Dose-Response Relationship of Prenatal Mercury Exposure and IQ: An Integrative Analysis of Epidemiologic Data
SUPPLEMENTAL MATERIAL

Daniel A. Axelrad[1], David C. Bellinger[2], Louise M. Ryan[3], Tracey J. Woodruff[4]

[1]U.S. Environmental Protection Agency, Office of Policy, Economics and Innovation, Washington, DC, USA

[2] Departments of Neurology, Harvard Medical School and Children's Hospital Boston, and Department of Environmental Health, Harvard School of Public Health, Boston, MA, USA

[3]Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA

[4]U.S. Environmental Protection Agency, Office of Policy, Economics and Innovation, San Francisco, California, USA (on sabbatical to the University of California, San Francisco)

Supplemental Material Contents

Neurological Tests Conducted in the Faroe Islands, New Zealand and Seychelles Studies

A variety of tests of neurological development were administered to the children in the Faroe Islands, New Zealand and Seychelles studies. Each study administered a different set of tests, though there was some overlap of tests across studies. The tests administered are shown in Table A1.

Rescaling the Scores of the Neurological Tests

Each of the tests of neurological development administered to the children in the Faroe Islands, New Zealand and Seychelles studies has a different scoring system. For example, intelligence quotient (IQ) is defined to have an expected mean score of 100 and expected standard deviation of 15. In contrast, the Boston Naming Test has an expected mean score of 43 with a standard deviation of 5, and the California Verbal Learning Test has a mean score of zero and standard deviation of 1.

In order to consider each of these tests in a single statistical model informative of the relationship between prenatal mercury exposures and IQ, it is necessary for scores on all tests to be expressed in the same terms. Thus we rescaled the estimated regression coefficients and standard errors for tests included in the statistical model so that they correspond to test scores with the same distribution as full-scale IQ (that is, a standard deviation of 15).

This rescaling allows all inputs and outputs of our model to be expressed in terms of the decrement in IQ associated with a one unit increase in mercury body burden. Since the analysis was done using linear models, this is easily achieved with a simple linear rescaling of the estimated regression coefficients and standard deviations. To see this, consider a simple linear model, $Y = \beta_0 + \beta_1 X + e$, where Y is the outcome of interest, X is the covariate of interest and e is an error term. The regression coefficient β_1 corresponds to the expected change in the outcome Y associated with a one unit change in the covariate X (in our case, ppm of mercury in hair). Suppose the outcome Y has a standard deviation of σ, but we would like to compare the results with a regression based on a different outcome that has a standard deviation τ. We can do so by rescaling Y to $Y^* = \tau Y/\sigma$. It follows that the regression equation for Y* will be $Y^* = \beta_0^* + \beta_1^* X + e^*$, where $\beta_0^* = \tau \beta_0/\sigma$, $\beta_1^* = \tau \beta_1/\sigma$ and e* is the rescaled error term. Thus, we can convert coefficients obtained from regressions done in one scale to the desired scale by multiplying them by the factor τ/σ. In our context, we have used $\tau = 15$ so that all regression coefficients are comparable with analyses done for full-scale IQ.

Actual test results for relatively small study populations will typically have some differences from the "expected" distributions. In our rescaling, the value used for σ is the actual standard deviation on each test score reported for the three mercury epidemiologic studies. In New Zealand, for example, observed standard deviations tended to be somewhat larger than the established population standard deviations for the tests being

used (see Tables 10 and 11 of Kjellstrom et al. 1989), reflecting the very heterogeneous population in that study. In the Seychelles, the observed standard deviations tended to be smaller than established population standard deviations (Myers et al. 2003).

Table A2 shows the derivation of the scaling factors for each of the regression coefficients used in the analysis. The numerator in the calculation of each scaling factor is 15, as this is the standard deviation for IQ. For the New Zealand and Seychelles studies, the denominator is the response standard deviation for each test – i.e. the standard deviation of the tests scores reported for each cohort. For the Faroe Islands study, the denominator consists of the observed standard deviation and two additional factors. First, an additional factor of 10 is incorporated because regression coefficients for the Faroe Islands study (Budtz-Jorgensen et al. 2005) are reported for 10 ppb changes in cord blood mercury. Second, there is a factor of 0.2 to account for the conversion of mercury in cord blood (in ppb) to mercury in hair (in ppm) (Budtz-Jorgensen et al. 2004).

Sample code

The analysis was conducted with the statistical package WinBUGS. Sample model code is shown in Table A3.

References

Budtz-Jorgensen E, Debes F, Weihe P, Grandjean P. 2005. Adverse mercury effects in 7 year old children expressed as loss in "IQ." Report to the U.S. Environmental Protection Agency. Document ID EPA-HQ-OAR-2002-0056-6046. Available: http://www.regulations.gov [accessed 20 January 2006].

Budtz-Jorgensen E, Grandjean P, Jorgensen PJ, Weihe P, Keiding N. 2004. Association between mercury concentrations in blood and hair in methylmercury-exposed subjects at different ages. Environ Res 95:385-393.

Kjellstrom T, Kennedy P, Wallis P, Mantell C. 1989. Physical and mental development of children with prenatal exposure to mercury from fish. Stage 2: Interviews and psychological tests at age 6. Solna, Sweden. National Swedish Environmental Protection Board.

Myers GJ, Davidson PW, Cox C, Shamlaye CF, Palumbo D, Cernichiari E, et al. 2003. Prenatal methylmercury exposure from ocean fish consumption in the Seychelles child development study. Lancet 361:1686-1692.

Table A1. Neurobehavioral tests administered in the Faroe Islands, New Zealand and Seychelles studies.

Test	Primary Domain Assessed
Faroe Islands Study (age 7 years)	
Wechsler Intelligence Scale for Children-Revised (WISC-R) (selected subtests)	
Digit Span	Short-term memory
Similarities	Abstract verbal reasoning
Block Design	Constructional praxis
Bender-Gestalt Test	Visual-motor integration
California Verbal Learning Test-Children	Verbal learning and memory
Boston Naming Test	Confrontational naming
Tactual Performance Test	Nonverbal memory
Neurobehavioral Evaluation System (selected tests)	
Finger tapping	Motor speed
Hand-eye coordination	Hand-eye coordination
Continuous performance test	Vigilance
Profile of Mood States	Mood
Child Behavior Checklist (selected items)	Behavior disorders
New Zealand Study (age 6 years)	
Wechsler Intelligence Scale for Children - Revised Edition (WISC-R)	General intelligence
McCarthy Scales of Children's Abilities (MCC)	General development
Test of Language Development (TOLD)	General verbal skills
Peabody Picture Vocabulary Test	Receptive language
Clay Reading Diagnostic Survey	Reading
Burt Word Recognition Test	Single word reading
Key Math Diagnostic Arithmetic Test	General math skills
Everts Behaviour Rating Scale	Behavioral disorders

Test	Primary Domain Assessed
Seychelles Study (age 9 years)	
Wechsler Intelligence Scale for Children - Third Edition (WISC-III)	General intelligence
California Verbal Learning Test-Children (CVLT)	Verbal learning and memory
Boston Naming Test (BNT)	Confrontational naming
Finger tapping	Motor speed
Continuous Performance Test	Vigilance
Developmental Test of Visual-Motor Integration (VMI)	Visual-motor integration
Bruininks-Oseretsky Test of Motor Proficiency	Gross and fine motor skills
Grooved Pegboard	Manual dexterity
Trail-Making Test	Visual tracking and executive function
Woodcock-Johnson Tests of Achievement (selected subtests)	
Letter-Word Identification	Single word reading
Applied Math	Quantitative problem-solving
Wide Range Assessment of Memory and Learning (WRAML)	
Design Memory subtest	Visual memory
Haptic Discrimination Test	Cross-modal integration
Child Behavior Checklist	Behavioral disorders
Conners Hyperactivity Index	Behavioral disorders

Table A2. Scaling factors used to convert estimated regression coefficients and associated standard errors to the same scale as full-scale IQ and hair mercury.

Study	Endpoint	Scaling Factor
Faroe Islands[a]	Full-scale IQ[b]	15 / (1.45*0.2*10) = 5.17
	Bender[c]	-15 / (5.29*0.2*10) = -1.42
	BNT	15 / (5.48*0.2*10) = 1.37
	CVLT	15 / (2.58*0.2*10) = 2.91
	Full-scale IQ, alternate estimate[b]	15 / (0.586*0.2*10) = 12.8
New Zealand[d]	Full-scale IQ	15 / 16 = 0.94
	Performance IQ	15 / 16 = 0.94
	TOLD	15 / 16 = 0.94
	MCC	15 / 10 = 1.5
Seychelles[e]	Full-scale IQ	15 / 11.6 = 1.29
	CVLT	15 / 1.04 = 14.42
	BNT	15 / 4.8 = 3.13
	WRAML	15 / 2.9 = 5.17
	VMI	15 / 11.7 = 1.28

BNT, Boston Naming Test; CVLT, California Verbal Learning Test; IQ, intelligence quotient; MCC, McCarthy Scales of Children's Abilities; TOLD, Test of Language Development; VMI, Developmental Test of Visual-Motor Integration; WRAML, Wide Range Assessment of Memory and Learning.

[a]Data from Table 2 of Budtz-Jorgensen et al. (2005).
[b]For the Faroe Islands study, full-scale IQ is derived from a Structural Equation Model that combines the three WISC-R subscales (Digit Span, Similarities, and Block Design). For the primary estimate, the scaling factor is derived using the standard deviation (1.45) of Digit Span, because the since the latent variable of SEM is assumed to be on the same scale as Digit Span. The alternate estimate is derived using the standard deviation (0.586) of the latent variable itself, obtained as part of the SEM fitting procedure.
[c]The scaling factor for Bender-Gestalt is negative because higher scores on this test represent poorer performance.
[d]Standard deviations from Tables 10 and 11 of New Zealand report (Kjellstrom et al. 1989).
[e]Data from Tables 2 and 3 of Myers et al. (2003).

Table A3. Sample WinBUGS Code.

```
model{
      for (i in 1:nrow)  {
        # create the scaling variable and scale the endpoint specific
dose
        # response estimates and standard errors
         scale[i] <-  scale1[i]/scale2[i]
         y[i] <- b[i]*scale[i]
         p.y[i] <- 1/(scale[i]*scale[i]*pow(bse[i],2))

             # specify model for observed data
         y[i]  ~ dnorm(mu[i], p.y[i])
         mu[i] <-  beta1[study[i]] + beta2[endpoint[i]]     }

        for (i in 1:nstudy) {   beta1[i] ~ dnorm(0,p.study)  }

        for (i in 1:nendpoint) { beta2[i] ~ dnorm(beta0,p.endpoint)  }

        # flat prior on overall mean
        beta0 ~ dnorm(0,.0001)

             # specify feasible range of values for variance component
             sigma.study  ~ dunif(0,.2)

             # specify R
        R <- 3.5

        # compute sigma.endpoint and corresponding precisions
             sigma.endpoint  <- sigma.study/sqrt(R)
        p.study <- 1 / pow(sigma.study,2)
        p.endpoint <- 1 / pow(sigma.endpoint,2)
        }

list(nrow=13,nstudy=3,nendpoint=9,

b=    c(-.53, -.54, -.60, -.53,  -.13, .013, -.012, -.021, -.010,
-.024, .073, -.190, -.058),

bse = c( .29,  .33,  .30,  .21,  .10, .010,  .046,  .029,  .12,
.011, .059,  .063,  .032),

scale1 = c(.94,.94,  .94, 1.5,   1.29, 14.42, 3.13,  5.17,  1.28,
10.34, -2.84,  2.74,  5.82),

study = c(1,1,1,1, 2,2,2,2,2, 3,3,3,3),
endpoint=c(1,2,9,3, 1,4,6,7,8, 1,5,6,4),
scale2 = c(1,1,1,1, 1,1,1,1,1, 2,2,2,2))


# Study codes
# 1 is New Zealand (4 endpoints)
# 2 is Seychelles (5 endpoints)
# 3 is Faroes (4 endpoints)


# Endpoint codes:
# 1 is fullscale IQ (all three studies)
# 2 is performance IQ (WISC_RP in NZ)
# 3 is MCC_PP (McCarthy perceptual performance)
# 4 is CVLT-short delay (Seychelles, Faroes)
# 5 is Bender Visual (Faroes)
# 6 is BNT total (Seychelles, Faroes)
# 7 is WRAML design (Syechelles)
```

```
# 8 is VMI (Seychelles)
# 9 is TOLD-SL (New Zealand)

# The vector "scale1" corresponds to 15 divided by the observed
standard deviation of the
# corresponding endpoint.
# The vector "scale2" converts the results from cordblood to hair
mehg.
```