CIRA Methodology and Biostatistics Seminar Series

Exposure to Polychlorinated Biphenyls (PCBs) and Health Status: A Critical Look at the Data

Dom Cicchetti, Ph.D.

Yale University School of Medicine

A comprehensive set of objective scientific criteria was developed to examine the widely accepted claim that exposure to PCBs both prenatally and postnatally produces cognitive, neuropsychological, and behavioral deficits in both children and adults. These scientific criteria assessed: (1) Whether reliable and valid measuring instruments were employed; (2) Whether appropriate reliability and validity assessments were made on study variables; (3) Whether corrections were used for multiple comparisons; (4) Whether the authors controlled for fundamental confounding variables; (5) Whether a distinction was made between statistical and clinical significance; and (6) Whether the studies were longitudinally designed and appropriately analyzed. It will be demonstrated that the numerous claims regarding the putative deleterious effects of maternal PCB exposure upon subsequent cognitive, neuropsychological, and behavioral development whether prenatally, or postnatally, in children, adolescents, or in adults, are not supported by the data.

Thursday, March 27th, 2003
2:30 - 3:45 - CIRA conference room
40 Temple Street, Suite 1B

CIRA

Yale Uni

Yale University Center for Interdisciplinary Research on AIDS

Outline of Presentation:

- 1. Purpose of Presentation
- 2. Brief History of PCBs in the Environment
- 3. The Types of Questions That Were Investigated
- 4. Data From Six Cohorts
- 5. The Set of Scientific Criteria
- 6. Application of Scientific Criteria
- 7. Summary and Conclusions
- 8. Broader Implications of the Critical Review

Purpose of Presentation:

- A Critical Examination of Research Relating Exposure to Polychlorinated Biphenyls (PCBs) to:
- 1. Neurobehavioral
- 2. Health-Related and
- 3. Cognitive Deficits

Major Source: Cicchetti, Kaufman, & Sparrow (in press, 2003)

Definition, Brief History of Polychlorinated Biphenyls (PCBs):

1. PCBs –synthetic chlorinated hydrocarbon compounds used mostly as insulating material for electrical transformers and capacitors
2. Fat-soluble
3. Some have long half-lives in humans
4. Cross the placenta
5. Present in breast milk so that
6. Both the fetus and nursing infant are exposed
7. Significant source of human exposure said to be consumption of fatty sports fish from contaminated waters (e.g., Lake Michigan)
8. Introduced in 1930
9. Used worldwide
10. Banned in U.S. , most industrialized nations in 1970s

The Critical Sources Of PCB Data Over Past Two Decades:

Six Major Cohorts of Infants and Children:

To Answer Whether:

PCB Exposure Is Significantly Associated With:

Intellectual, Neuropsychological, Behavioral, and/or Other Health-Related Problems

The Study Periods:

- Prenatally- Via PCB levels in Mothers' Umbilical Cord
- Infancy- VIA PCB Levels in Breast Milk and
- Childhood-Via Blood Samples

Location of the Six Infant-Child Cohorts: Study Period: Since 1983 and Continuing

- 1. Michigan (Jacobson, Jacobson, et al.)
- 2. North Carolina (Rogan, Gladen, et al.)
- 3. Holland (Koopman-Esseboom, Lanting, Patandin, et al.)
- 4. German (Winneke, et al.)
- 5. Oswego (Lonky, Stewart, Darvill, et al.)
- 6. Faeroe Islands (Steuerwald, et al.) –21 Islands Between England and Iceland

What The Authors of the Six Cohorts Conclude:The Claim

- Each Study:
- Alone-
- In Tandem-
- Consistent With Results of Animal PCB Studies-
- Show Evidence of Association Between PCB Exposure and Later Functioning

Some Quotations From The Authors: The Michigan Cohort

- "The Data Reported to Date Are Consistent With the Hypothesis that Parental Exposure to PCBs and Related Compounds Can Cause Persistent Changes in the Developing Brain that Adversely Affect Cognitive Function, at Least Through School Age" (Jacobson & Jacobson, 1997, p.345).

The Same Authors Concede That:

"None of the Prospective Studies Has Found an Increased Frequency of Mental Retardation" (Op. cit., 345)

From The Dutch Cohort:

"Our Data Demonstrate the Continuation of a Toxic Impact Received in Utero on Cognitive Functioning at Toddler Age"
(Patandin, et al., 1999, p.40)

Goal of Today's Presentation:

To Critically Evaluate The Claims About the Putative Adverse Effects of Ingesting PCBs at the Reported Levels of Exposure

<u>Method of Evaluation:</u>

Application of Six Objective Scientific Criteria to Each of the Published PCB Investigations

Rationale For Development and Application of the Six Criteria:

Their Critical Importance is That They Serve, In General, to Guard Against Threats to Both The Reliability and Validity of Reported Study Results

ANOTHER SCIENTIST EVALUATES OUR SIX CRITERIA:

A QUOTATION FROM THE HARVARD NEUROPSYCHOLOGIST NANCY HEBBEN (in press, 2003):

"These criteria are crucial in the design of sound scientific studies and in the appropriate analysis of study results. Errors in design and analysis weaken, and in some cases, eliminate the reliability and validity of scientific data."

The Criteria:

In the Form of Answers to Six Scientific Queries

The First Query:

Were State-of-the-Art Measuring Instruments and Techniques Consistently Applied?

The Second Query:

Were Appropriate Reliability Assessments Made on Study Variables?

The Third Query:

Did the Authors Control Appropriately for Findings that Could Have Occurred by Chance Alone?

The Fourth Query:

Were Attempts Made To Differentiate Statistically Significant Findings From Those That Were Both Statistically And Clinically Meaningful?

The Fifth Query:

Were Appropriate Attempts Made To Control For Potential Confounding Variables?

The Sixth Query:

Were The Studies, As Intended By The Authors, Longitudinally Designed And Analyzed Appropriately?

IN RESPONSE TO EACH QUERY:

A “YES” Response Indicated That The Criterion Was Met

WHILE,

A "NO" Response Indicated That The Criterion Was NOT Met

GENERAL FINDINGS:

NOT A SINGLE STUDY MET ALL SIX SCIENTIFIC CRITERIA

General Findings Across the Six Cohorts- Failure to Satisfy Criteria 1 and 2:

Investigators, As A Rule:

1. Selected Outdated and Poorly Normed Measuring Instruments Over More Recent Ones With Solid Psychometric Properties
2. Failed to Appropriately Establish the Reliability and Validity of the Poorly Chosen Test

General Findings Across the Six Cohorts- Failure to Satisfy Criteria 3 and 4:

3. Performed A Multitude of Data Analyses Until One or More Finally Reached a Level of Statistical Significance at the Conventional p=.05 Level and then Interpreted This As a Real or Valid Result

4. Interpreted Differences Between High and Low PCB Exposure Groups On Study Outcome Variables As Clinically Meaningful Even When They Were Within the Range of the Error of the Measuring Instrument

General Findings Across the Six Cohorts- Failure to Satisfy Criteria 5 and 6:

5. Failed to Control Adequately For Confounding Variables and

6. Conceived/Treated A Longitudinal Research Investigation As a Series of Independent Cross-Sectional Studies

SPECIFIC FINDINGS: CRITERION 1:

Were State-of-the Art Measuring Instruments And Techniques Consistently Used By The Authors?

NO.

Examples of Failure To Meet Criterion 1 In General:

1. The Instruments Used To Measure Outcome Variables Typically Had Very Poor Psychometric Properties: Unacceptably Low Levels of Reliability and/or Validity, or in Most Studies No Tests of Reliability Were Performed At All.

Specific Examples of Failure To Meet Criterion 1:

1. In the Earlier Studies, PCBs Were Often Not Measured With Great Accuracy:

<u>In the Authors" Own Words:</u>

" ... this inconsistency is presumably due to limitations in reliability of measurement. The packed column gas chromatographic analysis used here, although state of the art in the mid-1980s, provides relatively low resolution and high detection limits...In light of recent refinements in analytic methodology, replication of the present findings is warranted using more sensitive analytic techniques" (Jacobson, et al., 1992).

Yet, Ignoring Their Own Caveats:

The Jacobson's have continued to use these questionable data , based upon self-admitted unreliable and invalid PCB measurement instruments in more recent publications (Jacobson & Jacobson, 1996; Jacobson, et al., 2002).

Serious Problems Beyond Unreliability Of PCB Measurement:

Jacobson & Jacobson (1996, Table 7) :

Depended Solely Upon Maternal Recall For Information on The Following Variables:
Birth weight, Gestational Age, Weight Gain During Pregnancy, Smoking During Pregnancy, Smoking Before Pregnancy, Absolute Alcohol Per Day Prior To And During Pregnancy, And Both The Amount And Type Of PCB-Contaminated Fish Consumption Over A 5 Year Period During The Previous 16 Years

Specific Findings for Criterion 2: Were Appropriate Reliability Assessments Used on Study Variables? The Data

1. Specific Reliability Assessments of Study Variables Were Seldom Made

2. When Made, the Attempts Were Always Unsuccessful

Specific Findings for Criterion 2: Were Appropriate Reliability Assessments Used on Study Variables? The Data

3. Not a Single Study Deriving From Any of the Six Cohorts Ever Utilized A State-of-the-Art Reliability Statistic, such as an Appropriate Model of Kappa, Weighted Kappa or the Mathematically Similar Intraclass Correlation Coefficient

Specific Findings for Criterion 2: Were Appropriate Reliability Assessments Used on Study Variables? The Data

4 . Instead, the Invalid Pearson Product Moment Correlation Coefficient (PMCC) Was Used, Despite Its Known Potential for Exaggerating the Extent of Reliability of Any Given Variable to Which it is Applied.

SOME EXAMPLES:

1. The Brazelton (1984) Neonatal Behavior Assessment Scale (NBAS) Was Used to Assess the Putative Effects of PCB Exposure Upon Infant Behaviors.

Using the PPMC, the Reported Reliability Levels for Range of State, and Reflexes from the Original NBAS and its Revised Version Were -.13 (below 0), .07, and .05!

A Second Example:

2. The Fagin Test of Infant Intelligence (FTII) Has Also Been Utilized to Assess Putative Effects of PCB Exposure Upon Infant's Ability to Choose Novel Over Standard Stimuli.

Winneke et al. (1998, p. 425) reported:

"The re-test reliability of the novelty score for the mobile version in our study after 2 weeks for two observers and 10 children was found to be almost zero and even negative (R= -.195)."

A Third Example:

3. Jacobson, et al (1985) and Darvill et al. (2000) never reported inter-observer reliability data for the FTII.

They merely concluded that there were no statistically significant systematic differences in examiners on the FTII scores they obtained. They clearly implied that this meant that reliability levels were high.

The Error in Their Reasoning:

The scores of two independent examiners can be widely different, and still produce the exact same mean or average values.

As one example, if three subjects are assigned scores of 1 ,7, and 10 by the first examiner, and 5, 10, and 3 by the second, their means will be identical at 6.00, yet their level of chance corrected agreement (Ri) will only reach -.14 .

A Fourth Example:

4. In Their Discussion of The Bayley Scales of Infant Development to Measure Putative Effects of PCB Exposure Upon Mental and Psychomotor Development The Authors Said:

"We did not carry out formal intra-rater or inter-rater testing, but we did monitor scores and discuss any drift at meetings with study personnel" (Gladen, et al., 1988, p.992).

Specific Findings for Criterion 3: Did the Authors Control Adequately for Chance Findings? The Data

Overall Finding:

With Very Few Exceptions, Large Numbers of Statistical Tests Were Performed Without Controlling AT ALL for the Number That Could Have Been Expected by Chance Alone

Specific Findings for Criterion 3- Some Examples:

1. Jacobson & Jacobson (1997, Figure 5):

Attempted to Control for Multiple Comparisons in Their Analysis of WISC-R Full Scale IQ Data (as an Outcome Variable, Putatively Affected by PCB Exposure) BUT:

Used the Duncan Multiple Range Test , A Procedure That OVERESTIMATES the Number of Statistically Significant Results in a Given Investigation (Cicchetti, 1994; Petrinovich & Hardyck, 1969; Toothaker, 1991).

It Should Be Noted At This Point That:

The Authors In the Six Cohorts Had A Penchant For Consistently Choosing Those Statistical Techniques (Invalid as They All Were) That would Increase the Probability Of Producing a Statistically Significant Result that Would Implicate PCB Exposure As the Cause of Neuropsychological and Behavioral and Other Health Problems.

Another Example:

To Compound Results Further, Some Investigators, Such As , Again, Jacobson & Jacobson, 1997), Not Only failed To Use Appropriate Multiple Range Tests, But When Results Were Not Statistically Significant Using A Required Two Tailed Test . . .

Why They Just:

Doubled Their Chances of Finding A Statistically Significant Result By Simply Employing A One-Tailed Test .

This Continues To Be Justified By The "Argument" That the Investigation Under Study Is An "Exploratory" or "Pilot " Investigation (Following About 2 Decades Of Published PCB Research)

Another Example:

In an Attempt to Show the Effects of PCB Exposure Upon Cognitive Deficits, Gladen & Rogan (1991) Conducted 40 Multivariate Analyses. None of Them Produced Statistically Significant Results.

Yet Another Example:

Across Three of the Six Cohorts:

96 Analyses Were Conducted, Again, to Determine The Effects of PCB Exposure Upon Children'sCognitive Functioning. Four Significant Results Occurred Before Controlling For Chance!

Further Examples:

The Weakest of All Strategies is for Investigators to Finally Interpret Non-Significant Results As IF They Were Statistically Significant: (e.g., .087 in Jacobson & Humphrey, 1990; and .09 in Rogan & Gladen, 1991)

How The Strategy Works:

1. Attempt To Control For Chance Using Appropriate Multiple Comparison Procedures (e.g., Bonferroni Correction)

2. If That Doesn't Work, Try Multiple 2 Tailed Tests

3. If This Doesn't Work, Try One-Tailed Tests, AND, . . .

If None Of These Three Strategies Works:

Then, Applying Abelson's (1995, p.55) Delightful Parody On The Researcher Desperate To Find A Statistically Significant Result:

"State the Actual p Value, But Talk Around It ."

Specific Findings for Criterion 4: Were Attempts Made To Differentiate Between Statistical And Clinical Significance? Usually No. Examples:

The Most "Robust" Finding, A Four Point Difference Between Highest and Lowest PCB Exposure Groups in Cognitive Performance was Taken in the Dutch Cohort to Be Clinically meaningful (Pantandin, Lanting, et al., 1999).

Yet, The Error of Cognitive Measuring Instruments Is Known to Be of The Order of ±5 to ±6 Points on a Simple Re-Testing of The Same Individual By the Same Clinical Examiner (Kaufman, 2001) .

Other Failures To Differentiate Clinical From Statistical Significance:

Contaminated Fish Eaters, Compared to Control Mothers: Had Slightly Smaller:

-Birth Weights (7.6 lbs vs 8.1 lbs)-Both Normal

-Head Circumference (34.92cm vs 35.48 cm

-Gestational Age (40.31 wks vs 40.82 wks)

(Fein, Jacobson, et al., 1984)

Specific Findings for Criterion 5: Were Appropriate Attempts Made to Control For Potential Confounding Variables ? No.

Potential Confounding Variables That Were Most Usually Not Controlled Adequately or Not At All:

1. Parents' IQ or Other measure of Cognitive Level (the North Carolina Cohort)
2. Home Environment (Grandjean, et al., 2001)
3. Prenatal Factors (e.g., maternal alcohol, smoking, drug status)
4. Levels of Methyl Mercury in Contaminated Fish

Specific Findings for Criterion 6: Were The Studies Longitudinally Designed And Analyzed Appropriately ?

Without Exception, the Investigators in Each of the Six Cohorts Describe Their Research As Longitudinally Designed:

Yet ,All Authors in All Cohorts Report Only Cross-Sectional Study Results At Each Follow-Up Assessment Period.

To Obviate This Problem:

The Authors Should Have Used A Strategy Such As Individual Growth Curve Analysis, or in The Case of Little or No Missing Data Some Variant of A Repeated Measures Analysis of Variance Design (Cicchetti, Kaufman, & Sparrow, 2003)

Another Critical Methodologic Flaw:

Authors Failed to Define What Is Or Is Not a Toxic Level of PCB Exposure.

Rather, They Defined High , Low , and Moderate Levels of PCB Exposure, Not IN Absolute Levels, But Relative to What was Considered , High, Moderate, or Low in a Given Cross-Sectional Study

Another Critical Methodologic Flaw:

What Was Defined As High, Low, or Moderate PCB Exposure Differed Within A Given Cohort, From One Cross-Sectional Study to Another (Especially True of the Jacobson, et al., Cohort), As Well As Between Cohorts (Cicchetti, Kaufman, & Sparrow, in press, 2003; Schell, Budinsky, & Wernke, 2001)

Summary And Conclusions:

Based Upon An In-Depth Analysis of More Than 100 Published PCB and Related Articles:

The Multiple Claims and Conclusions Regarding the Putative Deleterious Effects of PCB Exposure Upon Subsequent Cognitive, Neuropsychological and Behavioral Development in Their Offspring, Whether the Exposure is *in utero* or Postnatal, Are Not Tenable

Procedure By Which The Six Criteria (ONE By One)Were Not Met:

THE INVESTIGATORS TYPICALLY:

1. Selected outdated and poorly normed measuring instruments over more recent ones with solid psychometric properties

Procedure By Which The Six Criteria Were Not Met:

2. Failed to establish the reliability and validity of the poorly chosen test

Procedure By Which The Six Criteria Were Not Met:

3. Performed a multitude of data analyses until one or more reached a level of statistical significance at $p \leq .05$

Procedure By Which The Six Criteria Were Not Met:

4. Interpreted differences within the standard error of a measuring instrument as if they were clinically meaningful

Procedure By Which The Six Criteria Were Not Met:

5. Failed to control adequately for confounding and potentially confounding variables , and

Procedure By Which The Six Criteria Were Not Met:

6. Treated longitudinal research investigations as a series of independent cross-sectional studies

The Logical Scientific Conclusion That Needs To Be Drawn Is:

The Numerous Claims and Conclusions Regarding the Putative Deleterious Effects of Maternal PCB Exposure Upon Subsequent Cognitive, Neuropsychological, and Behavioral Development in Their Offspring, Whether the Exposure Is *in utero* or Postnatal, are Scientifically Invalid.

However, A Very Important Caveat Needs To Be Made:

This Conclusion is Based Solely Upon the Low Levels of PCB Exposure That Have Been Reported.

It is Well Known (First Law of Toxicology), That:

Any Foreign Substance, Depending Upon Its Level, Will Be: Harmless, Slightly, Moderately, Highly Toxic, or Lethal (e.g. , the standards for lead exposure published by the CDC)

THEREFORE:

Given the Completely Unknown Level of PCB Toxicity That is Buried in The Silt of the Hudson River, Coupled With the Fact That PCB Levels Have Been Decreasing Over the Past 20 Year Period:

It Would Seem Foolhardy To Dredge Them Up, at This Point, Thereby Increasing the Probability of Placing Them in the Ambient Environment at Potentially Harmful Levels, Relative To the Harmless Levels at Which They Have Been Investigated and Reported, When Analyzed Objectively and Scientifically.

Broader Implications of This Research:

The Six Criteria Have Value Far Beyond the Evaluation of PCB Research.

They Have Relevance For Investigating Threats to the Reliability and Validity of Reported Research Results Across the Wider Field of Biobehavioral Research.