



Syddansk Universitet

Structural equation models for meta-analysis in environmental risk assessment

Budtz-Jørgensen, Esben; Debes, Frodi; Weihe, Pál; Grandjean, Philippe

Published in:
Environmetrics

DOI:
[10.1002/env.1000](https://doi.org/10.1002/env.1000)

Publication date:
2010

Document Version
Final published version

[Link to publication](#)

Citation for polished version (APA):
Budtz-Jørgensen, E., Debes, F., Weihe, P., & Grandjean, P. (2010). Structural equation models for meta-analysis in environmental risk assessment. *Environmetrics*, 21(5), 510-527. DOI: 10.1002/env.1000

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 15. jan.. 2017

Published in final edited form as:

Environmetrics. 2010 August 1; 21(5): 510–527. doi:10.1002/env.1000.

Structural equation models for meta-analysis in environmental risk assessment

Esben Budtz-Jørgensen^{1,*,†}, Frodi Debes², Pal Weihe^{2,3}, and Philippe Grandjean^{3,4}

¹ Department of Biostatistics, University of Copenhagen, Øster Farimagsgade 5B, DK-1014 Copenhagen, Denmark ² Department of Occupational and Environmental Health, The Faroese Hospital System, Sigmundargota 5, FO-100 Tórshavn, Faroe Islands ³ Department of Environmental Medicine, University of Southern Denmark, Winslowparken 17, DK-5000 Odense, Denmark ⁴ Department of Environmental Health, Harvard School of Public Health, 401 Park Drive, Boston, MA 02215 USA

SUMMARY

The potential of structural equation models for combining information from different studies in environmental epidemiology is explored. For illustration we synthesize data from two birth cohorts assessing the effects of prenatal exposure to methylmercury on childhood cognitive performance. One cohort was the largest by far, but a smaller cohort included superior assessment of the PCB exposure which has been considered an important confounder when estimating the mercury effect. The data were analyzed by specification of a structural equation model for each cohort. Information was then pooled based on a joint likelihood function with key parameters constrained to be equal in the different models. Modeling assumptions were chosen to obtain a meaningful biological interpretation of the joint effect parameters. Measurement errors in mercury variables were taken into account by viewing observed variables as indicators of latent variables. Adjustments for measurement error were also included for confounder variables. In particular, this example illustrates how to properly utilize that one study provided superior information about a confounder. A final more advanced model pooled information across different outcomes to gain power and to avoid multiple testing problems. In this model, the mercury effect remained statistically significant, while the effect of PCB was less certain.

Keywords

Environmental epidemiology; Confounding; Latent variables; Measurement error; Missing data; Multiple endpoints

1. INTRODUCTION

Modern environmental epidemiology often examines relatively weak exposure effects which may be difficult to detect. As replication is crucial, meta-analysis becomes particularly important in this field. However, the full potential of such methods can only be achieved if the statistical analysis properly addresses the most important weaknesses associated with the data. For example, the true causative exposure is typically difficult to measure and instead

*Correspondence to: E Budtz-Jørgensen, Department of Biostatistics, University of Copenhagen, Øster Farimagsgade 5B, DK-1014 Copenhagen K, Denmark.

†ebj@biostat.ku.dk

only imprecise exposure variables are available. To avoid underestimation of the exposure effect, measurement error in exposure markers must be allowed for (Carroll *et al.*, 2006). In addition, the significance of exposure effects may be overstated if the multiple testing problems induced by multiple exposure variables and multiple outcomes are not addressed. Effects of potential confounders must also be taken into account, and even here the effects of imprecision may need consideration. Structural equation models have previously been shown to be an useful approach to analyze multivariate data from environmental epidemiology (Budtz-Jørgensen *et al.*, 2002; 2003). Accordingly, we now explore the possibilities for combining information from different studies in these models. For illustration we consider two data sets on the long-term effect of prenatal exposure to methylmercury.

Methylmercury (MeHg) is a common food contaminant that is considered an important toxic risk to brain development (U.S.EPA, 2001; JECFA, 2003). Epidemiological studies of developmental MeHg neurotoxicity have applied different designs, exposure indicators, covariates, and outcome variables. Meta-analysis of such data may be difficult. However, results from two prospective studies of Faroese birth cohorts (Grandjean *et al.*, 1997; Steuerwald *et al.*, 2000) have used similar designs and may therefore be useful to examine jointly with modern statistical tools. While results from the first Faroese birth cohort at age 7 years have been reported in detail already (Grandjean *et al.*, 1997; 2001), a second, smaller cohort has subsequently been examined at the same age, and we now report the results for the first time. In the present paper, we include adjustment for concomitant exposure to polychlorinated biphenyls (PCBs) and for the beneficial influence of maternal fish intake during pregnancy.

An analysis of the joint data from the two Faroese cohorts poses some common challenges to the statistical methods. The main aim is to estimate the effect of prenatal mercury exposure as precisely as possible. This can only be achieved by taking into account effects of possible confounders, measurement error in mercury markers, and multiple testing problems. The possibility of confounding from PCB exposure has been considered likely when interpreting the results from the first cohort. However, adjustments for this effect are complicated by the fact that information about the PCB exposure was obtained in only half of the children. In addition, this information was based on the PCB concentration in cord tissue which may be imprecise. Failure to allow for measurement error in confounders will lead to bias in the estimated effect of exposure (Budtz-Jørgensen *et al.*, 2007). The direction of the bias depends on the relationship between study variables, and in the current study, an error component in the PCB variable would lead to an overestimation of the adverse mercury effect. An important advantage of including Cohort 2 is that in this cohort PCB was measured in maternal serum which is considered a very precise exposure marker. However, this also complicates the joint analysis as we have to allow for different degrees of imprecision in the two cohorts. This paper illustrates how the problems outlined can be solved using structural equation modeling to achieve a more accurate estimate of the impact of prenatal MeHg neurotoxicity on neurobehavioral performance at school age.

2. THE DATA

The first cohort was assembled in the Faroe Islands during a 21-month period in 1986–1987 (Grandjean *et al.*, 1992; 1997). As indicators of intrauterine exposure to MeHg, we used the mercury concentrations in cord blood and in maternal hair at parturition. Exposure to PCBs was assessed by analyzing cord tissue for major PCB congeners, and 50 whole blood samples from the cord were also analyzed (Grandjean *et al.*, 2001). The frequency of maternal intake of fish was available from a questionnaire administered shortly after parturition (Grandjean *et al.*, 1997). Cohort members were first invited for detailed

examination at school age (7 years), when a total of 917 of eligible children (90.3%) participated. The second cohort was formed during a 12-month period in 1994–1995 and consisted of 182 singleton term births at the National Hospital in Tórshavn (Steuerwald *et al.*, 2000). About 64% of all births were included, incomplete sampling being mainly due to logistic problems in the busy ward. In addition, four children were excluded because they were born before the 36th week of gestation, and two children because they had congenital neurological disease. For mercury analysis, we obtained whole blood from the umbilical cord and a hair sample from the mother. In this cohort, we analyzed the first 2 cm segment of the hair as a more precise measure of the fetal MeHg exposure. Analysis of PCB congeners was carried out on maternal serum from the 34th week of pregnancy and on transition milk obtained at days 4 or 5 after parturition. The total PCB concentration was based on twice the sum of congeners CB-138, CB-153, and CB-180. After a logarithmic transformation the correlation between the total PCB concentration and the cord blood mercury concentration was 0.44 and 0.43 in Cohort 1 and 2 (Grandjean *et al.*, 1997; Steuerwald *et al.*, 2000).

Outcome measures were similar in the two cohorts. The neuropsychological tests had been chosen to include tasks that would be affected by the neuropathological abnormalities described in congenital methylmercury poisoning and the functional deficits seen in children with early-life exposure to other neurotoxicants. Paper-and-pencil tests were administered by a Faroese clinical psychologist who had translated the tests into Faroese and verified their feasibility through pilot testing of Faroese children. Three computer-assisted tests were given at a separate session using a computer with the recommended joy-stick. Each child was asked about familiarity with computer games, and the answer was scored as none, some and much. The tests have been described in detail previously (Grandjean *et al.*, 1997) and are therefore only summarized here. The Neurobehavioral Evaluation System (NES2) Finger Tapping Test, where the maximum number of taps in 15 s was recorded for the preferred hand, the non-preferred hand, and both. The NES2 Continuous Performance Test with animal silhouettes was used for a 4-minute duration, and the median reaction time during the last three minutes and number of missed stimuli were recorded. For this test, we disregarded Cohort 1 data from the 2nd year of examination, as the quality was not acceptable. We also used the Wechsler Intelligence Scale for Children-Revised (WISC-R) Digit Spans, forward condition only, and the Similarities and Block Designs subtests. As visuospatial measure, we used the error score from the Bender Visual Motor Gestalt Test in the first cohort, and then switched to the Stanford-Binet Copying test (Thorndike *et al.*, 1986), a feasible test that appears to be sensitive to developmental MeHg exposure (Grandjean *et al.*, 1999). The test included the California Verbal Learning Test (children), learning condition (total number of correct responses during five trials), short-term and delayed reproduction, and recognition. For Cohort 2 a slightly updated version with four more item was applied. In the Boston Naming Test, we recorded both the score without cues and the total score after cues. These tests were grouped into two major categories of motor tests and verbally-mediated tests for the purpose of the statistical analyses (Budtz-Jørgensen *et al.*, 2002). Written informed consent was obtained from all mothers, and the study design and methodology were approved by the Faroese Ethical Review Committee and the institutional review board in the US.

3. STRUCTURAL EQUATION MODELS

The conditional distribution of the *dependent variables* $Y_i = (Y_{i1}, \dots, Y_{ip})^t$ given the *covariates* $Z_i = (Z_{i1}, \dots, Z_{iq})^t$ of subject i ($i = 1, \dots, n$) is modeled after specification of a measurement model and a structural model (Sánchez *et al.*, 2005). In the measurement model, the dependent variables are assumed to be linearly related to a set of latent variables ($\eta_{i1}, \dots, \eta_{im}$) and random error

$$\begin{aligned}
Y_{i1} &= \nu_1 + \sum_{j=1}^m \lambda_{1j} \eta_{ij} + \sum_{j=1}^q \delta_{1j} Z_{ij} + \varepsilon_{i1} \\
&\vdots \\
Y_{ip} &= \nu_p + \sum_{j=1}^m \lambda_{pj} \eta_{ij} + \sum_{j=1}^q \delta_{pj} Z_{ij} + \varepsilon_{ip}
\end{aligned} \tag{1}$$

In matrix notation $Y_i = \nu + \Lambda \eta_i + \Delta Z_i + \varepsilon_i$, where $\nu_{p \times 1}$, $\Lambda_{p \times m}$ and $\Delta_{p \times q}$ are parameter matrices. Typically, the number of latent variables (m) is much smaller than the number of observed variables (p). Furthermore, each of the dependent variables is usually considered a measure only of one latent variable. In this case, only one of the so-called factor loadings (λ_{ij}) in each row will be different from 0. The parameters δ model direct effects of covariates on observed variables. Covariates are often assumed to affect Y_i only through latent variables and therefore most δ -parameters are usually fixed at zero. The measurement errors ($\varepsilon_{i1}, \dots, \varepsilon_{ip}$) follow a normal distribution with a mean of zero and a variance matrix of Ω .

The structural part of the model describes how the latent variables are related to each other and to the covariates

$$\begin{aligned}
\eta_{i1} &= \alpha_1 + \sum_{j=1}^m \beta_{1j} \eta_{ij} + \sum_{j=1}^q \gamma_{1j} Z_{ij} + \zeta_{i1} \\
&\vdots \\
\eta_{im} &= \alpha_m + \sum_{j=1}^m \beta_{mj} \eta_{ij} + \sum_{j=1}^q \gamma_{mj} Z_{ij} + \zeta_{im}
\end{aligned} \tag{2}$$

In matrix notation $\eta_i = \alpha + B \eta_i + \Gamma Z_i + \zeta_i$, where $\alpha_{p \times 1}$, $B_{p \times p}$, and $\Gamma_{p \times r}$ are parameter matrices. Thus, each of the latent variables may depend on the covariates and other latent variables. Effects of latent variables are described by B , while Γ gives the effects of the covariates. Here we shall restrict attention to so-called recursive models which do not allow feedback loops, i.e., that a variable is affected by itself (Bollen, 1989). In such models, B can be arranged as a lower triangular matrix. The residuals ($\zeta_{i1}, \dots, \zeta_{im}$) are assumed to be independent of the measurement errors ($\varepsilon_{i1}, \dots, \varepsilon_{ip}$), while following a normal distribution with a mean of zero and a variance matrix of Ψ .

3.1. Inference

The model parameters are given by $\theta = (\nu, \Lambda, \Delta, \Omega, \alpha, B, \Gamma, \Psi)$. The distribution of Y_i given Z_i is $N_p\{\mu(\theta) + \Pi(\theta)z_i, \Sigma(\theta)\}$, where $\mu(\theta) = \nu + \Lambda(I - B)^{-1}\alpha$, $\Pi(\theta) = \Lambda(I - B)^{-1}\Gamma + \Delta$, and $\Sigma(\theta) = \Lambda(I - B)^{-1}\Psi(I - B)^{-1}\Lambda^t + \Omega$. Therefore the likelihood function is

$$L(Y, Z, \theta) \propto \prod_{i=1}^n \exp\{-[y_i - \mu(\theta) - \Pi(\theta)z_i]^t \Sigma(\theta)^{-1} [y_i - \mu(\theta) - \Pi(\theta)z_i] / 2\} / \sqrt{|\Sigma(\theta)|} \tag{3}$$

where $Y = (Y_1, \dots, Y_n)$, $Z = (Z_1, \dots, Z_n)$ and $|\cdot|$ denotes the determinant. Maximum likelihood estimates can be obtained using standard numerical maximization techniques such as Newton-Raphson or Fisher scoring (Bollen, 1989). Standard errors may be estimated from the inverse of the Fisher information and the significance of individual parameters can be assessed using Wald or likelihood ratio testing.

When many variables are modeled, the number of subjects with information on all variables (complete cases) may be limited and an analysis based only on these subjects will be inefficient and sometimes biased. Information from subjects with missing values on some of the dependent variables may be included in the analysis by using the likelihood theory of Little and Rubin (2002). Under the assumption that values are missing at random, the contribution to the likelihood function of the i 'th subject is determined by integrating the i 'th factor in (3) with respect to the missing variables, y_i^{mis} , i.e.

$\int \exp\{-[y_i - \mu(\theta) - \Pi(\theta)z_i]^t \Sigma(\theta)^{-1} [y_i - \mu(\theta) - \Pi(\theta)z_i]/2\} / \sqrt{|\Sigma(\theta)|} dy_i^{mis}$. The resulting density is (conditionally) normal with mean and variance given by deletions of rows and columns in $\mu(\theta)$, $\Pi(\theta)$, and $\Sigma(\theta)$ corresponding to the positions of the missing variables.

4. ANALYSIS OF MULTIPLE-STUDY DATA

Here we consider an extension of the standard model to allow meta-analysis. Thus, we assume that a vector of covariates Z_i^g and a vector of outcomes Y_i^g have been collected for the i 'th subject ($i = 1, \dots, n_g$) in study g ($g = 1, \dots, G$). The joint data is modeled by specifying a structural equation model for each study. The parameters are then estimated by maximization of the joint likelihood function given by the product of the study specific functions

$$L(\theta) \propto \prod_{g=1}^G L_g(Y^g, Z^g, \theta^g) \quad (4)$$

where $Y^g = (Y_1^g, \dots, Y_{n_g}^g)$ and $Z^g = (Z_1^g, \dots, Z_{n_g}^g)$. Here the likelihood function of each study L_g are all on the form given by equation (3) with parameters θ^g . The joint set of parameters is given by $\theta = (\theta^1, \dots, \theta^G)$. The main purpose of the analysis is to combine information from different studies. This is achieved by constraining specific parameters to be the same in some of the studies. A combined inference is then obtained based on maximization of the likelihood function (4).

The model above provides a very flexible framework for analysis of multiple-study data. Thus, studies can be compared in a likelihood ratio test of the hypothesis of no difference between studies in specific parameters. Given that a parameter is assumed to have the same value in all studies, the importance of this parameter can then be assessed again using likelihood ratio testing. Similar models have been considered before under the name *multiple group analysis* (e.g. Bollen, 1989). However, these models were developed for data sets where the same set of variables had been collected in all groups. This requirement is not satisfied in the Faroese data, so the standard framework is here modified to allow for group differences in the set of available variables.

5. JOINT ESTIMATION OF THE MERCURY EFFECT

For each of the neurobehavioral outcomes, this section compares mercury effects in the two cohorts. The analysis is based on a structural model allowing for measurement error in the mercury markers. This model is developed in two steps. First we build a model for exposures variables and covariates. This is done by viewing mercury concentrations in blood and hair as manifestations of a latent variable representing the true exposure. This exposure is then allowed to depend on covariates, and differences in model parameters between cohorts are explored. In the second step, the latent exposure and the covariates are related to the outcomes.

5.1. Exposure model for mercury

In both cohorts mercury exposure was assessed from concentrations in cord blood and hair ($B-Hg$, $H-Hg$). For the hair marker, Cohort 2 provides concentrations in the 2 cm closest to the scalp. This marker is also available in Cohort 1, but here we preferred concentrations in the full length of the hair as this variable has fewer missing values and because this marker has been used in previous analysis of Cohort 1 data. In a so-called MIMIC (Multiple Indicators-Multiple Causes) model, exposure data was linked to the covariates and to a common latent variable representing the true exposure (Hg). Thus, after a log transformation, both concentrations were assumed to depend linearly on true exposure

$$\begin{aligned}\log(B-Hg) &= \log(Hg) + \varepsilon_B \\ \log(H-Hg) &= \gamma_2^g + \lambda_2^g \log(Hg) + \varepsilon_H\end{aligned}\quad (5)$$

In the measurement model (5), the factor loading (λ) and intercept (ν) of the cord blood concentration is fixed at 1 and 0 to define the scale and the zero point of the latent variable. Thus, the true exposure has the same scale as the cord blood concentration in the sense that a one unit increase in $\log(Hg)$ will (on average) lead to a one unit increase in $\log(B-Hg)$. The structural part of the model allowed the true exposure to depend on the covariates, i.e.,

$\log(Hg) = \alpha_e^g + \gamma_e^g Z^g + \zeta_{Hg}$. Here the intercept α_e^g gives the true exposure level (in the cord blood scale) when covariates are all zero and γ_e^g gives the effects of covariates. Both sets of parameters were allowed to depend on the cohort g ($g = \{1, 2\}$). Residual terms ε_B , ε_H and ζ_{Hg} were assumed to be independent and normally distributed with mean zero and variance parameters that were allowed to differ between cohorts.

Potential confounders of the exposure-outcome relationship were included as covariates (Z^g). Here we used the set variables identified by Grandjean *et al.* (1997) in a regression analysis of Cohort 1 data. For Cohort 2 we used the same variables except that a variable indicating whether or not the child was in day-care was not available. Recent statistical analyses have indicated a beneficial effect of maternal fish intake during pregnancy therefore we also included this variable (Budtz-Jørgensen *et al.*, 2007). Finally, we also included the number of maternal pilot whale dinners during pregnancy. We did not assume that this variable had an independent effect on the outcomes but previous analysis has shown that it is a strong predictor of mercury exposure. Thus, this covariate was not included for confounder adjustment, but to improve identification of the true exposure.

The relationship between mercury biomarkers in the two cohorts was then compared using likelihood ratio testing. First, we tested a hypothesis of invariance of the factor loadings, i.e. $\lambda_2^1 = \lambda_2^2$. This was accepted ($p = 0.39$) so the slope between log-transformed concentrations in blood and hair were consistent in the two cohorts. The joint estimate was $\hat{\lambda} = 0.887$ (CI: 0.822–0.951), indicating that the blood to hair concentration ratio increases for higher exposure levels, i.e., $B - Hg / H - Hg \approx \exp(-\gamma_2^g) Hg^{(1-\lambda)}$ (Budtz-Jørgensen *et al.*, 2004). Factor loading invariance is of key importance for the interpretation of the following analysis. Had the slopes differed, a one unit increase in the latent exposure would have had different consequences in observed concentrations in the two cohorts, and therefore it would have been difficult to compare the effects of the latent exposure variable across cohorts. It should be noted that the choice of reference indicator is not important in the test of factor loading invariance. Thus, the same results would have been obtained had we used the

maternal hair concentration as the reference and then tested if the cord blood loadings were the same.

There was stronger evidence for a difference between the intercepts, i.e., $\widehat{\nu}_2^2 - \widehat{\nu}_2^1 = 0.071$ with $p = 0.059$. This means that for the same cord blood concentration the hair concentration level in Cohort 2 is 7.4% [$\exp(0.071) - 1$] higher. Although, this may be due to the shorter hair strains considered in Cohort 2, this difference was unexpected and since it was not formally statistically significant it was excluded from the models in the following analysis.

In agreement with the results of Budtz-Jørgensen *et al.* (2002; 2004) for Cohort 1, the measurement error term was less important for the cord blood concentration. Thus, in Cohort 1 the correlation to the true exposure was 0.913 and 0.857 for log-transformed mercury concentration in cord blood and hair, respectively. In Cohort 2 the corresponding numbers were 0.951 and 0.885. These coefficients depend on the variance of the true exposure and may therefore be difficult to compare between cohorts. Alternatively, the imprecision may be expressed using the standard deviation of the error term. Because of the logarithmic transformation these may be interpreted as (error) coefficients of variation on the original concentration scale. For the cord blood concentrations, the error standard deviations were 0.349 and 0.252 in Cohort 1 and 2 respectively, while the numbers for the hair concentration were 0.418 and 0.364. Thus, precisions were better in Cohort 2, but the differences were not statistically significant ($p = 7.33\%$ for blood, $p = 15.7\%$ for hair). For the hair concentration this was to be expected because Cohort 2 used shorter hair which are less likely to be affected by external contaminations.

5.2. Outcome model

Outcome variables were included one at a time through an additional regression equation $outcome = \alpha_o^s + \beta^s \log(Hg) + \gamma_o^s Z^s + \zeta_o$ (Figure 1). So the outcome in question was assumed to depend linearly on the covariates and the log-transformed true mercury exposure. Potential confounders of the exposure-outcome relationship were included as covariates (Z^s). As in the exposure part of the model, covariates were allowed to have different effects on the outcome in the two cohorts, i.e. $\gamma_o^1 \neq \gamma_o^2$. The parameter of main interest β^s represents the effect of a 10-fold increase in the true mercury exposure (in each cohort). Thus, measurement error in observed mercury concentrations was taken into account by estimating the effect of a common latent variable. As a further advantage of this analysis, the effects of the two exposure variables were modeled using only one effect parameter thereby reducing the multiple testing problem. We first tested whether mercury effects were the same in the cohorts, then a joint effect estimate was obtained by re-fitting the model under the constraint $\beta^1 = \beta^2$. This hypothesis is meaningful because the covariates used in the two studies were similar.

Generally, mercury effect estimates in the two cohorts were in good agreement (Table 1). Thus, except for one outcome (finger tapping with preferred hand) the test for equality of effects was accepted for all outcomes. A very close agreement between the cohorts was seen for the scores on the Boston Naming Test, which were the outcomes with the most significant effects in Cohort 1. For finger tapping with preferred hand an un-expected strong positive effect of mercury was observed in Cohort 2. This effect was significantly different from that of Cohort 1. As expected, the joint estimates were between the two cohort specific effects, but closer to the Cohort 1 value as a result of the larger sample size. Interestingly, all joint coefficients were in the anticipated direction and 8 of 14 were statistically significant at the two-tailed 5% level.

6. PCB ADJUSTMENT

This analysis was complicated by the fact the two cohorts contained different information about the prenatal exposure to PCB. In Cohort 2, the PCB concentration in maternal pregnancy serum (*S-PCB*) and transition milk was obtained in almost all children. The serum concentration is generally considered to provide a very precise measure of the true exposure to the fetus. Therefore, we shall view this variable as the gold standard. This marker was not available in Cohort 1. Instead for about half of the children PCB was measured in cord tissue (*Ct-PCB*). Previous regression analysis of the effect of PCB in Cohort 1 have included this variable as an additional covariate thereby greatly reducing the number of complete cases (Grandjean *et al.* 1997; 2001). Furthermore, the cord tissue measurement is technically demanding and this biomarker may be imprecise compared to the serum concentration. As a further complication, cord tissue concentrations were obtained by two different laboratories. Information about the degree of imprecision in the cord tissue concentration is available from 50 children in Cohort 1, where PCB measurements were made in both cord tissue and cord blood (*Cb-PCB*). Because lipid-based PCB concentrations in cord blood and serum are expected to be almost identical we will use information from these children to estimate the relationship between *S-PCB* and *Ct-PCB*.

Figure 2 shows the relationship between PCB concentrations in cord blood and cord tissue as measured in two laboratories. For each laboratory a linear relationship on the log scale is observed. The residual variance is seen to decrease somewhat with increasing concentrations. This tendency is not easily modeled using structural equations and is ignored in the following analysis. First we estimated the relationship between *Ct-PCB* and *Cb-PCB* for each laboratory (*l*) in a joint regression model

$$\log(Ct-PCB) = \alpha_l + \beta_l \log(Cb-PCB) + \varepsilon \quad (6)$$

where $\varepsilon \sim N(0, \sigma_l^2)$ and $l \in (1, 2)$. This analysis showed that measurement error variances in the two laboratories could be assumed to be identical ($\sigma_1^2 = \sigma_2^2$) and so could the slope in blood-tissue relationship ($\beta_1 = \beta_2$). However, a significant difference in the intercepts (α_l) was seen indicating that the *Ct-PCB* concentration levels were about 37% higher in Lab 1. As will be explained, these findings were taken into account in the structural equation analysis.

Figure 3 illustrates how information about the prenatal exposure to PCB was incorporated into the model. The main idea is that concentrations in serum and cord blood can be viewed as very precise markers of the child's true long term exposure which is modeled by the latent variable $\log(PCB)$. In the initial model, we therefore fixed the measurement error terms ε_S and ε_{Cb} at zero. Thus, in Cohort 2 almost all children have an error free PCB exposure variable. This is only true for 50 children in Cohort 1, but here exposure information is also provided by the cord tissue measurement. Using a measurement model similar to equation (5), this variable was included as an additional indicator with a measurement error variance to be estimated in the analysis. In accordance with the regression analysis above, the level of this marker was allowed to depend on the analytical laboratory. So this covariate is expected to affect a particular observed indicator directly and not through the latent variable. In the psychometrics literature this type of an effect is sometimes described as *item bias* or *differential item function*. Latent exposures to PCB and mercury were allowed to be correlated and, together with covariates, they were assumed to affect the outcome. Information from the two cohorts were pooled by constraining the effect parameters of environmental exposures to be equal in the two data subsets. Finally, it should

be noted that, in the Cohort 2 model, the milk concentration was excluded because this variable will not provide additional exposure in children who also have an *S-PCB* measurement.

6.1. Estimation and results

When fitting the proposed model to the Faroese data, complete case estimation methods cannot be used. In Cohort 1 very few children have a PCB concentration in cord blood and approximately half of the children are missing the cord tissue concentration. To include observations from in-complete cases we used the maximum likelihood technique outlined in Section 3.1. Thus, the likelihood function was determined by integrating the complete case density over the missing exposure concentrations. This approach leads to an efficient effect estimation as all available exposure information is utilized in a joint analysis.

After adjustment for PCB the mercury effect estimates were generally weaker and less significant (Table 2). However, all coefficients were in the direction of an adverse effect, and five mercury coefficients remained statistically significant. The PCB effect was less clear. Here the sign of the coefficients varied, and only the Digit Span coefficient had a *p*-value below 5%. However, this PCB effect was positive which may be due to the collinearity introduced by having both mercury and PCB in the model. A stronger analysis which will have a better chance of separating the effects of mercury and PCB can be obtained by borrowing information from different outcomes. In the next section, we describe a structural equation model for multiple outcome variables.

7. JOINT ANALYSIS OF MULTIPLE OUTCOMES

In this section different outcome variables are viewed as manifestations of underlying latent neurobehavioral functions. We shall extend the model developed by Budtz-Jørgensen *et al.* (2002) for estimation of mercury effects on the child's motor and verbal function. In that model, scores on the Boston Naming Test, CVLT scores and WISC(R) Digit Spans were considered indicators of a common verbally mediated function while the NES2 Finger Tapping scores and NES2 Hand-Eye Coordination (HEC) were assumed to measure motor function. Of these scores, all except the HEC were available in both cohorts. We therefore developed a multiple group version of this model with the exclusion of the HEC.

In the measurement part of the model, the three finger tapping scores were viewed as reflections of a latent motor function (*motor*)

$$FT1 = motor + \varepsilon_{f1}$$

$$FT2 = \nu_{f2}^g + \lambda_{f2} motor + \varepsilon_{f2}$$

$$FT3 = \nu_{f3}^g + \lambda_{f3} motor + \varepsilon_{f3} \tag{7}$$

Thus, finger tapping with preferred hand (FT1) was chosen as the reference indicator for the latent motor function. A similar set of equations was included to view the rest of the test scores mentioned above as indicators of a latent verbally mediated function (*verbal*). Here, the Boston Naming score with cues was used as the reference. Error terms (ε) were generally

assumed to be independent, meaning that the observed test scores must be uncorrelated given the level of the latent neurobehavioral function. This conditional independence is not likely to hold for the two Boston Naming scores and for the four scores on the CVLT. These scores are sub-scores of the same overall test and are therefore expected to be correlated beyond the degree explained by the latent variable. This type of association is known as local dependence (Budtz-Jørgensen *et al.*, 2002). Here we modeled local dependence by allowing the error terms of subscores on the same test to be correlated. Thus, the appropriate off-diagonal elements in the error covariance matrix Ω were allowed to be non-zero.

Observed exposure markers were included using the measurement model of the previous section. In the structural part, effects of environmental pollutants were modeled by allowing the latent outcome functions to depend on the latent exposure to mercury and PCB in addition to the covariates (Figure 4)

$$motor = \alpha_m^g + \beta_{m,hg} \log(Hg) + \beta_{m,PCB} \log(PCB) + \gamma_m^g Z^g + \zeta_m$$

$$verbal = \alpha_v^g + \beta_{v,hg} \log(Hg) + \beta_{v,PCB} \log(PCB) + \gamma_v^g Z^g + \zeta_v \quad (8)$$

Thus, again information were combined across the cohorts by assuming that the effects of mercury and PCB did not depend on the cohort. No such restriction was put on the effects of the covariates. Residual variation in the latent outcome functions were modeled by the variables ζ_m and ζ_v , which were assumed to follow normal distributions with zero mean. These terms were allowed to be correlated as verbal and motor function were expected to be positively associated (given the predictors considered).

7.1. Estimation and results

In equation (7) factor loadings are assumed not to depend on the cohort. As was the case for factor loadings in the exposure part of the model, this assumption is essential for the interpretation of the following analysis. Had this not been satisfied, the slopes reflecting the relation between two outcomes would have been different in the two cohorts. As a consequence, a one unit increase in the underlying latent function would not have had the same effect on the observed scores in the two cohorts. So even under the restriction that exposure effect parameters (β) are the same, the exposure would affect observed outcomes differently in the two cohorts. Therefore it does not make sense to compare effects of exposures unless factor loadings are the same in the two groups. As a check of the model, we compared it to a more flexible model allowing for cohort dependent factor loadings. The hypothesis of invariant factor loadings was easily accepted ($\chi_8^2 = 6.56, p = 0.58$).

In the measurement model with invariant factor loadings, we went on to test whether the intercepts were identical ($v^1 = v^2$). This was clearly rejected ($\chi_8^2 = 132, p < 0.0001$) indicating that children from different cohorts with the same (expected) score on one test may score systematically different on another test. Thus, for fixed values of the two reference indicators (NES2 Finger tapping with preferred hand and Boston Naming cued) children in Cohort 2 did better on the CVLT-scores. This advantage may have been a result of the fact that more items were included to the CVLT for this cohort, thereby allowing children to have higher tests scores. On the other tests, Cohort 1 children were superior. As a consequence of the differences between v^1 and v^2 , we cannot compare the level of latent outcome functions between cohorts. The difference between the intercepts of the latent

variables ($\alpha_m^1 - \alpha_m^2$) will depend on how well each cohort performed on the score chosen as the reference indicator. While some of these differences were highly statistically significant (e.g. the Cohort 1 advantage on finger tapping with non-preferred hand and Cohort 2 advantage on CVLT Recognition) others were far from being so. However, we decided not to do model selection for these parameters because the main aim of the analysis was to estimate exposure-related effects which can be achieved in the full model.

Table 3 provides estimated effects of prenatal exposure to mercury and PCB on a child's verbal and motor function. In this joint analysis all four effect parameters are negative indicating that, due to increased power, collinearity is less of a problem in the structural equation model. For mercury, the verbal effect is statistically significant while the motor effect is close to being so. The improved power of the joint analysis is illustrated by the fact that the overall verbal effect is more significant than each of the individual effects of Table 2. On the other hand, mercury effects are somewhat attenuated when compared to the structural equation results of Budtz-Jørgensen *et al.* (2007), which did not consider the PCB effect. The effect of prenatal exposure to PCB is less certain. Both coefficients are negative but especially the motor effect is far from reaching statistical significance.

The interpretation of the effect parameters of Table 3 depends on the choice of reference indicators. Thus, for mercury it was estimated that a 10-fold increase in the true cord blood concentration induces a verbal deficit corresponding to a loss of 1.52 points on the Boston Naming Test (cues). This effect is in nice agreement with the regression coefficient of 1.58 obtained when analyzing this outcome separately (Table 2). For the other outcomes a similar comparison can be conducted by expressing the mercury effect of the multiple response model in different outcome scales. This can be achieved by repeating the estimation procedure with different choices of reference indicator. However, under the model assumptions, the ratio of mercury effects in two different outcomes is equal to the corresponding ratio of factor loadings. Thus, the expected effect in a given outcome can be obtained by multiplying the effect of Table 3 with the estimated factor loading for that outcome. Table 4 compares the estimated mercury effects of the multivariate response model to those obtained in separate models for each outcome. A nice agreement is seen between the two sets of coefficients, indicating that the structure imposed by the multiple outcome model gives a good description of the effects of mercury in different outcomes. The increased power of the multivariate outcome model is illustrated by the fact that this model yields effect estimators with decreased standard errors.

The comparisons conducted of the health effects of prenatal exposure to mercury and PCB all depend on the assumption of no measurement error in PCB concentrations in cord blood and serum. To study the consequences of this assumption, we included a sensitivity analysis assuming different degrees of error in these concentrations. For both outcome functions the adverse mercury effect increased, while the PCB effect decreased when a higher degree of uncertainty was assumed for the cord blood/serum marker. The observed relationship between PCB in cord blood and cord tissue (Figure 2) limits the possible values for the error variance in cord tissue (Fuller, 1987). Thus, a higher degree of assumed uncertainty in cord blood will lead to a lower estimated error variance in cord tissue. The maximal error standard deviation in agreement with the data was 0.309. At this level of imprecision, the adverse mercury effect on verbal and motor function were -1.62 ($p = 0.014$) and -1.38 ($p = 0.036$). Although, the effect on motor function is now significant, these coefficients do not differ by an important extent from the results in Table 3; so our results seem quite robust to the error assumptions for PCB in serum/cord blood. Interestingly, the maximal error standard deviation of the structural equation model is supported by information from Cohort 2 on the PCB concentration in maternal milk. If this concentration is assumed to be error

free and used as a covariate to predict the cord serum concentration then the residuals had a standard deviation of error of 0.31.

8. DISCUSSION

The observational studies carried out so far to explore the low-level neurotoxicity of methylmercury offer strong, although somewhat uneven evidence of the public health impact of exposures from contaminated seafood (Grandjean *et al.*, 2005). A recent attempt to combine findings from three prospective studies chose to rely on only three of twenty neuropsychological outcomes from the Faroes Cohort 1 that overlapped with tests used in other studies (Axelrad *et al.*, 2007). Thus, meta-analysis may be hampered, when studies differ in outcome variables. Differences in the choice of confounders could also be of crucial importance for the apparently conflicting findings. While the results from one study were routinely adjusted for postnatal MeHg exposure (Myers *et al.*, 2003), most of the data available is unadjusted for the beneficial effects of seafood nutrients, thereby underestimating the true MeHg toxicity (Budtz-Jørgensen *et al.*, 2007). The present study overcame these problems by combining two studies that applied almost the same outcome tests, administered by the same neuropsychologist in the same language. While the statistical analysis allowed for a beneficial effect of maternal fish consumption, confounder adjustment efforts focused mainly on prenatal PCB exposure. The Faroese population is exposed to increased amounts of both MeHg and PCBs. Although developmental PCB exposure is associated with neurobehavioral deficits (Jacobson and Jacobson, 1996; Vreugdenhil *et al.*, 2000), previous analyses of the Faroes data failed to show any important impact of PCB exposure on the neurotoxicity outcomes (Grandjean *et al.*, 1997). The present study extends this information, because more children and more information on PCB exposure was included in the structural equation model. Inclusion of PCB exposure as a confounder attenuated the joint estimate of the mercury effect somewhat, but mercury remained statistically significant in a multivariate analysis. The effect of exposure to PCB was weaker and less certain.

Structural equation models are being used more frequently in epidemiology, and this study shows that they may also be successfully applied for meta-analysis. Here a structural equation model was first specified for each cohort. Then information from the cohorts were combined based on the joint likelihood function with key effect parameters constrained to be equal in the two sets of data. In addition, this structure was exploited to achieve a more sophisticated adjustment for prenatal PCB exposure which has been considered an important confounder in the Faroese data. Thus, the model took into account that different children had different exposure information. Most children had no or only a potentially imprecise marker value. Superior PCB exposure information was available for a small subset of Cohort 1 and in the smaller Cohort 2. The structural equation analysis estimated the imprecision in the Cohort 1 exposure variable and took this into account when combining information from the two cohorts to achieve a joint estimate of the PCB effect. In contrast, previous standard regression analysis have ignored the possibility of measurement error in the PCB concentrations. Not only may this lead to an underestimation of the PCB effect but it may also have resulted in a biased estimate of the health effects of mercury. If a confounder variable has measurement error, then a standard regression analysis will not provide full adjustment for its effect (Carroll *et al.*, 2006; Budtz-Jørgensen *et al.*, 2007). In the current study, this would imply that a part of an adverse PCB effect would be assigned to mercury. Thus, an unbiased analysis can only be obtained by allowing for measurement error in both the exposure and the confounder. Along with the improved adjustment for imprecision and confounding, the structural equation models also allowed information from different outcomes to be utilized in a joint model. In doing so power was gained and the multiple comparison problem of standard regression analysis was avoided.

Hierarchical regression models have been shown to be useful when combining dose-response information from different studies (Dominici *et al.*, 2000). In fact, Bayesian hierarchical models have recently been used to combine data from epidemiological studies carried out in the Seychelles, New Zealand and the Faroese Islands (Cohort 1) to assess the health effects of mercury exposure (Axelrad *et al.*, 2007; Coull *et al.*, 2003; Ryan, 2008). This analysis used standardized mercury regression coefficients for the different outcomes considered. The variability of these coefficients were studied in a random effects model allowing for study-to-study and endpoint-to-endpoint variation. Compared to this approach, the structural equation analysis presented here have some important advantages. First of all, individual outcome data were available here, which meant that the correlation between outcomes obtained in the same child could be taken properly into account. Perhaps more importantly, measurement error in study variables were handled more appropriately in the structural equation model. This analysis allowed for measurement error in exposure and confounder variables, which is essential for unbiased effect estimation. However, measurement errors in the outcome variables were also handled differently. Such errors were taken into account in the measurement part of the structural equation analysis by assuming that each outcome was a sum of a term depending linearly on the latent variable and random error. Instead the hierarchical analysis was based on standardized coefficients to overcome the problem that different outcomes may have different scales. However, standardized coefficients will tend to be closer to zero for outcomes with a high degree of measurement error. The joint effect estimate of the hierarchical regression will therefore depend on the precision in the outcome variables included in the model. In the structural equation analysis, outcome imprecision will affect only the standard error of the estimator.

Acknowledgments

The study was supported by grants from the National Institute of Environmental Health Sciences (ES09797, ES11687), and the European Commission through its Sixth Framework Programme for RTD (contract no FOOD-CT-2006- 016253, PHIME). The paper reflects only the authors' views. Neither the NIH nor the Community is liable for any use that may be made of the information contained therein. The authors are grateful to J. Hilden and N. Keiding for helpful comments.

References

- Axelrad DA, Bellinger DC, Ryan LM, Woodruff TJ. Dose-response relationship of prenatal mercury exposure and IQ: An integrative analysis of epidemiological data. *Environmental Health Perspectives*. 2007; 115:609–615. [PubMed: 17450232]
- Bollen, KA. *Structural Equations with Latent Variables*. John Wiley & Sons; Chichester, UK: 1989.
- Budtz-Jørgensen E, Grandjean P, Jørgensen PJ, Weihe P, Keiding N. Association between mercury concentrations in blood and hair in methylmercury exposed subjects at different ages. *Environmental Research*. 2004; 95:385–393. [PubMed: 15220072]
- Budtz-Jørgensen E, Grandjean P, Weihe P. Separation of risks and benefits of seafood intake. *Environmental Health Perspectives*. 2007; 115:323–327. [PubMed: 17431478]
- Budtz-Jørgensen E, Keiding N, Grandjean P, Weihe P. Estimation of health effects of prenatal mercury exposure using structural equation models. *Environmental Health*. 2002; 1:2. [PubMed: 12513702]
- Budtz-Jørgensen E, Keiding N, Grandjean P, Weihe P, White RF. Statistical methods for the evaluation of health effects of prenatal mercury exposure. *Environmetrics*. 2003; 13:105–120.
- Carroll, RJ.; Ruppert, D.; Stefanski, LA.; Crainiceanu, C. *Measurement Error in Nonlinear Models: A Modern Perspective*. Chapman & Hall; New York: 2006.
- Coull BA, Mezzetti M, Ryan LM. A Bayesian hierarchical model for risk assessment of methylmercury. *Journal of Agricultural, Biological, and Environmental Statistics*. 2003; 8:253–270.
- Dominici F, Samet JM, Zeger SL. Combining evidence on air pollution and daily mortality from the 20 largest US cities: a hierarchical strategy. *Journal of the Royal Statistical Society, Series A: Statistics in Society*. 2000; 163:263–284.

- Fuller, WA. Measurement Error Models. Wiley; 1987.
- Grandjean, P.; Cordier, S.; Kjellstrøm, T.; Weihe, P.; Budtz-Jørgensen, E. Health effects and risk assessments. In: Pirrone, N.; Mahaffey, KR., editors. Dynamics of mercury pollution on regional and global scales: atmospheric processes and human exposures around the world. Springer; Norwell: 2005. p. 499-523.
- Grandjean P, Weihe P, Burse VW, Needham LL, Storr-Hansen E, Heinzow B, Debes F, Murata K, Simonsen H, Ellefsen P, Budtz-Jørgensen E, Keiding N, White RF. Neurobehavioral deficits associated with PCB in 7-year-old children prenatally exposed to seafood neurotoxins. *Neurotoxicology and Teratology*. 2001; 23:305–317. [PubMed: 11485834]
- Grandjean P, Weihe P, White RF, Debes F, Araki S, Murata K, Sørensen N, Dahl D, Yokoyama K, Jørgensen PJ. Cognitive deficit in 7-year-old children with prenatal exposure to methylmercury. *Neurotoxicology and Teratology*. 1997; 19:417–428. [PubMed: 9392777]
- Grandjean P, White RF, Nielsen A, Cleary D, de Oliveira Santos CE. Mercury neurotoxicity in Amazonian children downstream from gold mining. *Environmental Health Perspectives*. 1999; 107:587–591. [PubMed: 10379006]
- Grandjean P, Weihe P, Jørgensen PJ, Clarkson T, Cernichiari E, Viderø T. Impact of maternal seafood diet on fetal exposure to mercury, selenium, and lead. *Archives of Environmental Health*. 1992; 47:185–195. [PubMed: 1596101]
- Jacobson JL, Jacobson SW. Intellectual impairment in children exposed to polychlorinated biphenyls in utero. *New England Journal of Medicine*. 1996; 335:783–789. [PubMed: 8703183]
- JECFA (Joint FAO/WHO Expert Committee on Food Additives). Summary and conclusions; Sixty-first meeting; Rome. 10–19 June 2003; [accessed, 27 January, 2007]. URL: <ftp://ftp.fao.org/esn/jecfa/jecfa61sc.pdf>
- Little, RJA.; Rubin, DB. *Statistical Analysis With Missing Data*. 2. John Wiley & Sons; New Jersey: 2002.
- Myers GJ, Davidson PW, Cox C, Shamlaye CF, Palumbo D, Cernichiari E, Sloane-Reeves J, Wilding GE, Kost J, Huang LS, Clarkson TW. Prenatal methyl-mercury exposure from ocean fish consumption in the Seychelles child development study. *Lancet*. 2003; 36:1686–1692. [PubMed: 12767734]
- Ryan LM. Combining data from multiple sources, with applications to environmental risk assessment. *Statistics in Medicine*. 2008; 27:698–710. [PubMed: 18069727]
- Sánchez BN, Budtz-Jørgensen E, Ryan L, Hu H. Structural equation models. A review with applications to environmental epidemiology. *Journal of the American Statistical Association*. 2005; 100:1443–1455.
- Steuerwald U, Weihe P, Jørgensen PJ, Bjerve K, Brock J, Heinzow B, Budtz-Jørgensen E, Grandjean P. Maternal seafood diet, methylmercury exposure, and neonatal neurological function. *The Journal of Pediatrics*. 2000; 136:599–605. [PubMed: 10802490]
- Thorndike, RL.; Hagen, EP.; Sattler, JM. *Stanford-Binet Intelligence Scale*. 4. IL: Riverside, Chicago; 1986.
- Vreugdenhil H, Duivenvoorden HJ, Weisglas-Kuperus N. The relative importance of prenatal PCB exposure, feeding type, and parental characteristics for cognitive and motor development in healthy children studied 3 to 84 months of age. *Organohalogen Compounds*. 2000; 48:139–142.
- U.S. EPA (Environmental Protection Agency), Office of Science and Technology, Office of Water. Washington: 2001 [accessed, 27 January, 2007]. *Water Quality Criterion for the Protection of Human Health: Methylmercury*, Final. EPA-823-R-01-001. URL: <http://www.epa.gov/waterscience/criteria/methylmercury/document.html>



Figure 1.
Path diagram showing the model fitted in each cohort. Mercury concentrations in cord blood (*B-Hg*) and hair (*H-Hg*) are indicators of true exposure (*Hg*). The outcome may be affected by the true exposure and covariates.

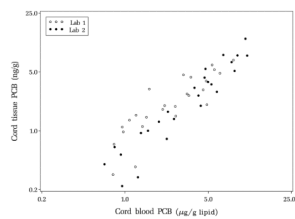


Figure 2. Relationship (on log scales) between PCB concentrations in cord blood and cord tissue measured by two different laboratories



Figure 3.

Path diagram illustrating a model with both prenatal exposure to mercury and PCB as predictors of the neurobehavioral outcome. In Cohort 1, PCB concentrations in cord blood (*Cb-PCB*) and cord tissue (*Ct-PCB*) were considered to be indicators of the true exposure PCB. The level of (*Ct-PCB*) was allowed to depend on the laboratory of measurement. In Cohort 2, only the PCB concentration in serum (*S-PCB*) was available. In the initial model, *S-PCB* and *Cb-PCB* were assumed not to have measurement error (i.e., $\varepsilon_S = \varepsilon_{Cb} = 0$). Latent exposures to PCB and mercury were allowed to be correlated and together with covariates they could affect the outcome. Effects of environmental exposures were assumed to be equal in the two cohorts.

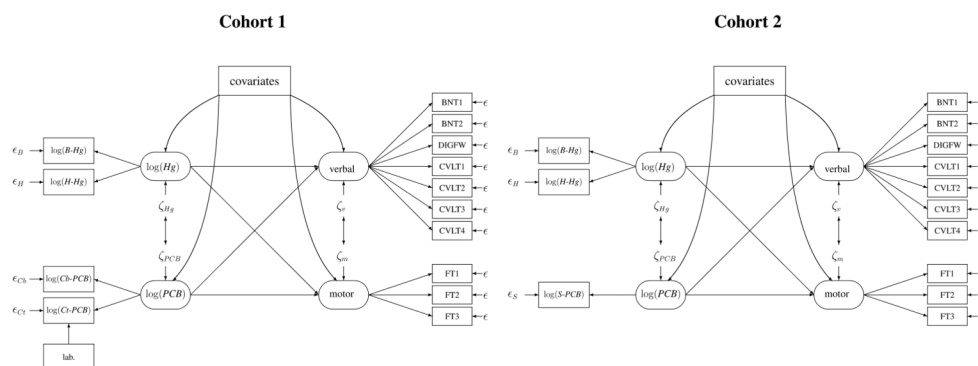


Figure 4. Path diagram illustrating a model for multiple neurobehavioral test-scores. Seven outcomes were considered manifestations of the child’s verbal functions while the three finger tapping scores reflected motor function. The latent outcome functions were allowed to depend on prenatal exposure to mercury and PCB. Effects of exposures were assumed to be the same in the two cohorts.

Table 1

Estimated effect of 10-fold increase in prenatal exposure to mercury. The results are shown separately for Cohort 1 and Cohort 2. A p-value is provided from the test that the two mercury effects are the same, and then a joint estimate is shown. The results are based on a latent mercury exposure variable and they have been adjusted for confounders other than PCB exposure. Number of children with outcome information was about 900 in Cohort 1 and 160 in Cohort 2.

Outcome	Cohort 1		Cohort 2		Cohort 1 & 2	
	β	95%-conf.	β	95%-conf.	β	95%-conf.
NES2 Finger tapping						
Preferred hand	-1.966	-3.228; -0.704	3.449	-0.226; 7.123	-1.424	-2.629; -0.219
Non preferred hand	-1.290	-2.510; -0.069	0.838	-2.108; 3.783	-0.985	-2.118; 0.148
Both hands	-2.975	-5.546; -0.404	-1.569	-8.188; 5.051	-2.795	-5.201; -0.389
NES2 CPT						
Reaction time	44.027	20.491; 67.563	7.680	-30.388; 45.748	34.080	13.832; 54.327
Ln total missed	0.139	0.027; 0.251	0.111	-0.066; 0.287	0.131	0.036; 0.225
Boston Naming Test						
No cues	-2.120	-3.239; -1.001	-2.020	-4.679; 0.639	-2.105	-3.137; -1.073
Cues	-2.223	-3.328; -1.119	-2.0780	-4.683; 0.527	-2.201	-3.219; -1.183
WISC(R)						
Similarities	-0.016	-0.908; 0.876	-0.969	-2.192; 0.253	-0.347	-1.071; 0.377
Sqrt. Block Designs	-0.234	-0.469; 0.002	0.097	-0.512; 0.706	-0.191	-0.411; 0.029
Digit Spans	-0.228	-0.539; 0.083	0.237	-0.348; 0.821	-0.126	-0.402; 0.150
CYLT-Children						
Learning	-1.637	-3.526; 0.253	0.255	-4.311; 4.820	-1.361	-3.110; 0.387
Short delay	-0.538	-1.097; 0.021	-0.510	-2.012; 0.992	-0.535	-1.059; -0.011
Long delay	-0.723	-1.351; -0.094	-0.244	-1.669; 1.181	-0.645	-1.221; -0.069
Recognition	-0.261	-0.725; 0.203	0.760	-0.358; 1.877	-0.114	-0.546; 0.318

Table 2

Effect of 10-fold increase in prenatal exposure to mercury and PCB after mutual adjustment in a joint analysis of Cohort 1 and Cohort 2 using latent exposure variables. Number of children with outcome information was about 900 in Cohort 1 and 160 in Cohort 2.

Outcome	Mercury			PCB		
	β	95%-conf.	<i>p</i>	β	95%-conf.	<i>p</i>
NES2 Finger tapping						
Preferred hand	-1.572	-3.280; 0.136	0.073	0.228	-2.020; 2.476	0.85
Non preferred hand	-0.833	-2.414; 0.748	0.30	-0.401	-2.426; 1.624	0.70
Both hands	-2.883	-6.254; 0.488	0.094	-0.064	-4.398; 4.271	1.00
NES2 CPT						
Reaction time	32.073	7.000; 57.145	0.013	4.044	-20.977; 29.065	0.75
Ln total missed	0.166	0.049; 0.282	0.0056	-0.061	-0.177; 0.055	0.31
Boston Naming Test						
No cues	-1.379	-2.849; 0.091	0.068	-1.479	-3.435; 0.477	0.14
Cues	-1.584	-3.025; -0.143	0.033	-1.268	-3.172; 0.636	0.19
WISC(R)						
Similarities	-0.373	-1.333; 0.587	0.45	0.037	-1.101; 1.176	0.95
Sqrt. Block Designs	-0.188	-0.495; 0.118	0.23	-0.024	-0.420; 0.372	0.90
Digit Spans	-0.403	-0.781; -0.026	0.035	0.548	0.084; 1.012	0.021
CVLT-Children						
Learning	-1.230	-3.664; 1.205	0.32	-0.280	-3.423; 2.864	0.86
Short delay	-0.795	-1.529; -0.061	0.034	0.519	-0.433; 1.471	0.29
Long delay	-0.472	-1.259; 0.315	0.24	-0.379	-1.385; 0.628	0.46
Recognition	-0.104	-0.712; 0.504	0.74	-0.043	-0.859; 0.773	0.92

Table 3

Effect of 10-fold increase in prenatal exposure to mercury and PCB on two latent neurobehavioral functions. The verbal effect is expressed in the scale of the Boston Naming Test cued, while the scale of finger tapping with preferred hand was used for the motor effect.

Outcome	Mercury			PCB		
	β	95%-conf.	<i>p</i>	β	95%-conf.	<i>p</i>
Verbal	-1.523	-2.868; -0.178	0.027	-0.739	-2.503; 1.025	0.41
Motor	-1.097	-2.459; 0.265	0.11	-0.186	-1.936; 1.564	0.83

Table 4

Comparison of estimated mercury effects and standard errors using separate univariate response models or a multivariate response model for the verbal outcomes

Outcome	Mercury effect		Standard error	
	univariate	multivariate	univariate	multivariate
Boston Naming Test				
Cues	-1.584	-1.523	0.735	0.686
No cues	-1.379	-1.476	0.750	0.665
WISC(R)				
Digit span	-0.403	-0.183	0.193	0.082
CVLT-Children				
Learning	-1.230	-1.843	1.242	0.830
Short delay	-0.795	-0.401	0.374	0.181
Long delay	-0.472	-0.462	0.401	0.208
Recognition	-0.104	-0.227	0.310	0.102