COMMENTS ON THE UNCERTAINTY AND VARIABILITY MODELING DONE FOR THE FDA FISH MERCURY NET BENEFITS ANALYSIS

WORKING PAPER NO. 2011-20
SEPTEMBER 2011
By Dale Hattis

NATURAL

CONSTRUCTED

1887 CLARK UNIVERSITY
George Perkins Marsh Institute

Comments on the Uncertainty and Variability Modeling Done for the FDA Fish Mercury Net Benefits Analysis

Dale Hattis
Clark University

Final Report
Order Number EP-11-H-001567, dated 7/18/11

September 7, 2011

Table of Contents

1. Overview

This brief comment presents initial conclusions on the strengths and weaknesses of the FDA uncertainty/variability modeling based on information gleaned from three sources:

> The revised FDA report, especially the technical appendices,
>
> The full Excel modeling and results worksheets provided on disk,
>
> Extended conversations with the principal developer of the model, Clark Carrington, on August 18 and 22, 2011.

The uncertainty/variability modeling framework created by FDA in consultation with EPA can potentially make valuable contributions not only to the immediate issues of framing fish consumption advisories, but also for future long term assessments of the likely benefits of incremental reductions to current mercury emissions from sources in the U.S. and internationally. Therefore it is potentially helpful to make the modeling tool as faithful as possible in incorporating available information about both mercury and fish consumption effects, and as readily adaptable as possible to different policy assessment questions and model assumptions. For example issues that could be addressed include consideration of national advisory policy vs decision-making on advisories for consumption of fish from local water bodies, vs policies for control of mercury emissions from coal fired power plants. To help address these ongoing questions it is important for the model to be both a good reflection of presently available information, and to be periodically updated to reflect changing circumstances in the model parameters such as fish consumption patterns for different population groups, and rates of consumption of different kinds of fish and fish oil supplements which may partially realize some of the benefits attributed in the model to fish consumption.

1.1 Philosophy of Science Background

The FDA modeling represents an extensive effort to represent available data and draw distributional inferences on the extent of neurodevelopmental seafood consumption benefits and risks for different portions of the overall U.S. population from commercial seafood. It does not appear to reflect obvious policy-related biases. The best judgment of this reviewer is that the analysis is an honest job from the perspective of the principal analyst, Dr. Clark Carrington. However it does reflect deep "positivist"/"frequentist" philosophy of science presumptions on the part of the analyst that appear to have steered the work away from directions that might have led to improved clarity, credibility, and usefulness for drawing policy-relevant conclusions.

The discussion below is shaped by a different philosophy of science orientation that can best be summarized as "critical realist"/"Bayesian". The key difference is

- Positivist philosophy of science enshrines "the data" as paramount in the scientific enterprise, and tends to view mathematical analyses as chiefly helpful in summarizing the

data and its direct implications. Theory, and in particular any postulation of constructs that cannot be directly observed and measured is discouraged.

- "Critical realist" philosophy of science views "the data" as a clouded window through which some underlying reality may be inferred, with the aid of mathematical analyses representing quantitative theory about causal relationships. Data and the comparative predictions of alternative theories about the world are both subjected to critical analysis to synthesize a tentative picture of the state of the world that is subject to continual updating as new information emerges. The prime object of analysis is not "the data" but the underlying reality that produced the data through the agency of causal processes that can only be understood by using the new information and causal inferences to update our prior understanding of how the world works.

A small illustration can help clarify this. In discussing the distributions of seafood mercury concentrations in different species groups, three different approaches were used depending on the amount and type of data available:

- Where only means and standard deviations were available for a specific set of seafood species, the "Analog" method was used to infer a range of parameters of either a lognormal or a gamma distribution that could produce the same summary variability statistics, with uncertainty represented as the difference between the two mathematical forms and the correlated ranges of fitting parameter values.

- For seafood categories where somewhat more individual observations of mercury concentrations were available, the approach was to "fit" a series of 11 distributional forms, and retain the best four (as judged by the sum of squares of the residual errors after using the best fitting parameter values for each model). The uncertainty in this case is represented by the differences among the four best-fitting models. A critical realist perspective would tend to place greater weight on distributional models that have some underlying mechanistic basis—such as the lognormal, which is expected to be produced when there are many factors that cause differences in mercury concentrations among seafood samples within a category, and when these factors tend to act multiplicatively in influencing concentrations.

- For seafood categories where relatively extensive mercury concentration data were available, (hundreds of individual data points), an "empirical distribution" was used directly from the data--that is the frequency of the different concentration levels themselves was used to represent variability. In this case, the analyst assumed that there was no uncertainty in the derived variability distribution, because the empirical distribution perfectly describes the data.

By contrast, a critical realist philosophical orientation would lead the analyst to expect in the last case above that there is both some remaining uncertainty due to (likely small) sampling error that uses the data to represent a large universe of real seafood samples, but also bias from some inevitable measurement errors that must in principle increase the variation of the distribution of the observed data relative to the true underlying mercury concentrations. This effect can be corrected if an estimate of the measurement error can be made, and the measurement variance

subtracted from the observed variance. The error in this particular case is not likely to have distorted the analysis in a major way because the analyst has taken the precaution of comparing his derived mercury distribution with some biomarkers of mercury in people derived from nationally representative NHANES data (Figures AB-1 and AB-2 in Appendix A of the FDA report.) However this example does illustrate the philosophy difference and the potential to at least modestly expand the analysis in some respects.

More serious difficulties are possible in the estimation of fish benefits using the Hibbeln study (Hibbeln et al., 2007) without assessing the possibility of imperfectly controlled confounding with socioeconomic status and other influences on IQ, although there has been a post-hoc attempt to factor in the expected adverse effects of methylmercury on the offspring of mothers consuming various amounts of fish in the United Kingdom, where the study was done. Implicitly the FDA analysis has left the findings of different authors from the individual studies, and different distributional models applied to the same data, as a representation of uncertainty rather than attempting a more integrative synthesis of the information. Further, a variety of mathematical forms are often applied to data to represent different possibilities for the shapes of dose response relationships, or the distributions of fish mercury concentrations within species groups. These analyses generally reflect weightings only according to statistical goodness-of-fit criteria, and no attempt at weighting from fundamental mechanistic considerations. Such mechanism-based weighting could help influence potential integrative "bottom line" analyses that may be helpful to promote clearer understanding by decision-makers and the public of the likely outcomes of alternative policy choices.

1.2 Outline and Objective of This Report

The balance of this report discusses the uncertainty/variability modeling in the FDA analysis under two major headings:

- Exposure and Methylmercury-IQ Dose Response Assessment
- PUFA-IQ Benefits and Net Benefits Assessment

To the extent possible, the analysis assesses both inputs and the outputs for the variability and uncertainty dimensions. First, however, it is helpful to devote a few paragraphs to a general clarification of how these two probabilistic dimensions differ, and the ways they pose distinct and in some ways contrasting challenges for analyses of data.

By "interindividual variability" is meant the real variation among individuals or cases in exposure-producing behavior, in exposures, or some other parameter (such as differences among individual municipal solid waste incinerators in emissions). Variability is of interest in policy analysis because, among other things, it helps confront equity issues—to what degree does a risk or policy option differentially affect different portions of the population? By contrast "uncertainty" is a description of the imperfection in knowledge of the true value of a particular parameter or its real variability in an individual or a group. In general uncertainty is reducible by additional information-gathering or analysis activities (better data, better models) whereas real variability will not change (although it may be more accurately known) as a result of better or

more extensive measurements. Uncertainty is of interest to decision makers in part because it helps confront the robustness of the available information on which a policy choice can be made. How likely is it that the expected risks and benefits from a specific policy scenario could be different enough from the central estimates that decision makers might wish to choose another option? And what improvement in the bases for decision-making might be obtained if time and resources were invested in reducing specific sources of uncertainty (this latter goes by the name of value-of-information analysis).

Wide acceptance by professional risk assessors of this basic distinction between variability and uncertainty is now about a quarter century old (Bogan, 1987) (Hattis, 1987) (Hattis, 2003). However it is still not yet widely appreciated that standard statistical techniques for making inferences about these two probabilistic dimensions from available data typically suffer from different kinds of systematic inaccuracies (Hattis, 1994).

On the one hand, standard statistical inferences of variability, such as the standard deviation, of a set of observations, nearly always tend to overstate real variability. This is because nearly all sets of observations contain some spreading from measurement/estimation errors beyond the real variation of the true underlying values of the measured parameter. However estimation and correction for this extra spreading is still rare in risk analyses.

On the other hand, standard statistical inferences of uncertainty, such as standard errors, almost always underestimate real uncertainty because they exclude sources of systematic errors that affect all the data points simultaneously, such as imperfect calibration of the instruments used to make the measurements, etc. Empirical observations suggest that this is often a large error, which means that distributions of uncertainty may often be non-gaussian in shape, with fatter tails and much larger probabilities of large deviations than would be expected from standard theory based solely on random errors analyzable from the fluctuations of individual points within data sets (Shlyakhter, 1992).

From a policy context the most critical of these parts of the FDA analysis are the Methylmercury-IQ and the PUFA-IQ Benefits assessments. The major policy-related conclusion that FDA draws from the analysis is that women of reproductive age should be encouraged to eat a minimum amount of fish to obtain the PUFA-related benefits for their children (preferably relatively low-mercury fish). The principal ways this conclusion could be wrong are that either,

- The PUFA-IQ related developmental benefit of fish consumption is greatly over-estimated, or

- The developmental IQ harm from mercury-containing fish consumption is greatly under-estimated.

Therefore it is sensible for an analysis of the FDA document to focus on these particular components of the work. Section 4 of this report briefly summarizes some conclusions and suggestions for improvement of the FDA analysis.

An important obstacle that has hindered timely completion of this analysis is that there is no single set of "bottom line" results in section V of the FDA report. Rather, there are multiple tables presenting results of both population and species-by-species net benefit analyses that

incorporate different combinations of neurological harm and neurological benefit dose response relationships as analyzed in different studies, without a clear articulation of which component studies have contributed to which table, and which sources of uncertainty are represented in the confidence limits. It is possible to at least partially reconstruct the mapping of data/analysis references to specific tables and columns of results in Section V by tracing the flow of the detailed results presented in Appendix B, with the proviso that the Appendix B results are stated in terms of changes of Z-scores, whereas for Section 5 these have been multiplied by 15 to make the conversion to IQs or putative IQ equivalents.

2. Exposure and Methylmercury-IQ Dose Response Assessment

2.1 Comments on Methodology

Table V-4 on page 95 presents the estimated contributions of methylmercury to different neurodevelopmental metrics related to IQ by approximately age 6-9, as estimated from the integrative studies of (Axelrad, 2007) and (Cohen et al., 2005). Of these, there is reason to prefer the lower estimates from the EPA-sponsored Axelrad et al. study because these authors had the benefit of fits of the raw data to low dose linear dose response relationships, whereas Cohen et al. needed to rely on interpretations of the Faroe Island study authors' analyses based on log-transformed estimates of individual methylmercury exposures (Grandjean, 1997), likely leading to overestimates of the low dose effect observed in that study. The results in this table do not include estimates of the adverse methylmercury effects derived by FDA itself for subjects of much younger ages (Carrington, 2000) that relied on translation of the Iraqi "late walking" observations to putative IQ equivalents, combined with information from the Seychelles study. The expected IQ-equivalent decrements from the Carrington/Bolger 2000 analysis, in the form of Z-score changes[1], do appear in Appendix B, in Table AB-4 on page 173, and are presented later in section V (Table V-7, based on the Iraqi poisoning observations for "late talking", and Table V-8 based on the similar observations for "late walking". Elsewhere in Appendix A (pp. 145-155) there is a comparison between dose response curves derived from the Axelrad 2007 and Carrington 2000 analysis which yields a strong impression that the derived dose response relationships for full IQ change in relation to hair mercury are very similar (See Figure AA13 on page 155). Somewhat larger estimates of effect are also given in these tables for estimates of the verbal component of IQ.

Within the Axelrad study results, the uncertainties were reportedly calculated from a normal distribution derived from stated 95% confidence limits from the study authors' meta-analysis of data from the New Zealand (Crump, 1998), Seychelles (Myers, 2003), and Faroe Island (Budtz-Jorgensen, 2005) (Budtz-Jorgensen, 2007) studies. Within the Cohen et al. study results, uncertainties were reportedly derived from a probability tree analysis using the same weights for different tests contributing to IQ as were used by the study authors. Unfortunately this probability tree analysis does not appear to have been documented in the FDA report.

[1] Numerically, these Z-score changes, in units of standard deviations of the population distribution, are just 1/15th of the values stated in terms of IQ points. IQ is conventionally defined as a measure of "general intelligence" with a mean of 100 and a standard deviation of 15.

One possible source of error in the dose response distributions for neurodevelopmental harm derived from both the Axelrad and Cohen meta-analyses is that there is no mention of a correction of the harm estimate for the putative simultaneous benefits of fish/polyunsaturated fatty acid consumption that should have been present, to varying amounts, in the component studies. (By contrast, the derivation of the beneficial effects of fish consumption does mention a correction for the estimated simultaneous effects of methylmercury from fish.) Such a correction would tend to increase the estimates of neurodevelopmental impairment from methyl mercury, acting by itself. The original paper of Axelrad et al. (2007) does not mention the possibility of partial offsetting effects from fish consumption, so the estimates in that paper should be regarded as net of any beneficial effects of that type that may be present. Recently Stern and Korn (Stern, 2011) have discussed approaches for obtaining unconfounded estimates of benefit and harm from fish consuming populations.

Variability in Table V-4 is primarily represented by different percentiles of U.S. mothers, arranged in order of mercury hair levels, which in turn reflect differences in dietary exposure and hair to blood ratios. Uncertainty is reflected in the difference between the two model analyses, and also some appreciable uncertainties captured within each analysis and stated in terms of 95% confidence limits for each of the variability cut points included in the table. The similar results for the Carrington (2000) analyses reflect a wider range of uncertainties resulting from 200 different dose response model forms applied to the data (selected from approximately 1000 which were originally tried).

2.2 Comments on Distributional Findings

To get a more quantitative handle on the variability/uncertainty results, we can make some probability plots of the reported model findings for both exposure and the estimated adverse effect of methylmercury on IQ. In this type of plot, when log scales are used for the y axis, and standard deviation units (Z-scores) are used for the x axis, the regression line represents the hypothesis that the variability or uncertainty giving rise to the percentiles follows a lognormal distribution. The correspondence of the points to the regression line serves as a quick qualitative indicator of the degree to which the results correspond to the plotted distribution. Systematic departures of the points from the line can indicate whether the data points suggest skinnier or fatter tails than would be expected for a true lognormal. To avoid rounding errors in creating these plots, it would have been desirable to work from the detailed uncertainty X variability tables given on the disk containing the model. Unfortunately it was found that the labeling of the sets of data and parameters in these files is too limited to allow unambiguous identification of the spreadsheet output data with the results presented in different tables in the FDA document.

It is helpful to lay the groundwork for consideration of the uncertainty and variability in the expected methylmercury neurological harm by first plotting the estimated distributions of internal doses as indicated by the central estimates of the distributions of hair and blood mercury levels. Figures 1a and 1b show plots of these variability distributions derived from the data in Table V-3. (The model-predicted population distributions of blood and hair mercury levels themselves are fairly successfully compared to representative observations from U.S. populations in Figures AB-1 and AB-2 on pages 171-172.)

It can be seen in Figures 1a and 1b that lognormal models describe the central estimates of the variability distributions very well, although there is a slight convex curvature in the data points for each plot. The spread (geometric standard deviations) of the fitted distributions are also similar. The corresponding geometric mean and geometric standard deviation for the modeled blood mercury levels are 0.58 ppb and 3.22, respectively, whereas the modeled hair mercury variability distribution is described by a geometric mean of 0.14 ppm and a geometric standard deviation of 3.55. The slightly larger geometric standard deviation for the hair levels is likely produced by modest estimated population variation in hair/blood ratios, which rises from 200 for the lowest percentiles in the table to about 312 at the 99.9th percentile.

Use of these central estimate distributional statistics could allow the generation of some estimates of hair and blood mercury levels beyond the 99.9th percentile, a stated limitation of the current analysis. To judge the likely errors in such estimates we can compare the modeled and distribution-predicted values for the 99.9th percentiles of the distributions:

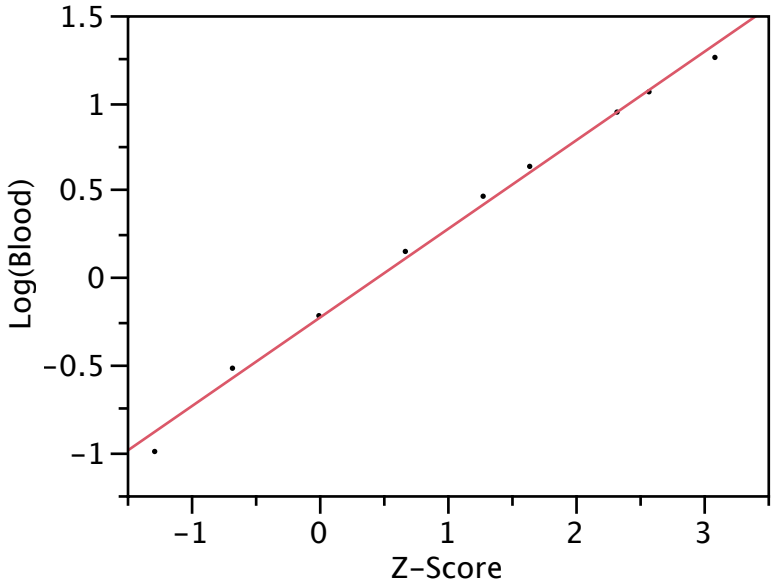
Percentile	Blood 99.9th	Blood 99.99th	Hair 99.9th	Hair 99.99th
Modeled	18.1	--	5.64	--
Projected	21.5	44.8	6.88	15.3

We can now proceed to do comparable log probability plots of the projected effects on IQ for different portions of the population from Table V-4. Figures 2a, 2b, and 2c show the central estimates, lower 95% uncertainty confidence limit, and upper 95% uncertainty confidence limits for the projected IQ effect, respectively for the full scale IQ effects inferred from the Axelrad (2007) meta-analysis. Figure 2a represents the central estimates of the effects on IQ at different exposure variability percentiles as estimated from hair levels, whereas the differences between Figure 2a and the two other figures represents uncertainty in the expected effect at each percentile of the variability distribution.

It can be seen in Figure 2a that when log(IQ change) is plotted vs the Z-score for hair mercury levels the points conform to the lognormal regression line just as well as the plot of the log(hair levels) themselves in Figure 1b. Moreover the amount of interindividual variability indicated by the geometric standard deviation, 3.37, is no greater than for the log(IQ) plot than the corresponding value derived for the log(hair level) plot, 3.55. This means that all the variability in the expected IQ detriment of methyl mercury has flowed through from the estimated exposure variability. No additional variability in susceptibility per unit of exposure has been added that is related to the model representation of the Axelrad (2007) dose response findings, which are low dose linear. On its face, this lack of added variability is might appear dubious in the light of the substantial interindividual variability in pharmacodynamics seen for a variety of other responses to chemicals and drugs (Hattis, 2007), including values corresponding to geometric standard

Figure 1a

Lognormal Probability Plot of the Variability Distribution for Mercury in Blood (ppb) of 16-45 Year Old Women (Central Estimates)



Summary of Fit

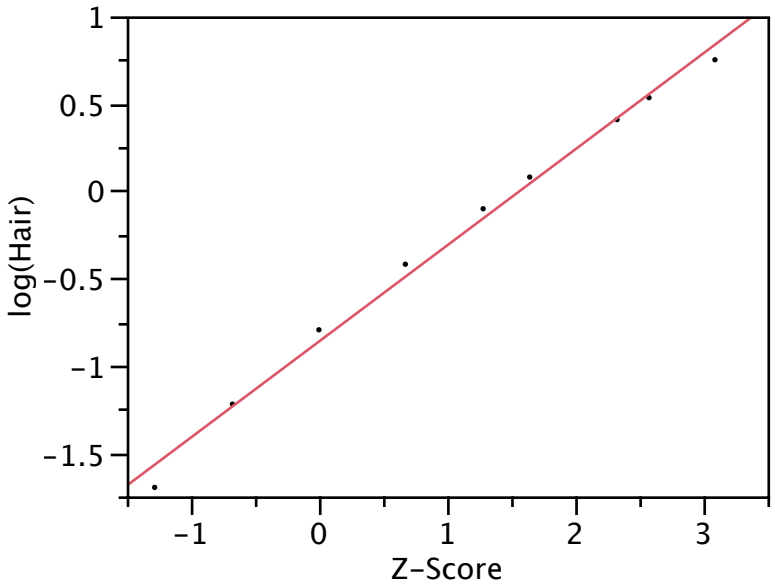
RSquare	0.9943
RSquare Adj	0.9935
Root Mean Square Error	0.0620
Mean of Response	0.3067
Observations (or Sum Wgts)	9

Parameter Estimates

Term	Antilog(Estimate)	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.580	-0.237	0.026	-9.15	<.0001
Z-Score	3.219	0.508	0.015	34.88	<.0001

Figure 1b

Lognormal Probability Plot of the Variability Distribution for Mercury in Hair (ppm) of 16-45 Year Old Women (Central Estimates)



Summary of Fit

RSquare	0.9929
RSquare Adj	0.9919
Root Mean Square Error	0.0749
Mean of Response	-0.2741
Observations (or Sum Wgts)	9

Parameter Estimates

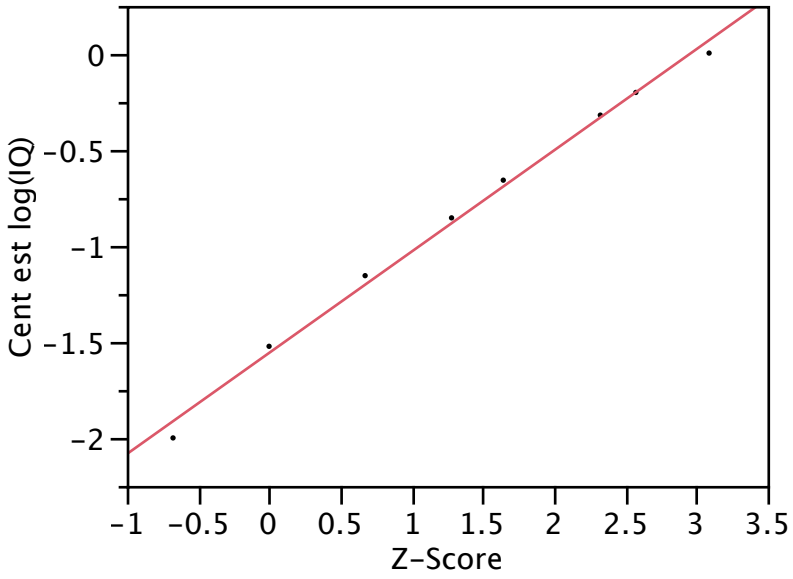
Term	Antilog(Estimate)	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.137	-0.863	0.031	-27.63	<.0001
Z-Score	3.552	0.550	0.018	31.34	<.0001

deviations of 1.3-2.9 for methylmercury effects in adults, and 3.7-15 for various neurological effects in children exposed during gestation, according to log probit analyses of the Iraqi poisoning data published in an Institute of Medicine Report [(IOM, 1991), p. 206]. However, even though appreciable interindividual variability might be present in the linear slopes of neurological effect vs dose for different people, this would not be expected to show up in the population variability plot of Figure 2a unless there were some correlation between the mercury dose and the mean sensitivity of the population to the IQ-lowering effect.

Comparison of Figure 2a with 2b and 2c, indicates that there is appreciable uncertainty in the modeled dose response relationship. At the medians of the variability distributions (Z-Score = 0) the upper 95% confidence limit suggests as much as a 1.8 fold larger effect on IQ than the central estimate; and the lower 95% confidence limit suggests that the effect on IQ might be as small as 3.8 fold less than that expected from the central estimate relationship. This is a reasonable, but not quite perfect reflection of the main conclusion of the Axelrad et al. (2007) study itself, which finds "a central estimate of –0.18 IQ points (95% confidence interval, –0.378 to –0.009) for each parts per million increase of maternal hair mercury…."—that is, a confidence range from 2.1 fold larger than the central estimate to 2.0 fold less than the central estimate. The stated Axelrad 95% confidence limits are multiplicatively symmetric (2.1 fold up vs 2.0 fold down). However the 95% confidence range in the FDA model results for the IQ effect at the median hair level is approximately symmetric arithmetically (0.0509 - 0.0277 = 0.0232 IQ points for the upper 95% confidence limit – the central estimate, vs 0.0277 – 0.0074 = .0203 IQ points for the central estimate – the lower 95% confidence limit. This difference has apparently resulted from the FDA model's use of a normal distribution to represent the uncertainty of the Axelrad IQ effect findings, whereas the confidence limits stated in the summary of the Axelrad paper itself are closer to lognormal. If the desire of the FDA model is to replicate the Axelrad findings on their own uncertainty, then future runs of the model should restate the uncertainty in the Axelrad IQ effect estimates to be lognormal in form rather than normal. It is not clear that this will make a substantial difference in the results, but this change probably should be recommended to FDA for future work.

Figure 2a

Lognormal Probability Plot of the Estimated Negative Effects of Commercial Fish-Borne Methylmercury on IQ Inferred from the Axelrad (2007) Meta-Analysis (Central Estimates)



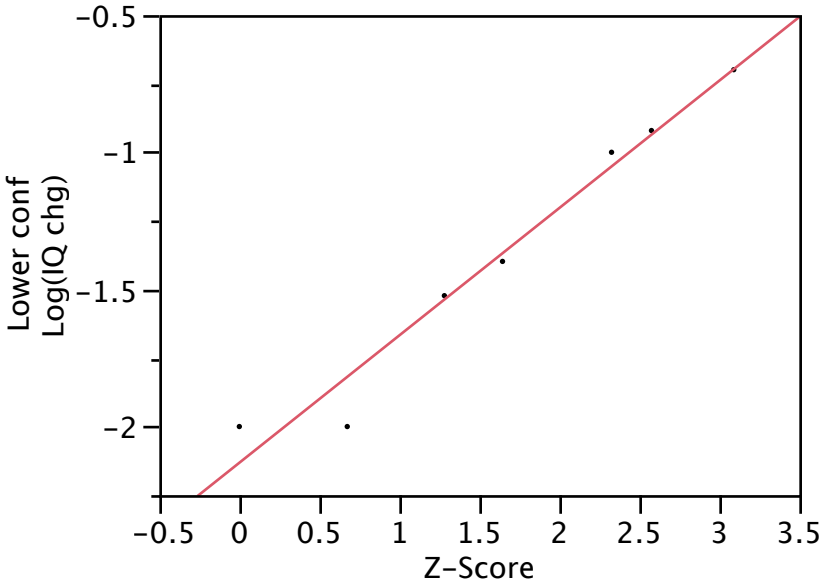
RSquare	0.9948
RSquare Adj	0.9939
Root Mean Square Error	0.0540
Mean of Response	-0.8380
Observations (or Sum Wgts)	8

Parameter Estimates

Term	Antilog(Estimate)	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.028	-1.557	0.029	-54.4	<.0001
Z-Score	3.366	0.527	0.016	33.7	<.0001

Figure 2b

Lognormal Probability Plot of the Estimated Negative Effects of Commercial Fish-Borne Methylmercury on IQ Inferred from the Axelrad (2007) Meta-Analysis (Lower 95% Confidence Limits at Each Variability Percentile)



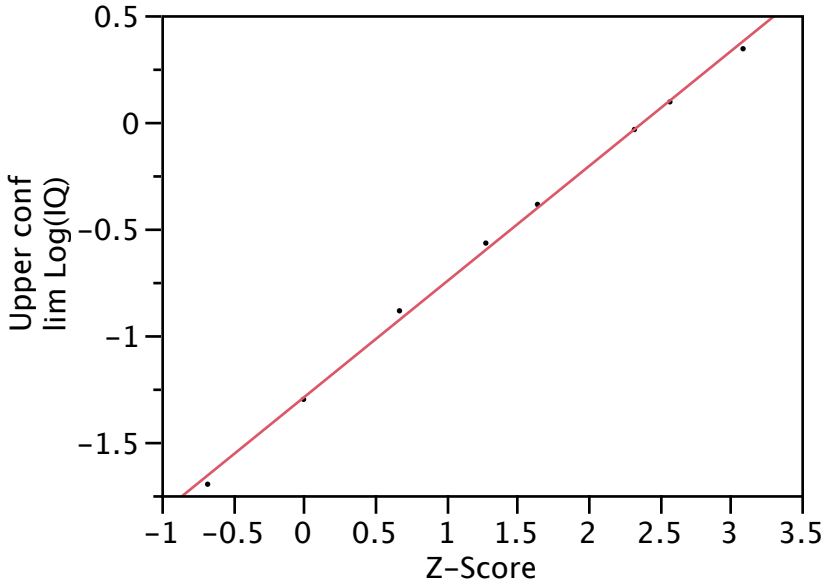
RSquare	0.9662
RSquare Adj	0.9595
Root Mean Square Error	0.1042
Mean of Response	-1.3629
Observations (or Sum Wgts)	7

Parameter Estimates

Term	Antilog(Estimate)	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.007	-2.132	0.075	-28.3	<.0001
Z-Score	2.913	0.464	0.039	12.0	<.0001

Figure 2c

Lognormal Probability Plot of the Estimated Negative Effects of Commercial Fish-Borne Methylmercury on IQ Inferred from the Axelrad (2007) Meta-Analysis (Upper 95% Confidence Limits at Each Variability Percentile)



RSquare	0.9983
RSquare Adj	0.9980
Root Mean Square Error	0.0319
Mean of Response	-0.5553
Observations (or Sum Wgts)	8

Parameter Estimates

Term	Antilog(Estimate)	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.051	-1.293	0.017	-76.5	<.0001
Z-Score	3.472	0.541	0.009	58.6	<.0001

3. PUFA-IQ Benefits Assessment and Net Benefits Calculations

3.1 PUFA/Fish Consumption-IQ Benefits

The FDA appears to have done considerably more than simply adopt the summary findings from the literature to model the putative beneficial effects of fish consumption as an offset to the adverse effects of methylmercury in their net benefits analysis. They report:

> "We obtained ALSPAC summary data from 5,407 mother-child pairs that included maternal fish consumption and both verbal and full IQ in their children. These data show children's mean IQ scores along with the standard error of the mean at six levels of maternal fish consumption. The results were adjusted by the ALSPAC statisticians for cofounders such as maternal smoking and alcohol use during pregnancy, maternal education, and other factors. Details regarding these adjustments can be found in Hibbeln et al. (2007)."

The Hibbeln et al. (2007) study is not completely clear in its explanation of its methods for confounder adjustment. The summary of the paper says:

> "Multivariable logistic regression models including 28 potential confounders assessing social disadvantage, perinatal, and dietary items were used to compare developmental, behavioural, and cognitive outcomes of the children from age 6 months to 8 years in women consuming none, some (1–340 g per week), and >340 g per week."

The body of the paper reports,

> "We identified potential confounding variables by review of published data (table 1). Two continuous variables were used to assess the cumulative e! ects of adverse social and developmental factors during defined developmental periods: the family adversity index during pregnancy and a measure of parenting, one based on facilities for child care in the home[27] at 6 months of age. The family adversity index was calculated from the scores of 38 adverse factors (webpanel 2). Perinatal variables were birthweight (<2500 g; >=2500 g), and gestation at delivery (<37 weeks, $37 weeks). Additionally, 12 individual categorical covariates were included: sex of the child, age of the mother (<20 years or >=20 years), parity, highest maternal educational attainment, educational attainment (based on the UK examination system, and referred to qualifications that the parent might have obtained at school or at later ages,[28] catergorised as low, medium, or high), housing status (council [subsidized public housing], other rented, owned/mortgaged), crowding (%1 or >1 person per room), stressful life events at 18 weeks of gestation (upper 10%, lower 90% of cohort), had partner at time of birth (no, yes), smoking status during pregnancy (never, smoked before but not at 18 weeks of gestation, or still smoking at 18 weeks of gestation), alcohol use during pregnancy (non-drinker, drank before 18 weeks of gestation, still drinking alcohol at 18 weeks of gestation), breastfeeding (none, some) and ethnicity (white, black, Asian). Perinatal variables were low birthweight (<2500 g, !2500 g), and preterm delivery (<37 weeks, >=37 weeks).

> To investigate the possibility that seafood intake might be a marker for other differences in diet affected by social patterning, we also adjusted for 12 other food groups noted to be socially patterned in a previous analysis of diet in pregnancy in this cohort (sausages/burgers, pies/pasties, red meat, poultry, green leafy vegetables, other vegetables, salad, chips, fresh fruit, fruit juice, crisps, and biscuits).[15] All these 28 confounding variables (12 categorical and two continuous social variables, two perinatal variables, plus 12 dietary food groups) were used uniformly for all multivariable logistic regression analyses."

Hibbeln et al. (2007) are therefore certainly aware of the potential for social class and other correlates of IQ to distort their results. And it is clear from their Table 1 ["Characteristics of confounders used in the analyses by seafood intake groups (n=8916) with tests of the differences between the three groups"] (which is partly reproduced in this report's Table 1 on the next page) that greater fish consumption is associated with higher social class. Social class in turn is likely to be correlated with IQ via both genetic and environmental mechanisms. However their brief Lancet paper does not include any detailed presentation of the forms of the assumed relationships between the potential confounding variables and IQ or reasoning that documents the relative merits of different mathematical forms for predicting the effects of those confounders. The fact that all of the variables shown in Table 1 were used in 2- or 3-factor categorical variables rather than as continuous variables raises concerns because any categorization of this type throws away information that could help in the causal prediction of the dependent variable (IQ). Variance in the dependent variable that is not "explained" by the categorical form of the confounders or possibly misspecified relationships between the confounders and IQ measurements is available to be attributed by the regression model to another correlated independent variable (in this case, fish or fish-derived polyunsaturated fatty acid consumption).

To assess whether other researchers investigating influences of on neurodevelopmental impairment measured as IQ used apparently better techniques to control for confounders, three papers from the recent literature were examined—one relating to the postnatal effects of lead (Wasserman, 2001), one on effects of organophosphate exposure (Bouchard, 2011), and the other dealing with effects of exposure of 6-7 year old children to violence/trauma (Delaney-Black, 2002). In addition to continuous variables related to the quality of the home environment, one variable that seems to be routinely included in such studies is maternal or caregiver IQ (Wasserman, 2001), which was not apparently included as a confounder in the Hibbeln et al. (2007) study or a prior study by the same research group (Daniels, 2004).

One other issue that arises from an examination of the Daniels (2004) paper is the accuracy of the assessment of fish consumption from the questionnaire studies used by this group. The claim in Hibbeln is that their fish consumption questionnaire ascertainment is "validated" by biomarker measurements such as mercury levels in cord blood. There is indeed an indication in Daniels (2004) that mercury levels measured in cord blood are higher in the groups of women reporting some fish consumption than in the women reporting no fish consumption. However there is little difference in the cord blood mercury levels in groups reporting rather rare fish consumption (one

Table 1

Some[2] Confounders Used in the Hibbeln et al. Analyses by Seafood Intake Groups

Level of Potential Confounder	No Fish Consumption (N = 1059)	1-340 g Fish per week (N = 5770)	> 340 g per week (N = 2087)	P
Low Maternal Education	392 (37%)	1521 (26%)	329 (16%)	<0.0001
Middle Maternal Education	345 (33%)	2164 (38%)	727 (35%)	
High Maternal Education	322 (30%)	2085 (36%)	1031 (49%)	
Mortgage/Owned Housing	740 (70%)	4650 (81%)	1809 (89%)	<0.0001
Council Housing	167 (16%)	594 (10%)	121 (6%)	
Other Housing	152 (14%)	526 (9%)	157 (8%)	
<1 person/room (less crowding)	821 (78%)	4879 (85%)	1866 (89%)	<0.0001
1+ person/room (more crowding)	238 (22%)	891 (15%)	221 (11%)	
Non-smoking mother	643 (63%)	4063 (70%)	1612 (77%)	<0.0001
Ex-smoking mother	162 (15%)	751 (13%)	234 (11%)	
Current smoking mother	254 (24%)	956 (17%)	241 (12%)	
Non-drinking mother	173 (12%)	366 (6%)	88 (4%)	<0.0001
Mother stopped drinking	485 (46%)	2431 (42%)	873 (42%)	
Mother still drinking	451 (43%)	2973 (52%)	1126 (54%)	
Brest Feeding	766 (72%)	4564 (79%)	1821 (87%)	<0.0001
No Breast Feeding	293 (28%)	1206 (21%)	266 (13%)	

[2] Not included here are borderline significant relationships between fish consumption and "life events", significant relationships to parity and ethnic origin (white, black, asian), and a nonsignificant relationship to the gender of the child.

fish meal per two weeks) and those reporting considerably greater amounts (1-3 fish meals per week, and 4 or more meals per week).

In the FDA modeling, beneficial effects after adjustment for the Hibbeln (2007) study confounders were fit using two different saturating dose response functions (a Hill model and a hockey-stick model). (Two other dose response models were rejected for lack of fit.) Of these the Hill model is to be preferred *a priori* because it allows some nonlinearity in the effect at low doses, and incorporates a gradual, rather than an abrupt approach to an asymptote at high doses. However the distributional analysis underlying Tables AB-6 and V-5 in the main text reportedly gave equal weights to fits of the two dose response forms.

3.2 Net Benefits Calculations

The methodology here is a straightforward summation of the estimated adverse effects from the methylmercury exposure and the beneficial effects from either fish consumption, considered as a whole, or polyunsaturated fatty acids (PUFA) from fish in the species-specific analyses. Summary results for the population as a whole for 6-9 year olds (an age where full scale IQ can be measured) are presented in Table V-6. A key result is that through the 99.9th percentile of fish consumption the benefits of fish consumption are assessed to be greater than the harm from methylmercury exposure, leading to an expectation of population net benefits. This, of course, depends on the assumption that the estimates of IQ benefits from different amounts of fish consumption are free from distortion by either under-control of correlated confounders that are more closely causally related to children's IQ such as the genetic and environmental influences arising from maternal IQ. In further work, it might be helpful to elicit a range of estimates of the probable magnitude and uncertainty in this effect from knowledgeable experts in the measurement of the effects of different prenatal factors on IQ.

Similar results for the consumption of various amounts of specific seafood species are presented in an extended series of tables near the end of section V of the FDA report (Tables V-12 through V-18 on pp. 106-126). Finally, detailed net benefit results under different policy scenarios hypothesizing different restrictions on fish consumption and different calculation assumptions are explored in Appendix B, Tables AB-8 through AB-15 on pages 175-180. Unfortunately there does not appear to be a succinct summary of policy-relevant conclusions from these analyses.

4. Summary of Conclusions and Recommendations

Briefly, the major conclusion of this work is that the FDA analysis represents an important and honest effort to elucidate the relative neurodevelopmental benefits and risks from the consumption of different types of seafood by reproductive-age women. Benefits and risks from potential effects of methylmercury-containing seafood for cardiovascular disease at later ages will need to be part of a different analysis. Major conclusions and suggestions for improvement of the neurodevelopmental analysis are:

- The overall FDA finding that maternal seafood consumption is net-beneficial for neurodevelopment (as measured by IQ and some other indices) over nearly the entire range of population seafood exposures rests on the adequacy of control of confounders associated with higher social class in analyses of observations of a single population. Absence of specific statistical control of the genetic and environmental effects of maternal IQ, and the expression of many confounders as discrete categories, rather than continuous variables produces uncertainties that are difficult to quantitatively analyze. The most rapid way to quantitatively estimate these may be by way of an expert elicitation by professionals with relevant experience in the determinants of IQ. Preferably these experts should have access to the underlying epidemiological data, so that effects of different model forms for the available confounders can be considered.

- The FDA estimates of the adverse effects of methylmercury from fish consuming populations do not appear to have been corrected for the presence of the hypothesized beneficial effects from fish polyunsaturated fatty acids in the studied groups. This can and should be remedied in further work. Such a correction can be expected to increase the estimates of the separated adverse neurodevelopmental effects from prenatal methylmercury exposure.

5. References

-

Axelrad, D., Bellinger, DC, Ryan, LM, Woodruff, TJ, 2007. Dose-Response Relationship of Prenatal Mercury Exposure and IQ: An Integrative Analysis of Epidemiological Data. Environmental Health Perspectives. 115**,** 609-615.

Bogan, K., Spear, RC, 1987. Integrating Uncertainty and Interindividual Variability in Environmental Risk Assessment. Risk Analysis. 7**,** 427-436.

Bouchard, M., Chevrier, J, Harley, KG, Kogut, K, Vedar, M, Calderon, N, Trujillo, C, Johnson, C, Bradman, A, Barr, DB, Eskenazi, B, 2011. Prenatal Exposure to Organophosphate Pesticides and IQ in 7-Year-Old Children. Environmental Health Perspectives. 119**,** 1189-1195.

Budtz-Jorgensen, E., Debes, F, Weihe, P, Grandjean, P, Adverse mercury effects in 7 year-old children expressed as loss in "IQ." Final report for the U.S. EPA., 2005.

Budtz-Jorgensen, E., Grandjean, P, Weihe, P, 2007. Separation of risks and benefits of seafood intake. Environmental Health Perspectives. 115**,** 323-327.

Carrington, C., Bolger, PM, 2000. A Pooled Analysis of the Iraqi and Seychelles Methylmercury Studies. Human and Ecological Risk Assessment. 6**,** 323-340.

Cohen, J. T., et al., 2005. A quantitative analysis of prenatal methyl mercury exposure and cognitive development. Am J Prev Med. 29**,** 353-65.

Crump, K., Kjellstrom, T, Shipp, AM, Silvers, A, Stewart, A, 1998. Influence of Prenatal Mercury Exposure Upon Scholastic and Psychological Test Performance: Benchmark Analysis of a New Zealand Cohort. Risk Analysis. 18**,** 701-713.

Daniels, J., Longnecker, MP, Rowland, AS, Golding, J, and the ALSPAC Study Team--University of Bristol Institute of Child Health, 2004. Fish Intake During Pregnancy and Early Cognitive Development of Offspring. Epidemiology. 15**,** 394-402.

Delaney-Black, V., Covington, C, Ondesma, SJ, Norsrom-Klee, B, Templin, T, Ager, J, Janisse, J, Sokol, MD, 2002. Violence Exposure, Trauma, and IQ and/or Reading Deficits Among Urban Children. Arch Pediatr Adolesc Med. 156**,** 280-285.

Grandjean, P., Weihe, P, White, R, Debes, F, Araki, S, Yokoyama, K, MUrata, K, sorensen, N, Dahl, R, Jorgensen, P, 1997. Cognitive deficit in 7-year-old children with prenatal exposure to methylmercury. Nerotoxicology and Teratology. 19**,** 417-428.

Hattis, D., The Conception of Variability in Risk Analyses--Developments Since 1980. In: M. a. Small, (Ed.), Risk Analysis and Society in the 21st Century: An Interdisciplinary Characterization of the Field. Cambridge University Press, Cambridge, 2003, pp. 15-45.

Hattis, D., Burmaster, DE, 1994. Assessment of Variability and Uncertainty Distributions for Practical Risk Analyses. Risk Analysis. 14**,** 713-730.

Hattis, D., Erdreich, L, Ballew, M, 1987. Human Variability in Susceptibility to Toxic Chemicals--A Preliminary Analysis of Pharmacokineic Data from Normal Volunteers. Risk Analysis. 7**,** 415-426.

Hattis, D., Lynch MK, Empirically Observed Distributions of Pharmacokinetic and Pharmacodynamic Variability in Humans—Implications for the Derivation of Single Point Component Uncertainty Factors Providing Equivalent Protection as Existing RfDs. In: J. L. a. E. Ohanian, (Ed.), Toxicokinetics in Risk Assessment. Informa Healthcare USA, Inc., 2007, pp. 69-93.

Hibbeln, J. R., et al., 2007. Maternal seafood consumption in pregnancy and neurodevelopmental outcomes in childhood (ALSPAC study): an observational cohort study. Lancet. 369**,** 578-85.

IOM, C. o. E. o. t. S. o. F. P., 1991. Seafood Safety. National Academy Press, Washington, DC.

Myers, G., Davidson, PW, Cox, C, Shamlaye, DF, Palumbo, D, Cernichiari, E, Sloane-Reeves, J, Wilding, GE, Kost, J, Huang, L, Clarkson, TW, 2003. Prenatal methylmercury exposure from ocean fish consumption in the Seychelles child development study. Lancet. 361**,** 1686-1692.

Shlyakhter, A. I., Kammen, DM, 1992. Sea-level rise or fall? Nature. 253**,** 25.

Stern, A. H., Korn, L. R., 2011. An Approach for Quantitatively Balancing Methylmercury Risk and Omega-3 Benefit in Fish Consumption Advisories. Environmental Health Perspectives. 119**,** 1043-6.

Wasserman, G., Factor-Litvak, P, 2001. Methodology, Inference and Causation: Environmental Lead Exposure and Childhood Intelligence. Archives of Clinical Neuropsychology. 16**,** 343-352.