increase (16% to 26%).

One problem that arises in research on PCBs and other ubiquitous environmental contaminants is the difficulty of identifying a truly nonexposed population to determine the test score cutoff associated with a  $p_0$  of 0.05 or 0.16. Many neurobehavioral tests lack general population norms, and even where norms exist, such as for IQ tests, they may not be relevant for the population being assessed in a given study. In studies of populations in which virtually all individuals have some degree of exposure, we can extrapolate from the dose–response curve for the study sample to determine a test score that would correspond to the cutoff for the bottom 5th or 16th percentile in a truly nonexposed population. The value of the test score where the dose–response curve crosses the  $y$ -axis (at the point of zero exposure) would

be considered the mean for a hypothetical nonexposed population. A normal curve with that mean and a variance equal to the mean square error from the regression line can be used to represent the nonexposed population. From this distribution one can determine the cut point below which the bottom 5% or 16% of the hypothetical nonexposed population might be expected to score. If  $p_0$  is 0.05 and the BMR is 0.10, the BMD will be the level of exposure at which 15% (i.e., an additional 10%) of the population would be expected to score below the cut point. In benchmark analysis the BMDL, defined as the lower bound of the 95% confidence interval (CI) for the BMD, is used as the principal criterion for regulatory purposes. The BMDL is used instead of the BMD to provide a margin of safety and ensure that the most sensitive individuals in the population are protected.

Our study is the first to apply a BMD analysis to data on the effects of prenatal

exposure to PCBs. We examined four end points from the Michigan data set—three from the 11-year follow-up and one from the 4-year assessment. For one end point, full-scale IQ, we compared three methods for computing the BMDL: maximum likelihood estimate, likelihood ratio, and bootstrapping. For all four end points, we compared the BMDs and BMDLs using two cut points,  $p_0$  of 0.05 and 0.16, and two response criteria, BMRs of 0.05 and 0.10.

## Materials and Methods

**Sample.** The Michigan cohort was recruited in four maternity hospitals in western Michigan in 1980–1981, a period during which Lake Michigan fish were relatively heavily contaminated with PCBs (14). Two hundred forty-two mothers who had eaten at least 11.8 kg of Lake Michigan fish during the previous 6 years and 71 mothers who did not eat these fish participated in the newborn phase of the study. We used three biologic

**Table 1.** Sample characteristics, levels of exposure, and outcome score.

Characteristic	Total no.	Mean	SD	No. (%)
Sociodemographic characteristics				
Socioeconomic status <sup>a,b</sup>				
Executive or professional	203			36 (17.7)
Middle management or semiprofessional	203			66 (32.5)
Skilled workers, clerical, or sales	203			57 (28.1)
Semiskilled workers	203			38 (18.7)
Unskilled workers	203			6 (3.0)
Mother's age at child's birth (years) <sup>a,b</sup>	203	27.4	4.7	
Marital status—(no. married) <sup>a,b</sup>	203			170 (83.7)
Sex of child <sup>a,b</sup>				
Male	203			104 (51.2)
Female	203			99 (48.8)
Parity <sup>a,b</sup>	203	2.0	1.2	
Gravidity <sup>a,b</sup>	194	2.3	1.4	
No. of children <sup>a,b</sup>	203	2.6	1.1	
Maternal education (years) <sup>a,b</sup>	203	13.9	2.1	
Maternal vocabulary <sup>a,b</sup>	203	99.7	15.6	
HOME <sup>a,b</sup>	194	48.7	3.8	
Nursery school attendance <sup>b</sup>	194			41 (20.2)
Maternal employment (hours/week) <sup>b</sup>	194	14.0	16.5	
Family stress at 4-year interview <sup>b</sup>	194	4.1	1.9	
Family stress at 11-year interview				
Past year <sup>a</sup>	178	3.6	2.1	
Past 5 years <sup>a</sup>	178	3.8	2.1	
Child's grade in school at 11 years <sup>a</sup>	178	5.5	0.5	
Prenatal risk				
Maternal drinking before pregnancy <sup>a,b</sup>	178	0.1	0.3	
Maternal drinking during pregnancy <sup>a,b</sup>	178	0.0	0.1	
Maternal smoking before pregnancy <sup>a,b</sup>	178	0.4	0.6	
Maternal smoking during pregnancy <sup>a,b</sup>	178	0.2	0.4	
Delivery complications <sup>a,b</sup>	194			48 (24.7)
Levels of exposure				
Cord serum PCB concentration	144	2.6	2.0	
Maternal serum PCB concentration	158	5.7	3.6	
Maternal milk PCB concentration	124	829.7	384.3	
Duration of breast-feeding	203	24.6	29.5	
Outcome scales				
WISC full-scale IQ	178	107.6	12.6	
Wechsler reading word comprehension	177	100.9	11.6	
Mental rotation reaction time	177	1889.9	546.3	
McCarthy memory scale	194	51.3	8.7	

HOME, home observation for measurement of the environment.

<sup>a</sup>Tested for potential confounding in relation to the 11-year end points. <sup>b</sup>Tested for potential confounding in relation to the 4-year end point.

samples to assess prenatal PCB exposure: umbilical cord serum collected at delivery and maternal serum and milk collected shortly thereafter. A composite measure of prenatal exposure was constructed by averaging the data from these assays, as described below. The 4-year follow-up sample assessed here consisted of the 194 children assessed on the McCarthy Scales for Children's Abilities (15) for whom there were sufficient data to construct the composite measure of prenatal PCB exposure. The 11-year sample consisted of the 178 children tested on the Wechsler Intelligence Scale for Children-Revised (WISC-R) (16) for whom the composite prenatal exposure measure was available. Demographic characteristics, levels of exposure, and outcome scores for these children are summarized in Table 1.

**Measures of exposure.** Cord and maternal serum samples were obtained shortly after delivery, and maternal milk samples within 0.2–4.5 months postpartum (median = 1 month). All specimens were analyzed for PCBs by packed column gas chromatography, using the Webb-McCall method (17,18). Because of the limitations of this analytic methodology, PCB values were not detectable in 70% of the cord and 22% of the maternal serum samples. Because placental transfer provides the sole route of fetal exposure to these compounds, which are in equilibrium in fat deposits throughout the body, maternal serum and milk concentrations provide alternatives to cord serum for evaluating prenatal exposure (1). To improve reliability and sensitivity in the assessment of fetal exposure, the cord serum and maternal serum and milk values were converted to  $z$ -scores and averaged to provide a single measure; serum values were included only if they exceeded the detection limit (9). Eleven children whose cord and maternal serum PCB values were both nondetectable and for whom no milk specimen was available were assigned a prenatal exposure score at the 10th percentile of the distribution. The composite  $z$ -scores generated by these analyses were converted to their equivalent values in terms of maternal milk PCB concentration (micrograms per gram on a fat basis) by multiplying each  $z$ -score by the sample standard deviation for maternal milk PCB concentration (0.39) and adding the sample mean (0.84).

Because virtually all the children in this sample were exposed to measurable quantities of PCBs, we used the dose–response curve to extrapolate the test scores that would correspond to the cutoffs for the bottom 5th and 16th percentiles in a truly nonexposed population. To determine the equivalent of no exposure on our composite measure, the  $z$ -scores corresponding to zero

on each of the components of that measure were determined and averaged together, yielding a composite  $z$ -score of  $-1.65$ . The value of each end point on its dose–response curve corresponding to an exposure of  $-1.65$  was considered the mean for a hypothetically nonexposed population. For full-scale IQ, a normal curve derived from that mean and a variance equal to the mean square error of the regression yielded a cutoff of 92.9 for a  $p_0$  of 0.05 and 100.3 for a  $p_0$  of 0.16. Cutoffs for these two values of  $p_0$  were derived for each of the other end points as well.

**Outcome measures.** We performed BMD analyses on four cognitive outcomes previously found to be related to prenatal PCB exposure in the Michigan cohort (2,8). The three end points from the 11-year follow-up study were full-scale IQ from the WISC-R, word comprehension from the Woodcock Reading Mastery Tests–Revised (19), and average reaction time from Kail's (20) mental rotation task. In the mental rotation task, the child must determine whether a letter displayed on a computer screen is forward or backward (i.e., mirror image). Latency to press a forward or backward button on the computer is tabulated. Task difficulty is varied by rotating the stimulus letter at varying angles (e.g., 30°, 60°, 90°) clockwise from the vertical. One end point, the McCarthy Memory Scale (15), was examined from the 4-year assessment.

**Control variables.** Twenty control variables were evaluated for the analyses of each of the end points: the 18 indicated in Table 1, age of child when tested, and child examiner. We used correlational analysis to determine which control variables needed to be adjusted statistically to control for confounding (except for examiner, where one-way analysis of variance was used). Because a control variable cannot be the true cause of an observed deficit unless it is related to both exposure and outcome (21), association with either exposure or outcome can be used as the criterion for statistical adjustment. In this study we selected control variables in relation to outcome, as recommended Kleinbaum et al. (22). All control variables even weakly related to each outcome (at  $p < 0.10$ ) were controlled statistically by regressing the outcome in question on the control variables related to it. We used the residualized outcome scores in the benchmark analyses reported here. We added the mean value for each outcome to its residual score to restore it to its original units.

**Calculation of BMDs and BMDLs.** Calculation of the BMDs followed the framework given by Crump (12), a brief outline of which is presented here. For a given dose  $d$ , there is a resulting continuous

response  $X$  governed by a distribution function  $F$ . One of the parameters of  $F$ ,  $\Theta(d)$ , depends upon  $d$ , and the remaining parameters, represented by  $\alpha$ , do not. Given the distribution function  $F$  and the value of  $p_0$ , a cutoff score  $x_0$  can be calculated with responses more extreme than  $x_0$  considered abnormal. Using the above notation when larger responses are more adverse, the probability of having an abnormal response as defined by  $x_0$  given a subject's dose  $d$  is given by

$$P(d) = 1 - F[x_0; \Theta(d), \alpha],$$

where  $F$  is the distribution function that governs the continuous response  $X$ ;  $d$  is the dose the subject received;  $x_0$  is the score on the outcome that corresponds to the chosen  $p_0$ ;  $\Theta(d)$  is the dose-dependent parameter that describes the dose–response relationship; and  $\alpha$  is the vector of parameters that does not depend on  $d$ .

The BMD is calculated by solving the following equation:

$$F[x_0; \Theta(0), \alpha] - F[x_0; \Theta(BMD), \alpha] = BMR.$$

The outcomes assessed in this paper were all approximately normally distributed, and a Gaussian distribution function was used. The dose-dependent parameter  $\Theta(d)$  was assumed to be a linear function of  $d$ , and the standard deviation of the errors  $\alpha$  was assumed not to depend on the dose. We used maximum likelihood estimation (MLE) to obtain the regression estimates used to calculate the BMDs.

We used three methods to calculate the BMDLs. The first, MLE, is based on the asymptotic normal distribution of the maximum likelihood estimate of the BMD. The second is based on the asymptotic chi-square distribution of the likelihood ratio statistic, which Crump (12) recommends for computing the BMDL because it is more robust than MLE. For low-dose extrapolation problems, the distribution of the BMD is skewed and the lower limit based on the asymptotic normal approximation is unreliable. We calculated the likelihood ratio-based confidence limits using the profile likelihood method (23), a general approach for calculating confidence limits for parameters such as the BMD that are defined in terms of more basic parameters. The profile likelihood method reduces the log-likelihood function to a function of a single parameter of interest by treating the other parameters as nuisance parameters and maximizing over them. In the third method, a bootstrap approach (24), the BMD is estimated from 10,000 samples drawn at random (with replacement) from the data set. Each bootstrap sample has the



same  $n$  as the original study sample, but in any given bootstrap sample some subjects are randomly selected for inclusion more than once, while others are omitted. Using the percentile method (25), we calculated the BMDL as the cut point below which 5% of the bootstrapped BMD values fell.

## Results

Figure 1 presents a scatterplot relating the 11-year full-scale IQ scores (residualized for control variables and rescaled as described above) to the composite prenatal PCB exposure measure. Table 2 compares the BMDs and BMDLs for full-scale IQ for four  $p_0$ -BMR combinations, with the BMDLs derived using the three computational methods described above. Not surprisingly, the BMDs vary considerably depending on the criterion for adverse effect ( $p_0$ ) and the increase in risk (BMR) being stipulated. The BMD for the least protective criteria, a BMR of 0.10 for moderate-to-severe deficit ( $p_0 = 0.05$ ), is 1.44  $\mu\text{g/g}$ . Only 9.8% of the cohort were exposed above that level. The BMD for the most protective criteria, a BMR of 0.05 for subtle deficit ( $p_0 = 0.16$ ), is 0.58  $\mu\text{g/g}$ . Using those criteria, 74.1% of the Michigan sample would be considered at increased risk from prenatal PCB exposure. It is interesting that the BMDs for a smaller increase in severe risk ( $p_0 = 0.05$ , BMR = 0.05) are very similar to those for a greater increase in the risk of a subtler deficit ( $p_0 = 0.16$ , BMR = 0.10). The reason for this similarity is that both sets of criteria

represent a 0.35 standard deviation shift on the normal curve.

The three computational methods generated similar BMDLs for the least-protective criteria ( $p_0 = 0.05$ , BMR = 0.10), possibly because the values estimated by all three methods are well within the range of the observed data (Figure 1). By contrast, the BMDL estimates were quite different for the most protective criteria ( $p_0 = 0.16$ , BMR = 0.05), especially the MLE estimate, which was much lower than those generated by the other two methods. For the two intermediate sets of criteria, the likelihood ratio and bootstrap approaches provided generally similar BMDL estimates, whereas the MLE estimates were lower, confirming Crump's (12) concern that under some circumstances MLE estimates might be unreliable. One major advantage of the bootstrap approach is that it makes no assumptions about the shape of the BMD distributions. For that reason and because it is much less difficult to compute, we used bootstrapping to determine the BMDLs for the other three end points (Table 3).

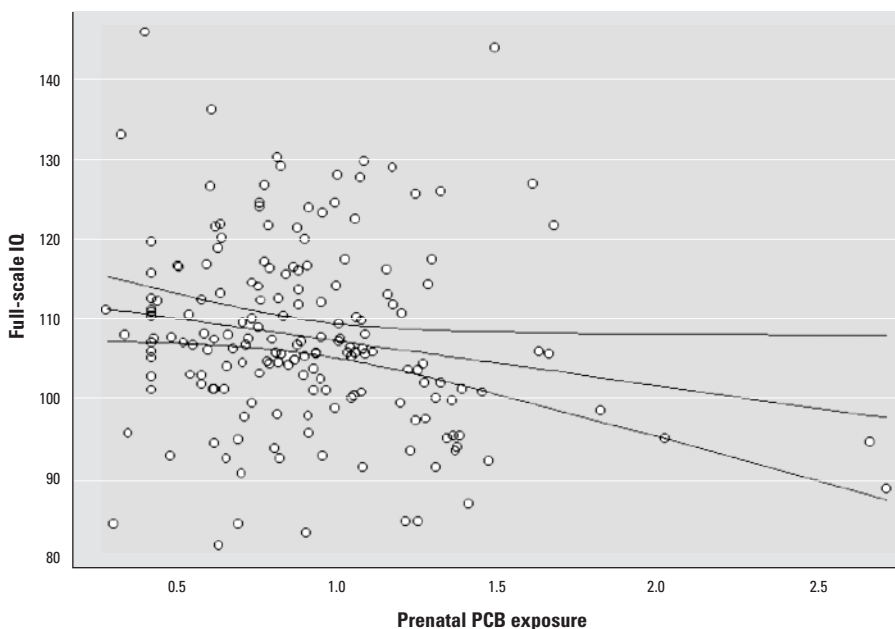
Figure 2 shows the dose-response relations generated when the four end points included in this article were examined in relation to prenatal PCB exposure divided *a priori* into five groups. For three of these end points, visual inspection of the bar graphs suggests that no adverse effect is seen below 1.25  $\mu\text{g/g}$ , whereas the threshold for reading word comprehension appears to be notably lower (1.0  $\mu\text{g/g}$ ). By contrast, with the BMD

methodology (Table 3), the BMDLs for reading word comprehension are virtually indistinguishable from those for full-scale IQ. Although the BMDLs for the 4-year McCarthy Memory Scale are consistently the highest across all four sets of analyses, the magnitude of the differences between the 4-year and 11-year BMDLs is very small.

## Discussion

BMD analysis is particularly well suited for risk assessment based on continuous data from human exposure studies, in which it is often difficult to identify discrete dose-response thresholds. It has been used, for example, in recent risk assessments for prenatal methylmercury exposure in both the United States (26) and Canada (27). Crump (12) recommends MLE analysis to generate the BMD values, but because standard methods for computing confidence limits based on MLE are often unreliable for calculating BMDLs, he recommends using likelihood ratio-based confidence limits for the BMDL. The latter are computationally complex, however, and not available in standard computer packages. This study is among the first to use a bootstrap approach to generate BMDLs. The likelihood ratio and bootstrap approaches generated similar BMDLs for three of the four sets of  $p_0$ -BMR criteria examined in this study. Two major advantages of bootstrapping are that it makes no assumptions regarding the distribution of the data, and it is relatively easy to program in packages such as SAS (28), S-PLUS (29), and Resampling Stats (30).

*A priori* division of the continuous distribution of prenatal PCB exposure levels into distinct exposure groups appears to reveal nonlinearities in the dose-response relationships (Figure 2) that might provide the basis for identifying a NOAEL. These nonlinearities may be misleading, however, because they may be influenced by the selection of



**Figure 1.** Dose-response regression line and 95% CIs relating full-scale IQ (residualized for the control variables indicated in Table 3) to prenatal PCB exposure, based on the asymptotic normal distribution of the maximum likelihood estimate. The confidence limits flare at both ends of the distribution because the regression line is measured most reliably at the center of the dose-response distribution.

**Table 2.** BMD analyses for full-scale IQ scores comparing three methods for computing the BMDL.<sup>a</sup>

	BMR = 0.05	BMR = 0.10
$p_0 = 0.16$		
BMD	0.58	0.91
BMDL		
Maximum likelihood	-1.80	0.26
Likelihood ratio	-0.20	0.56
Bootstrap	0.42	0.62
$p_0 = 0.05$		
BMD	0.94	1.44
BMDL		
Maximum likelihood	0.36	1.08
Likelihood ratio	0.50	1.10
Bootstrap	0.63	0.93

<sup>a</sup>Values are composite measures of prenatal PCB exposure expressed in terms of micrograms of fat in maternal milk. Full-scale IQ scores were residualized for the control variables listed in the footnote to Table 3.

the exposure group cut points, which have no biologically meaningful basis. Because the number of children in each group is relatively small, there is also a risk that a few individuals may unduly influence a group mean.

A major advantage of the benchmark methodology is that the BMD and BMDL are derived from the slope of the entire dose–response function. Although most benchmark analyses performed to date have used a linear model, this methodology can also be applied to nonlinear dose–response curves (12). In this study the BMDs and BMDLs derived from benchmark analysis were remarkably consistent across four different end points at two different ages. Benchmark analyses performed on data sets from three different studies of prenatal methylmercury exposure also showed remarkable cross-end point consistency within each of the studies examined (27). The BMDs and BMDLs differed considerably, however, among the three methylmercury studies. Similarly, we would anticipate that BMDs and BMDLs derived from the more recent Netherlands (3) and Oswego studies (4,5), which used more sensitive measures of prenatal PCB exposure than were available for the Michigan study (2), might well be lower than those reported here.

Another methodology that has sometimes been used to identify no-effect levels is nonparametric regression (31,32). Nonparametric regression fits a series of curves in overlapping segments corresponding to small regions of the dose–response relationship, to which a scatterplot smoothing technique is then applied. The resulting curve is often nonlinear, relatively flat at lower levels of exposure, and becomes steeper as exposure increases. However, it is usually not possible to discern a discrete no-effect cutoff in this continuous distribution. Moreover, because each of the segments used to construct the nonparametric curve is based on a relatively small number of cases, the curves generated for different end points frequently bend at quite different thresholds (32).

Besides providing BMD criteria likely to be stable across a range of end points (at least within a given study), the BMD approach enables the risk assessor to derive regulatory standards for exposures for which no reliable NOAEL values have been detected. The absence of a reliable NOAEL is also common in laboratory animal experiments, where adverse effects are often evident even at the lowest doses tested. Moreover, a NOAEL derived from one study is often superseded by evidence of adverse effects at lower exposure levels in subsequent studies that use more reliable exposure measures or more sensitive end points. In most cases,

therefore, it may be most reasonable to assume that there is no true no-effect level and to derive BMDs and BMDLs from the dose–response curve.

One important feature of the benchmark methodology is that it focuses the attention of the policy maker more directly on the cost–benefit tradeoffs that environmental

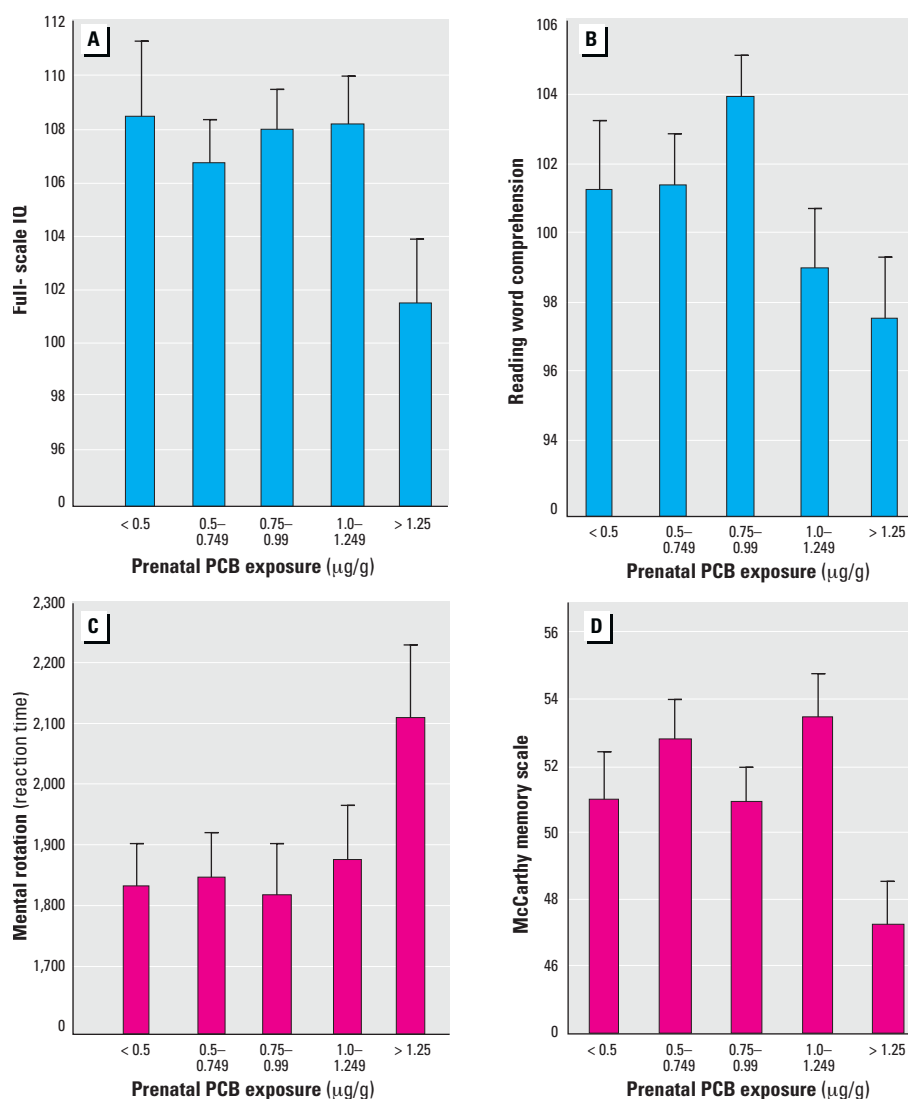
**Table 3.** BMD analyses for four cognitive end points based on the bootstrap computation.<sup>a</sup>

	BMD		BMDL	
	BMR = 0.05	BMR = 0.10	BMR = 0.05	BMR = 0.10
$p_0 = 0.16$				
Full-scale IQ <sup>b</sup>	0.58	0.91	0.42	0.62
Reading word comprehension <sup>c</sup>	0.57	0.90	0.44	0.64
Mental rotation <sup>d</sup>	0.63	1.00	0.43	0.63
McCarthy memory scale <sup>e</sup>	0.64	1.02	0.46	0.69
$p_0 = 0.05$				
Full-scale IQ <sup>b</sup>	0.94	1.44	0.63	0.93
Reading word comprehension <sup>b</sup>	0.93	1.43	0.65	0.95
Mental rotation <sup>d</sup>	1.03	1.59	0.64	0.94
McCarthy memory scale <sup>e</sup>	1.05	1.63	0.71	1.04

<sup>a</sup>Values are composite measures of prenatal PCB exposure expressed in terms of micrograms of fat in maternal milk.

<sup>b</sup>Residualized for socioeconomic status, maternal education, maternal vocabulary, and HOME Inventory. <sup>c</sup>Residualized for socioeconomic status, child's sex, parity, maternal education, maternal vocabulary, HOME Inventory, maternal alcohol consumption and smoking during pregnancy, and child's age when tested. <sup>d</sup>Residualized for child's grade.

<sup>e</sup>Residualized for socioeconomic status, maternal age, child's sex, maternal education, maternal vocabulary, HOME Inventory, and child's age when tested.



**Figure 2.** Relation of each developmental end point (residualized for the control variables indicated in Table 3 and rescaled as described in the text) to prenatal PCB exposure divided *a priori* into five groups.

regulation almost always entails. Given that it is unlikely to be economically feasible to completely rid the environment of every substance for which there is some evidence of adverse effect, the benchmark approach requires the regulator to determine the level of risk of increased adverse effect he or she is willing to tolerate. The data reported here make clear how dramatically the selection of regulatory criteria can alter the proportion of the population deemed to be at risk. Analyses based on visual inspection of the dose-response data (Figure 2) suggest that 15–35% of the Michigan cohort was at risk from prenatal PCB exposure. Based on the least-protective benchmark criteria we tested, only about 10% of the Michigan cohort would have been considered at risk, whereas using our most stringent criteria, almost 75% would have been deemed at risk. It is surprising that the least-protective criteria examined in this paper are the ones that have been used most frequently in benchmark analyses performed to date (33,34). These criteria (a  $p_0$  of 0.05 and a BMR of 0.10) are designed to protect against the tripling of the risk of moderate-to-severe deficit (from 5% in a nonexposed population to 15%). It would seem more appropriate to tolerate at most a doubling of the incidence of moderate-to-severe deficit and/or to base risk assessment on the prevention of more subtle deficit.

## REFERENCES AND NOTES

- Rogan WJ, Gladen BC, McKinney JD, Carreras N, Hardy P, Thullen J, Tinglestad J, Tully M. Polychlorinated biphenyls (PCBs) and dichlorodiphenyl dichloroethene (DDE) in human milk: effects of maternal factors and previous lactation. *Am J Public Health* 76:172–177 (1986).
- Jacobson JL, Jacobson SW. Intellectual impairment in children exposed to polychlorinated biphenyls *in utero*. *N Eng J Med* 335:783–789 (1996).
- Pantandin S, Lanting CI, Mulder PGH, Boersman ER, Sauer PJJ, Weisglas-Kuperus N. Effects of environmental exposure to polychlorinated biphenyls and dioxins on cognitive abilities in Dutch children at 42 months of age. *J Pediatr* 134:33–41 (1999).
- Stewart P, Darvill T, Lonky E, Reihman J, Pagano J, Bush B. Assessment of prenatal exposure to PCBs from maternal consumption of Great Lakes fish: an analysis of PCB pattern and concentration. *Environ Res Sec A* 80:S87–S96 (1999).
- Darvill T, Lonky E, Reihman J, Stewart P, Pagano J. Prenatal exposure to PCBs and infant performance on the Fagan Test of Infant Intelligence. *NeuroToxicol* 21:1029–1038 (2000).
- Rogan WJ, Gladen BC, McKinney JD, Carreras N, Hardy P, Thullen J, Tinglestad J, Tully M. Neonatal effects of transplacental exposure to PCBs and DDE. *J Pediatr* 109:335–341 (1986).
- Jacobson SW, Fein GG, Jacobson JL, Schwartz PM, Dowler JK. The effect of intrauterine PCB exposure on visual recognition memory. *Child Dev* 56:853–860 (1985).
- Jacobson JL, Jacobson SW, Humphrey HEB. Effects of *in utero* exposure to polychlorinated biphenyls and related contaminants on cognitive functioning in young children. *J Pediatr* 116:38–45 (1990).
- Jacobson JL, Jacobson SW. Evidence for PCBs as neurodevelopmental toxicants in humans. *NeuroToxicol* 18:415–424 (1997).
- Gladen BC, Rogan WJ. Effects of perinatal polychlorinated biphenyls and dichlorodiphenyl dichloroethene on later development. *J Pediatr* 119:58–63 (1991).
- Jacobson JL, Jacobson SW. Prenatal alcohol exposure and neurobehavioral development: Where is the threshold? *Alcohol Health Res World* 18:30–36 (1994).
- Crump KS. Calculation of benchmark doses from continuous data. *Risk Anal* 15:79–89 (1995).
- Butz-Jorgensen E, Grandjean P, Keiding N, White RF, Weihe P. Benchmark dose calculations of methylmercury-associated neurobehavioural deficits. *Toxicol Lett* 112–113:193–199 (2000).
- Jacobson JL, Jacobson SW, Fein GG, Schwartz PM, Dowler JK. Prenatal exposure to an environmental toxin: A test of the multiple effects model. *Dev Psychol* 20:523–532 (1984).
- McCarthy D. Manual for the McCarthy Scales of Children's Abilities. New York:Psychological Corporation, 1972.
- Wechsler D. Wechsler Intelligence Scale for Children-Revised. New York:Psychological Corporation, 1979.
- Sawyer LD. Quantitation of polychlorinated biphenyl residues by electron capture gas-liquid chromatography. *J Assoc Off Anal Chem* 61:272–291 (1978).
- Needham LL, Burse VW, Price HA. Temperature-programmed gas chromatographic determination of PCB and PBB serum. *J Assoc Off Anal Chem* 64:1131–1137 (1981).
- Woodcock Richard W. Woodcock Reading Mastery Tests—Revised. Circle Pines, MN:American Guidance Services Inc, 1987.
- Kail R. Sources of age differences in speed of processing. *Child Dev* 57:969–987 (1986).
- Schlesselman J. Case-control studies: Design, conduct, analysis. New York:Oxford University Press, 1982.
- Kleinbaum DG, Kupper LL, Muller KE. Applied regression analysis and other multivariable methods. 2nd ed. Boston:PWS-Kent, 1988.
- McCullagh P, Nelder JA. Generalized Linear Models. 2nd ed. London:Chapman Hall, 1989.
- Efron B, Tibshirani RJ. An Introduction to the Bootstrap. New York:Chapman & Hall/CRC, 1998.
- Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Stat Med* 9:1141–1164 (2000).
- Keating MH. Mercury Study for Congress. Vol I: Executive Summary. EPA-452/R-97-003. Research Triangle Park:U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards and Office of Research and Development, 1997.
- National Research Council. Toxicological Effects of Methylmercury. Washington, DC:National Academy Press, 2000.
- SAS. Version 6.12. SAS Macro Language: Reference. Cary, NC:SAS Institute Inc, 1996.
- Data Analysis Products Division. S-PLUS 2000 Programmer's Guide. Seattle, WA:MathSoft Inc, 1999.
- Resampling Stats, Inc. Resampling Stats Standalone User's Guide. Arlington, VA:Resampling Stats Inc, 2000.
- Cox C, Clarkson TW, Marsh DO, Amin-Zaki L, Tikriti S, Myers GJ. Dose-response analysis of infants prenatally exposed to methylmercury: an application of a single compartment model to single strand hair analysis. *Environ Res* 31:640–649 (1989).
- Jacobson JL, Jacobson SW, Sokol RJ, Ager JW. Relation of maternal age and pattern of pregnancy drinking to functionally significant cognitive deficit in infancy. *Alcohol Clin Exp Res* 2:345–351 (1998).
- Crump KS, Kjellstrom T, Shipp AM, Silvers A, Stewart A. Influence of prenatal mercury exposure upon scholastic and psychological test performance: benchmark analysis of a New Zealand cohort. *Risk Anal* 18:701–713 (1998).
- Crump KS, Van Landingham C, Shamlaye C, Cox C, Davidson PW, Myers GJ, Clarkson TW. Benchmark concentrations for methylmercury obtained from the Seychelles Child Development Study. *Environ Health Perspect* 108:257–263 (2000).