

Manuscript submitted for Toxicology Letters (EUROTOX Proceedings) - revised

Benchmark Dose Calculations of Methylmercury-Associated Neurobehavioural Deficits

Esben Budtz-Jørgensen,¹ Philippe Grandjean,^{2,3} Niels Keiding,¹ Roberta F. White,^{2,3} Pal Weihe^{2,4}

¹Department of Biostatistics, Institute of Public Health, Panum Institute, University of Copenhagen, Copenhagen, Denmark

²Institute of Public Health, Odense University, Odense, Denmark

³Departments of Neurology and Environmental Health, Boston University Schools of Medicine and Public Health, Boston, Massachusetts

⁴Department of Occupational and Public Health, Faroese Hospital System, Tórshavn, Faroe Islands

Please address correspondance to: Dr. P. Grandjean, Institute of Public Health,

Winslowparken 17, 5000 Odense, Denmark. Phone: +45-65573769. Fax: +45-65911458.

Email: p.grandjean@winsloew.ou.dk

Abstract

Prenatal methylmercury exposure is associated with neuropsychological deficits in Faroese children at age 7 years. Lower confidence bounds of benchmark doses (BMDLs) have now been calculated. With the cord-blood mercury concentration as the dose parameter, a logarithmic dose-response model tended to show a better fit than a linear dose model for the attention, language and verbal memory tests. The lowest BMDLs averaged about 5 µg/l cord blood, which corresponds to a maternal hair concentration of about 1 µg/g. However, most BMDLs for hair mercury concentrations were higher. Thus, the results of the benchmark calculations depend on the assumed dose-response model.

Key words: food contamination; prenatal exposure delayed effects; risk assessment

Methylmercury is a common contaminant of seafood and freshwater fish (EPA, 1997). While adverse effects have been unequivocally demonstrated in poisoning incidents, the implications of lower-level exposures in fish-eating populations have been controversial (Davidson et al., 1998; Mahaffey, 1998). A birth cohort was generated in the Faroe Islands during 1986-1987 and is being studied prospectively to examine the possible adverse effects of prenatal exposure to methylmercury (Grandjean, et al., 1997). Deficits in several domains of brain function were found to be associated with increased prenatal exposures to methylmercury, while postnatal exposures appeared less important (Grandjean et al., 1997; Grandjean et al., 1999a). Regression analyses were used to determine these associations, but such data are of limited use in the consideration of thresholds of effects or safe levels of exposure. An attractive approach in this regard is to calculate so-called benchmark doses and their lower confidence limits (Crump, 1995). The present paper therefore presents benchmark calculations for indicative findings of this study.

Study population

The cohort of 1022 singleton births in the Faroe Islands was assembled during 1986-1987, and mercury concentrations were measured in maternal hair at parturition ($\mu\text{g/g}$) and in cord blood ($\mu\text{g/l}$) as biomarkers of the prenatal methylmercury exposure (Grandjean et al., 1997). The latter was thought to be the best indicator of the amount of this neurotoxicant that had reached the fetal circulation and the fetal brain during susceptible periods of development.

Because the effects of fetal childhood exposure to methylmercury are persistent (EPA, 1997), detailed examination of the children was carried out at school age, i.e., in 1993-1994, when advanced neurobehavioral testing would be feasible. Seven cohort children had died,

and a total of 917 of the surviving children (90.3%) completed the examinations, 443 of them in 1993.

Neuropsychological tests

Neuropsychological tests were 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 neurotoxicants (Grandjean et al., 1997; White et al., 1994). Tests with advantageous psychometric properties, i.e., with a wide range of scores possible without floor or ceiling effects, and acceptable test-retest reliability, were preferred, including computer-assisted tests. Based on the scores obtained, one test was selected for detailed study for each of five domains of brain function (Grandjean et al., 1997):

Motor speed: Neurobehavioral Evaluation System (NES2) Finger Tapping Test, maximum number in 15 s with the preferred hand. This task is a measure of manual motor ability that focuses specifically on speed.

Attention: NES2 Continuous Performance Test, average reaction time during the last three of the four minutes of testing. This test measures vigilance/attention.

Visuospatial performance: Bender Visual Motor Gestalt Test, number of errors in copying the 9 figures as scored by the Göttingen system. This visuospatial test has been used extensively as a non-specific measure of brain damage but may also be informative in regard to the functioning of the right cerebral hemisphere.

Language: Boston Naming Test, total number of correct responses, including semantic and phonemic cues. This language test has been shown to be sensitive to metal

neurotoxicity.

Verbal memory: California Verbal Learning Test (Children), number correct at spontaneous long-delay recall of 12 words. This test assesses short-term memory.

Benchmark calculations

In examining dose-response relationships, we have used the approach suggested by Crump (1995). Thus, the benchmark dose (BMD) is the dose of a substance that results in an increased probability of abnormal test performance by a benchmark response (BMR), i.e., from P_0 for an unexposed child to $P_0 + \text{BMR}$ for a child at the BMD. In accordance with previous applications of this method (Crump, 1995; EPA, 1997), P_0 is kept at 5%. Thus, the abnormal test performance is defined by a probability of 5% in an unexposed population. A lower confidence limit (BMDL) is then calculated as the statistical 95% lower bound of the BMD. The BMR is often set at 10%, and the corresponding BMD then results in a tripling of the probability of an abnormal response. We have used BMRs of both 5% and 10%.

The benchmark concept was originally developed in the context of highly standardized carcinogenesis trials with a dichotomous outcome. In the present application of the method, the graded responses are influenced by confounders in addition to the exposure of interest. The preferable adaptation of the benchmark concept to this situation is to use the same unexposed risk (P_0) for each child and a linear multivariate model for the dependence on the confounders. Accordingly, for each child the level of abnormal test performance depends on the individual values of the confounders for that child. This approach avoids dependence of the BMD and BMDL on the confounders (Budtz-Jørgensen, unpublished data).

The dose is expressed as the mercury concentration in cord blood or in maternal hair, i.e., the two exposure biomarkers reflecting prenatal methylmercury exposure (Grandjean et al., 1997).

Values of BMD and BMDL may be calculated using a statistical dose-response model based on power functions (Crump, 1995) for the dependence of a child's expected test score on the mercury dose (d):

$$\mu(d) = \beta \cdot d^K$$

The power parameter K is restricted to values equal to or above 1, thus allowing the dose-response curve to be nonlinear. This approach has the advantage of avoiding unreasonably low BMDLs, which may otherwise occur if the dose-response curve becomes infinitely steep at low doses. Given these restrictions (Crump, 1995), a power of 1 generally provided the best fit to the Faroese data, i.e., a linear association between dose and response.

A logarithmic dose parameter has been used in previous regression analyses (Grandjean et al., 1997). To examine the robustness of the benchmark calculations, both a linear and a logarithmic dose parameter were therefore used. To avoid problems with an infinite slope of the logarithmic curve at zero dose, the dose was transformed to the logarithm of the mercury concentration + 1. The following two dose models were therefore used:

Linear dose-response model: $\mu(d) = \beta \cdot d$

Logarithmic dose-response model: $\mu(d) = \beta \cdot \log(d + 1)$

To test the goodness of fit of the two dose-response models, each was tested against an expanded model that included both the linear term and the logarithmic term.

The confounders used for adjustment were identical to the ones previously identified (Grandjean et al., 1997). Thus, obligatory confounders were: Age, sex and the maternal

Raven score (as a measure of maternal intelligence). When relevant, the child's computer experience was included as well. Additional covariates included for confounder adjustment were selected if they were significantly associated with at least three of the 20 neuropsychological response variables: medical risk, daycare, maternal and paternal education, paternal employment (Grandjean et al., 1997). This uniform group of confounders was also used in the present calculations to adjust all responses.

Results

The number of children completing the tests varied between 837 and 901 for four of the tests. For the attention test, the results differed between the first and the second year, and a joint regression equation could not be generated (Grandjean et al., 1997). Since supervision was stringent only during the first year, the results from the 428 children examined during the first year were used for the purpose of detailed statistical analyses. The numbers available for analyses were slightly reduced because of missing information on some of the covariates.

The BMDLs for cord-blood concentrations are shown in Table 1. The results for the two models are clearly different, especially for the 5% BMR. The logarithmic transformation tended to show a better fit than the linear dose model for the attention, language and verbal memory tests. These tests also showed the lowest BMDLs. With half as many observations for the reaction time parameter, this outcome variable showed particularly low BMDLs. For motor speed and visuospatial performance, the two models could not be distinguished with regard to goodness of fit. Yet, the BMDLs obtained with the two models are clearly different.

When the maternal hair-mercury concentration is used as the exposure variable, no certain difference between the two different curve functions was found. The BMDLs for the

two models (Table 2) generally show a difference less than 3-fold for both the 5% and the 10% BMRs.

A serious disadvantage of the K power model and the linear model is that they are very sensitive to single observations at high dose levels. Thus, the highest cord-blood mercury concentration was 351 $\mu\text{g/l}$, and the results for this child had a profound influence on the BMDL results using the linear dose function. The BMDLs for this model decreased by 10-20% if this single child was excluded from the calculations. The effect on the logarithmic model results was negligible.

The results obtained with the two sets of exposure biomarkers (Tables 1 and 2) may be compared when taking into account that 1 $\mu\text{g/g}$ hair corresponds to about 5 $\mu\text{g/l}$ cord blood. If applying this conversion factor, the linear BMDLs are lower for the hair-mercury dose parameter than for cord-blood, while the opposite is true for the logarithmic dose-response model BMDLs.

Using the logarithmic dose-response model with a BMR of 10%, the lowest BMDLs are obtained for the attention measure (6.1 $\mu\text{g/g}$ hair and 4.0 $\mu\text{g/l}$ cord blood). The results for the language test are somewhat higher, i.e., 9.1 $\mu\text{g/g}$ hair and 9.7 $\mu\text{g/l}$ cord blood. The results for the linear curve function are higher, up to a factor of 20 for the cord-blood values. For both sets of dose indicators and for both curve functions, lower BMDLs are obtained if BMR is set at 5%, rather than 10%.

Discussion

Benchmark calculations are meant to provide an approximate threshold level which can be interpreted as a No-Observed Adverse Effect level (NOAEL) from experimental studies

(Crump, 1995). The lower confidence limit takes into account the uncertainty in the estimation of the relation between dose and effect given that the dose-response model is known. The uncertainty associated with the model determination is not taken into account. Thus, different results may be obtained with other model assumptions. Also, the population is dealt with as a uniform group, and individual hypersusceptibility is disregarded. For calculation of safe exposure levels, an uncertainty factor may therefore be needed. The U.S. Environmental Protection Agency (1997) used this approach when using data from the Iraqi poisoning incident as basis for their estimation of a Reference Dose for methylmercury exposure. In this case, with subacute exposure and crude assessment of neurological dysfunctions, an uncertainty factor of 10 was applied to the average of the BMDLs calculated from several model functions.

In the absence of detailed a priori knowledge on the shape of the dose-response curves, two different models have been used in the present study. One is similar to the transformation used for the regression analyses previously conducted (Grandjean et al., 1997), although we in the present study added 1 to the mercury concentration to avoid problems with increasing steepness of the curve at low doses. The other model used is in accordance with the recommendations by Crump (1995). Results for both models are provided (Tables 1 and 2), but it is possible that a third function may provide a better fit and perhaps lower BMDL values.

When the maternal hair-mercury concentration is used as the exposure variable, no certain difference between the different curve functions appeared. However, for the cord-blood concentration, the logarithmic transformation seems to fit better than the linear model for the attention, language and verbal memory measures. This finding therefore suggests that

power functions (including the linear function) used by Crump (1995) may not provide the best fit to the data. At least for the present data set, it would therefore seem appropriate to include the logarithmic dose-response model in the consideration of benchmark doses. With the wide range of mercury concentrations, this transformation also normalizes the distribution. The results therefore become less sensitive to extreme data points. Thus, the linear model is highly affected by the results for the child with the highest mercury exposure, while this effect is not seen in the logarithmic model.

The reaction time measure and the naming test showed the lowest BMDLs. The same two tests also appeared to be most sensitive to prenatal methylmercury exposure (Grandjean et al., 1997). Thus, it would be meaningful to base risk considerations on the results obtained for these two tests. Further, as the cord-blood mercury concentration is a better risk indicator than is the maternal hair-mercury concentration (Grandjean et al., 1997; Grandjean et al., 1999a), the blood concentration would seem most appropriate for risk assessment purposes.

The Reference Dose for methylmercury corresponds to a hair-mercury concentration of 1.1 $\mu\text{g/g}$ (EPA, 1997). That level would translate to about 5.5 $\mu\text{g/l}$ cord blood. For the attention and language variables, which showed the strongest association with the cord-blood mercury concentration (Grandjean et al., 1997), the average of the 5% and 10% BMDLs is approximately 5 $\mu\text{g/l}$ cord blood, i.e., the same order of magnitude as Reference Dose. These calculations would therefore suggest that the Reference Dose is in overall accordance with the present BMDL results, although without taking into account any uncertainty factor. At the same time, it should be stressed that the BMDL calculations are highly sensitive to choices of P_0 , BMR, and curve model, and the results therefore should only be taken as indicative of approximate orders of magnitude.

The Faroese cohort study may be particularly relevant for risk assessment purposes for several reasons. The culture is Scandinavian, i.e., western European, with a traditional and stable family structure, and a highly efficient societal system of health care, education and social support. Dietary deficiencies are nonexistent, and the intake of essential fatty acids and selenium is high because of their occurrence in seafood. The median birth weight in the Faroes is 3700 g, i.e., among the highest in the world, and only 14 cohort children had a birth weight below 2500 g. The rate of preterm birth is very low - only 29 cohort children were born before the 37th week of gestation. Breastfeeding in the Faroes is mostly continued for at least six months, with more than 97% of the cohort children being breastfed for at least one month. Alcohol intake during pregnancy is minimal, with more than 75% of the mothers being total abstainers. Thus, the population is relatively homogeneous, and socioeconomic confounding is expected to be limited.

Most of the methylmercury exposure in the Faroes originates from pilot whale, a traditional food item which is eaten as thin slices of cured meat or larger steaks. Many Faroese also eat whale blubber which contains polychlorinated biphenyls (PCBs) (Borrell and Aguilar, 1993). These substances have been linked to neurobehavioral dysfunctions (Jacobson et al., 1990), although the specific causative congeners have not been identified, and confounding from methylmercury has not been excluded. Nonetheless, we determined the PCB concentrations in cord tissue for half of the cohort (Grandjean et al., 1997). Detailed statistical analyses showed that confounding is limited, and interaction is unlikely with the methylmercury-associated effects (Budtz-Jørgensen et al., 1999).

Another cohort is being followed in the Seychelles, but results have only been published from examinations up to 5.5 years of age (Davidson et al., 1998). In this study, a

variety of developmental tests have been used, and the prenatal methylmercury exposure was determined from the maternal hair concentration only. With a population from a mid-income developing country, where the family structure is different, it is perhaps not surprising that subtle mercury neurotoxicity has not been demonstrated (Grandjean and White, 1999).

Although the Seychelles study may offer some confidence that methylmercury effects may be of a negligible magnitude in some fish-eating populations, the Faroese data provide information on subtle effects in a homogeneous population as determined with sensitive methods. Support for this notion has now emerged from studies of newborns in the Faroe Islands (Steuerwald et al., in press), and school-age children from Madeira (Murata et al., 1999) and the Brazilian Amazon (Grandjean et al., 1999b).

Acknowledgments

This cohort study was supported by grants from the U.S. National Institute of Environmental Health Sciences (ES06112), the European Commission (Environment Research Programme), the Danish Medical Research Council, and the Dannin Foundation. The benchmark calculations were supported by the Danish Health Foundation and the U.S. Environmental Protection Agency.

References

- Borrell, A., Aguilar, A., 1993. Pollution by DDT and PCB in blubber and muscle of long-finned pilot whales from the Faroe Islands. In: Donovan, G.P., Lockyer, C.H., Martin, A.R., (Eds.), *Biology of Northern hemisphere pilot whales (Special Issue 14)*. International Whaling Commission, Cambridge, pp. 351-367.
- Budtz-Jørgensen, E., Keiding, N., Grandjean, P., White, R.F., Weihe, P., 1999. Methylmercury Neurotoxicity Independent of PCB Exposure (letter). *Environ. Health Perspect.* 107, A236-237.
- Crump, K.S., 1995. Calculation of benchmark doses from continuous data. *Risk Anal.* 15, 79-89.
- Davidson, P.W., Myers, G.J., Cox, C., Axtell, C., Shamlaye, C., Sloane-Reeves, J., Cernichiari, E., Needham, L., Choi, A., Wang, Y., Berlin, M., Clarkson, T.W., 1998. Effects of prenatal and postnatal methylmercury exposure from fish consumption on neurodevelopment. *JAMA* 280, 701-707.
- EPA, 1997. *Mercury Study Report to Congress*. U.S. Environmental Protection Agency, Washington, D.C.
- Grandjean, P., Weihe, P., White, R.F., Debes, F., Araki, S., Murata, K., Sørensen, N., Dahl, D., Yokoyama, K., Jørgensen, P.J., 1997. Cognitive deficit in 7-year-old children with prenatal exposure to methylmercury. *Neurotoxicol. Teratol.* 19, 417-428.
- Grandjean, P., Budtz-Jørgensen, E., White, R.F., Jørgensen, P.J., Weihe, P., Debes, F., Keiding, N., 1999a. Methylmercury exposure biomarkers as indicators of neurotoxicity in 7-year-old children. *Am J Epidemiol* 150, 301-305.
- Grandjean, P., White, R.F., 1999. Effects of methylmercury exposure on neurodevelopment

(letter). JAMA 281: 896.

Grandjean, P., White, R.F., Nielsen, A., Cleary, D., de Oliveira Santos, E.C., 1999b. Mercury neurotoxicity in Amazonian children downstream from gold mining. *Environ. Health Perspect.* 107, 587-591.

Jacobson, J.L., Jacobson, S.W., Humphrey, H.E.B., 1990. Effect of in utero exposure to polychlorinated biphenyls and related contaminants on cognitive functioning in young children. *J. Pediatr.* 116, 38-45.

Mahaffey, K.R., 1998. Methylmercury exposure and neurotoxicity (editorial). *JAMA* 280, 737-738.

Murata, K., Weihe, P., Renzoni, A., Debes, F., Vasconcelos, R., Zino, F., Araki, S., Jørgensen, P.J., White, R.F., Grandjean, P. 1999. Delayed evoked potentials in Madeiran children exposed to methylmercury from seafood. *Neurotoxicol. Teratol.* 21, 343-348.

Steuerwald, U., Weihe, P., Jørgensen, P.J., Bjerve, K., Brock, J., Heinzow, B., Budtz-Jørgensen, E., Grandjean, P., 1999. Maternal seafood diet, methylmercury exposure, and neonatal neurological function. *J. Pediatr.* (in press).

White, R.F., Debes, F., Dahl, R., Grandjean, P., 1994. Development and field testing of a neuropsychological test battery to assess the effects of methylmercury exposure in the Faroe Islands. In: *Proceedings of the international symposium on assessment of environmental pollution and health effects of methylmercury*, Kumamoto, Japan, pp. 127-40.

Table 1. Lower 95% confidence limit for the benchmark dose expressed as the cord-blood mercury concentration ($\mu\text{g/l}$) with a background prevalence of abnormal response (P_0) of 5%, as calculated for a linear and a logarithmic dose-response model at benchmark responses of 5% and 10%.

Test	N	Benchmark response				p*	
		5%		10%		Linear	Logarithmic
		Linear	Logarithmic	Linear	Logarithmic		
Motor speed	901	79	7.9	132	38	0.64	0.41
Attention	428	46	1.6	76	4.0	0.02	0.73
Visuospatial	895	104	12.7	174	79	0.46	0.92
Language	865	58	3.1	96	9.7	0.06	0.67
Verbal memory	837	103	7.6	173	35	0.07	0.42

*p for difference between the single curve function model and an expanded model that includes both functions

Table 2. Lower 95% confidence limit for benchmark dose expressed as the maternal hair-mercury concentration ($\mu\text{g/g}$) with a background prevalence of abnormal response (P_0) of 5%, as calculated for a linear and a logarithmic dose-response model at benchmark responses of 5% and 10%. Neither model differs from an expanded model that includes both dose-response functions. The results may be compared with those in Table 1 if taking into account that cord-blood mercury concentrations (in $\mu\text{g/l}$) are approximately 5.0 times greater than maternal hair-mercury concentrations (in $\mu\text{g/g}$).

Test	N	Benchmark response			
		5%		10%	
		Linear	Logarithmic	Linear	Logarithmic
Motor speed	901	11.3	4.3	19	15
Attention	428	9.4	2.2	16	6.1
Visuospatial	895	14.0	6.8	24	30
Language	865	9.6	3.0	16	9.1
Verbal memory	837	13.4	4.8	22	18