A Pooled Analysis of the Iraqi and Seychelles Methylmercury Studies

CD Carrington and PM Bolger
U.S. Food and Drug Administration
200 C St SW HFS-308
Washington, DC 20204

Abstract

Several epidemiology studies have investigated the impact of maternal exposure to methylmercury (MeHg) on childhood development of the central nervous system (CNS). In the present report, data from the Iraqi episode that occurred in 1970 from contaminated grain are integrated with those from a more recent study of a population with a high fish intake in the Seychelles Islands. The latter study had many more subjects whose mercury hair levels that were much lower and more representative of levels typically found in consumers whose MeHg exposure is from fish. The age of onset of talking (AOT), the age of onset of walking (AOW) and a combined measure (CM) that integrated the two were used as common scales of MeHg effect for the two studies. The first step of the analyses involved the construction of separate two-dimensional cumulative frequency tables for each study for different groups spanning the range of hair levels and observed effect for each measure. Models were then fit to the values in the tables that were constructed from four components: 1) A dose-effect function that related hair MeHg to the effect measure; 2) a frequency distribution describing population variability; 3) parameters to represent dose-independent influences on effect; and 4) parameters to represent study dependent influences on effect. When the four submodels were assembled, a series of 1092 candidate models resulted which contained 3 to 7 parameters (e.g. slope, standard-deviation, dose-independent age of talking) whose value could be adjusted to improve the fit. After optimizing the fit of each model, a weighting algorithm that rewards for fit and penalizes for the number of parameters in the model was used to identify the best 200 models. The same algorithm was then used to assign a probability to each model in a probability tree. A two-dimensional Monte-Carlo simulation using the resulting function in combination with exposure values typical of U.S. consumers yielded predicted delays in AOT, AOW, and CM attributable to fish consumption in a variable and uncertain range of 0.000 to 1 day.

Introduction

For many environmental contaminants, The relevant health effects information for a public health problem often does not come from a single study. The consideration of the potential harm to a developing fetus that may result from maternal exposure to methylmercury is a prime example. Several epidemiology studies have been directed towards ascertaining the impact of maternal exposure to methylmercury on childhood development of the central nervous system (CNS). These studies have used maternal hair levels as a marker for exposure to methylmercury, but the observations on health effects collected in each study have varied.

Data

Sources

The concern for exposure to mercury is primarily a result of two poisoning epidemics that occurred in Japan and Iraq. The latter epidemic, that occurred after exposure to contaminated grain, was the subject of an extensive epidemiological investigation that included an effort to relate the magnitude of exposure to methylmercury to health impact (Marsh et al, 1987). Because these were not prospective studies, the reports concerned with the Iraqi episode do not reflect the same degree of experimental control as subsequent studies. For risk assessment purposes, perhaps the major shortcoming of the Iraqi study is the presence of relatively few individuals at low doses. For instance, because there is little data on the extent of normal variation for the observed measures of development, it is difficult to discern whether a slightly higher frequency of “abnormal” responses (e.g. delayed walking) is attributable to mercury effects or normal variation. However, in spite of numerous shortcomings, the Iraqi study has a major advantage over more recent reports – there were high-dose health effects that were unequivocally attributable to methylmercury exposure.

More recent prospective studies have searched for health effects of methylmercury in populations consuming whale or fish with much lower levels of exposure than those encountered in Iraq (Kjellstrom, et al, 1986; Marsh et al, 1995a , Grandjean et al, 1997). For the present analysis, the results of the Iraqi study are combined with a more recent study in the Seychelles Islands (Marsh et al, 1995a), where the exposure to methylmercury is from the consumption of marine fish. Data from the Seychelles study were used because of the presence of some of the same measurements as those collected from Iraq and because the individual subject data was made available to us. The Seychelles study has many more individual subjects and the range of mercury hair levels were much lower and more representative of levels typically found in consumers of fish, but which are still much higher than those typical of infrequent consumers of fish.

Response Measures

To combine results from two or more studies in an analysis, it is necessary that there be a common measure. For the present analysis, two endpoints that were collected in both the Iraqi and Seychelles studies were used as the common measure: 1) Age of Talking (AOT) – the age at which the infants started talking, and 2) Age of Walking (AOW) -- the age at which the infants became toddlers. Not only were these measures available from both studies, they have the advantage of being simple measures of neurological development. In addition, as a measure of a general developmental delay, the measures were averaged to produce a third combined measure (CM), which may be thought of as a generalized developmental measure.

Construction of Cumulative Frequency Tables

The data were used to construct separate (one for each study) two dimensional cumulative frequency tables for each study which tabulated frequency for groups spanning the range of hair levels and observed response. These were constructed by grouping the subjects from each study by dose, and calculating the frequency at which each of series of response levels were exceeded. Tables were then constructed for AOT, AOW, and the combined measure (CM) that averaged the two from each individual. Plots of cumulative frequency tables for both the Iraqi and Seychelles studies for each endpoint are shown in Figures 2-4.

Modeling

Comparative Modeling

The analysis presented here is an exercise in comparative modeling where a large number of alternative mathematical models are examined with respect to their ability to describe historical data. Analyses of epidemiological data often undertake evaluation that are designed to identify which of a number of different parameters (e.g. confounding variables or modifying factors) are to be included in a final model. The present analysis differs in two important respects. First, it evaluates models that are different in form rather than just complexity. Second, rather than concluding the analysis with a final or best model, a probability tree that employs probabilities for a set of alternative models to characterize the uncertainty associated with an estimation.

To conduct a comparative modeling exercise, the first step is to assemble a list of candidate models. Dose-response models often have multiple sources of theoretical uncertainty. These include the dose-response relationship itself, the influence of factors other than dose on the outcome, and the extent of the variability among individual subjects. In addition, when multiple studies are being used to evaluate the models, it may be desirable to accommodate differences in the studies within the model. As a result, models were formulated from four submodels, each of which had several theoretical alternatives. Each of the four submodels represent a potential source of model uncertainty: 1) A dose-response function (relating hair level to AOT, AOW, or CM)); 2) a statistical distribution describing population variability; 3) dose-independent factors; and 4) study dependent factors. With several variations of the mathematical form (see Table 1) and relative position of each of the submodels (Figure 1), a series of 1092 candidate models were assembled. All the models were relatively simple and contained 3 to 7 adjustable parameters (*e.g.* slope, standard deviation, dose-independent AOT) which could be altered to improve the fit.

As an example of what the dose response equations looked like when assembled, a model that fit the data well was constructed from a linear dose response function, a background response parameter, a background study parameter, and a Weibull distribution to account for population variability at position 4. To predict cumulative frequency as a function of dose and response, this yielded the following function:

$$= 1 - \exp((-\text{Response} / (\text{Dose}*P_1 + P_3 + P_4) / ((\log 2) \wedge (1/P_2))) \wedge P_2)$$

To predict response as a function of dose and frequency, the following function was used:

$$= (\text{Dose}*P_1 + P_3 + P_4) / ((\log 2) \wedge (1/P_2)) (\log(1/(1- \text{Frequency})]) \wedge (1/ P_2)$$

where

P_1 is the dose-response slope

P_2 is the Weibull alpha parameter

P_3 is the background response (*i.e.* age in months)

P_4 is a study-dependent background term (also age in months)

Software

The analysis was conducted in Microsoft Excel using procedures written in Visual Basic for Applications, which are available on request.

Goodness-of-Fit

Fitness was judged by a composite least residual squares measure that gave equal weight to residuals for predicted population percentiles (frequency as a function of dose and response) and for predicted magnitude of effect (response as a function of frequency). The fit for each dose-group was weighted by the original number of observations – which gave the values from the Seychelles considerably more weight in the low dose regions.

Optimization

The parameters were adjusted to fit the data (minimize the measure of fit) with Excel Solver. Simple equations were used to assign initial estimates for the parameters – some of these used information from the study such as the range of doses and responses. If an obviously poor fit was obtained, different initial estimates were used in order to find a better fit – usually by adopting estimates from simpler models with the same parameters that produced a better fit.

Model Weighting and Model Uncertainty

The models were judged with an algorithm that rewards a model for goodness-of-fit and penalizes for the use of extra parameters:

$$\text{Model Weight} = (((1 + n / Pn)^{O}) * ((1 - gof)^{H})$$

where

n = number of observations

Pn = Number of Model Parameters

gof = Goodness-of-Fit

O = The Parameter Penalty, an arbitrary constant that determines the relative importance of model simplicity

H = The Association factor, an arbitrary constant that determines the relative importance of goodness-of-fit.

In the present analysis, values of 0.3 and 100 were used for O and H, respectively. These values were chosen because they appeared to generate a reasonable balance between fit and model simplicity (see Carrington, 1996 for further discussion of this approach). The uncertainty associated with the predictions made was represented by weighting the 200 best models. The algorithm used for model weighting was also used to select the best models.

Two of the dose-response models employed have a biochemical heritage – the Mass Action model is an equation that is able to describe reversible (ionic) competitive ligand-receptor binding interactions. The first order equation is a function that describes irreversible (covalent) ligand-receptor interactions. Evidence that methylmercury acts by either of these mechanisms could be construed as an increase in the weight (and probability) accorded theories that employ those functions. However, it should be noted that even if a particular biochemical mechanism of action is conclusively established, the in vivo reaction will often be vastly more complicated that the in vitro one (Tallarida and Jacob, 1979). As a result, a model reflecting the wrong mechanism, or no mechanism, may still describe the data and still make a better prediction. Although it would be possible to include theoretical support for a theory in the calculation of each models evidential weight, the biochemical mechanism for methymercury is presently unknown.

Results

Age of Talking. For the AOT endpoint, the best model was comprised of a linear dose-response relationship, a Weibull population distribution, and a background response parameter, and a study-dependent dose parameter (see Figure 1). The exponential, hockey stick, and mass action dose-response relations were also heavily represented among the top-rated models (seeTable 2). The first-order and logistic models tended to not fit as well. The Weibull distribution was clearly the best fitting population distribution - regardless of the dose-response function used. The lognormal distribution consistently provided a better fit than the other two distributions. The poorer fit with either the normal or logistic distribution functions is indicative of a skewed distribution. All the top rated models included parameters for both dose-independent and study-dependent effects, reflecting the notions that a) children do not speak at age 0, and that there are differences in the Iraqi and Seychelles studies that are not attributable to methylmercury.

Age of Walking. For the AOW endpoint, the best model was comprised of a linear dose-response relationship, a Weibull population distribution, a background dose and response parameters, and a study-dependent dose parameter (see Figure 3). All the dose-response

functions were represented among the top-rated models. The Weibull and lognormal distributions were again the clear favorites for modeling population variability. All the top rated models included two parameters for dose-independent effects, reflecting the notion that children do not speak at age 0. All of the best models also included a study dependent parameter, again reflecting differences in the Iraqi and Seychelles populations.

Combined Measure. For the combined measure, the best model was comprised of a linear dose-response relationship, a Weibull population distribution, and a background dose and response parameters, and a study-dependent response parameter (see Figure 4). All the dose-response functions were represented among the top-rated models. The Weibull and lognormal distributions were again the clear favorites for modeling population variability. All the top rated models included parameters for both dose-independent and study-dependent effects.

Function Output. The output of the best model for each of the three endpoints is plotted in Figures 2, 3, and 4 -- for both the Iraqi and Seychelles studies. Probability trees comprised of the top 200 models yield an uncertainty distributions when used as a predictive tool. Sample output from a function that weights the frequency of use of the best 200 models is given in Tables 3-5. In a two-dimensional Monte-Carlo simulation used to simulate both variability and uncertainty, this function will impact the distribution in both dimensions.

Because the models contain study-dependent variables, the study for which a prediction is required must be specified. If the resulting models are to be used in a risk assessment, this requires a decision about which study population is more representative of the population of concern to the assessment. This decision would revolve around speculation about the source of the differences between the studies (e.g. cultural or genetic), and would be a source of both variability and uncertainty. For instance, the population of concern may be variable with regard to the percentage of the population for which each study is more appropriate, while the extent of that frequency for each may be uncertain.

Discussion

Comparative Modeling and Model Uncertainty

The present analysis differs from other analyses that have been conducted with the same data in several respects. The most important is that the analysis is designed to be part of a decision paradigm that includes scientific uncertainties that are presently unresolvable (Evans et al, 1994; Carrington, 1996). Other analyses of the same data have proceeded with a single model (Crump et al, 1995) or have examined two alternatives (Cox et al, 1995). In contrast, the present analysis began with over a thousand potential mathematical models.

Since the results of an analysis can be highly dependent on model choice, representing uncertainty arising from model selection can greatly impact the range of plausible interpretation portrayed by the analysis. Since it is generally not possible to conclude from such an analysis that one model must be preferred to the exclusion of all others, a probability tree that distributes the use of the models in making uncertain predictions can be used to integrate model uncertainty into an analysis. The present analysis used two hundred of the initial candidate models to characterize the range of plausible interpretations of the data.

It should be noted that the uncertainties described in this analysis reflect only model uncertainty. Statistical characterizations of uncertainty, such as those used by Cox et al (1989; 1995) and Crump et al (1995) reflect error arising from known (e.g. measurement error) or presumed (e.g. sampling error) characteristics of a larger sample from which the data is drawn. Since it is these presumptions that are often called into question by different investigators, an analysis of model uncertainty operates at a more basic level that must precede statistical characterization of uncertainty. Were a description of the potential effects of sampling error included in the analysis, it would be expected that the distribution of predicted effects would be somewhat broader and less discontinuous.

Dose-Response Models

Another major difference is that the models in the present analysis compared are somewhat more complex than those used by either Cox et al (1989; 1995) or Crump et al (1995). First, the models used here describe both the dose-response relationship for an individual and population variability. In contrast, the Cox et al (1989) and benchmark dose analyses (e.g. Crump et al, 1995) used models that relate dose to the population frequency (variability) for a discrete event. The analysis of the Iraqi data by Cox et al (1989) begins by asserting that walking or talking after a particular age constitutes an abnormal response -- with the implicit result that all values above that point are equivalent and all values below that point are equivalent. Similarly, Crump et al (1995) define an abnormal response as a two standard deviations from the mean. In contrast, the analysis by Myers et al (1997) examined the relationships for maternal hair mercury versus continuous individual measures, but did not model population variability. Using models that maintain both the individual measure and population variability yields more information to a decision maker.

Confounding Factors

In addition to accounting for the dose-response relationship, the models used in the present analysis provide a cursory account of dose-independent influences on the outcome measure and possible differences between studies. In contrast, dose-independent effects are often treated as parameters that must be estimated or presumed before the dose-response analysis, rather than as an integral part of the analysis. The analysis was also designed to acknowledge the possibility that there are unknown differences between the Iraqi and Seychelles studies that may influence the outcomes measured. This was a problem that arose because the two studies were analyzed together.

However, the population component of the models used in the present analysis are relatively simple compared to those commonly used in epidemiological studies that may include many potential confounding variables or modifying factors, each of which contributes to the overall population variability. Although epidemiological studies often do compare models that differ in the parameters they contain, they usually do not

compare models that differ in form. This is perhaps attributable to the difficulty in solving equations that do not presume linear relationships with normally distributed variances. For example, the models employed in the present exercise could perhaps be improved by having separate descriptions of the variability contributed by dose-dependent and dose-independent influences – if the equations could be solved.

Risk Management

This analysis is designed to be part of a risk assessment/risk management decision paradigm where the goal of the assessment is to make a statement of the present state of knowledge – including the attendant uncertainties. As a result, the analysis does not attempt to dictate the impact of the uncertainties on any subsequent decision that may be made. For example, using an outer bound as a criteria for a credible statement can often be used in two directions. First, in some conclusions drawn by Cox et al (1997), the emphasis is on whether there is significant support for saying that there is an effect. On the other hand, the benchmark dose calculation (Crump et al;1995) shifts the burden the other way by using an outer bound that reflects relative surety that the effect is less than a certain value. However, neither of these approaches relies entirely on an outer bound in drawing conclusions -- they also report central values.

The present analysis is more generally applicable because it leaves choices involving acceptable degrees of harm, population variability, and uncertainty for a later step in the decision process (i.e. risk management). Comparisons to other analyses may still be made by applying similar decision rules. For instance, a benchmark corresponding to a 5% increase in the frequency of the Iraqi population with an AOT in excess of 24 months can be calculated from the functions generated in the present analysis -- the median estimate is 39 ppm, with a plausible range (0.05 to 0.95) of 26 to76 ppm. Similarly, for an AOW in excess of 18 months, the median estimate is 30 ppm, with a plausible range of 15 to 157 ppm.

The uncertainty described in tables 3-5 and the benchmark calculations above reflect only model uncertainty. Other sources of uncertainty were not included because the analysis

is already technically complex, and we preferred to emphasize model uncertainty as it is too often neglected. The most straightforward way to introduce consideration of parameter uncertainty would be to use a bootstrapping procedure prior to the generation of the cumulative incidence tables. The bootstraps could also incorporate any a priori doubts concerning the accuracy of the measurements. For instance, there are reasons to doubt the accuracy of both the dose and response measurements from the Iraqi study (Cox et al, 1989).

A novel aspect of the present analysis is that it includes a 'study' variable to account for unexplained differences among studies. This allows some commonality to be found among studies without forcing a model to assume that the two populations are entirely equivalent. This does not resolve the problem of having to decide the relevance of each of the studies, but it does postpone the decision and allow it to be considered after the analysis. This may be beneficial if the relevance of the study varies with the population of concern – it may not be possible to anticipate whether any differences in the studies represents variability or uncertainty.

Data and Measures

The present analysis integrates results from the Iraqi or Seychelles studies. It would perhaps be better still to include other studies in the analysis as well. In particular, there are other studies of maternal exposure to mercury and childhood development that have been conducted since the Iraqi episode in Canada (McKeown-Eyssen et al., 1983), New Zealand (Kjellstrom et al., 1986), Peru (Marsh et al, 1995b) and the Faroe Islands (Grandjean et al., 1997). There are two obstacles that must be overcome to accomplish this. The first is to accumulate the data from all the individual investigators. The availability of the data was a major influence on the choice of studies used in the present analysis.

A second obstacle is to devise a common measure that would allow all the studies to be placed on the same scale. Although there is considerable overlap in the tests conducted in each of the studies, there are differences as well. The biggest difference, however, is

between the Iraqi study and the more recent studies. To some extent the studies aren't comparable because the simply aren't measuring the same thing. Nonetheless, at the very least it should be possible to devise a scale that reflects some abstract notion, like "development time" that incorporates whatever information is provided by a particular study. Like the measure of intelligence with the IQ, such a measure would not precisely gauge every aspect of neurological and behavioral development that anyone may care about. But, as it can serve as the basis for further discussion, any measure is preferable to none. Should it become apparent that a measure is combining two very different phenomena, a more refined measure could be devised. For example, separate scales for motor and sensory development might be preferable.

Conclusions

If an analysis is defined by the context in which the analysis is to be used, there are no limits to how studies may be combined. If there is reason to believe that a particular study has some bearing on a particular inference such as a dose-response relationship then the model used to draw the inference may be altered to reflect this. The main obstacle in accomplishing may not lie in devising the model. Rather, the problem may be that there are numerous models that may reasonably be employed for the purpose. Our solution to this problem is to examine as many of the alternatives as we can, using the evidence provided by the data to discount as many as we can. Since we cannot show that a single theory is clearly preferable to all the others, the range of predictions yielded by the plausible remainder constitutes our uncertainty about the inference of a dose-response relationship.

Taken as a whole, the predictions made by the models suggest a small effect on the AOT, AOW, and CM that is a very small percentage of the normal variation in these measures. Based on this analysis, the magnitude of the effect would be expected to be undetectable in even a large prospective epidemiology study. Whether or not the effect is large enough to merit some consideration in regulating personal or institutional behavior is another matter, that we leave to be discussed elsewhere.

References

Carrington, C.D. (1996). Logical Probability and Risk Assessment. *Hum Ecol Risk Assess 2*:62-78.

Cox, C., Clarkson, T.W., Marsh,D.O., Amin-Zaki, L., Tikriti, S., and Myers, G. (1989). Dose-response analysis of infants prenatally exposed to methylmercury: An application of a single compartment model to single strand hair analysis. *Environ Res 49*:318-332.

Cox, C., Marsh, D.O., Myers, G.J., and Clarkson, T.W. (1995). Analysis of data on delayed development from the 1971-1972 outbreak of methylmercury poisoning in Iraq: Assessment of influential points. *Neurotoxocol 16*:727-730.

Crump, K., Viren, J., Silvers, A., Clewell, III H., Gearhart, J., and Shipp, A. (1995). Reanalysis of dose-response data from the Iraqi methylmercury poisoning episode. *Risk Anal 15*:523-532.

Evans, M. Hastings, N., and Peacock, B. (1993). *Statistical Distributions, 2nd Ed*. John Wiley & Sons, New York.

Grandjean, P., Weihe, P. White, R., Debes, F., Araki, S., Yokoyama, K., Murata, K., Sorensen, N., Dahl, R. and Jorgensen, P. (1997). Cognitive deficit in 7-year-old children with prenatal exposure to methylmercury. *Neurotox. Terat. 19*:417-428.

Kjellstrom, T., Kennedy, P., Wallis, S. and Mantell, C., Physical and mental development of children with prenatal exposure to mercury from fish. Stage 1: preliminary tests at age 4. National Swedish Environmetnal Protection Board Report 3080, 1986

Marsh, D.O., Clarkson, T.W., Cox,C., Myers, G.J., Amin-Zaki, L., and Al-Tikriti ,S. (1987). Fetal methylmercury poisoning. Comparison of mercury levels in maternal blood, fetal cord blood, and placental tissues. *Arch. Neurol. 44*:1017-1022.

Marsh, D.O., Clarkson, T.W., Myers, G.J., Davidson, P.W., Cox, C., Cernichiari, E., Tanner, M.A., Lednar, W., Shamlaye, C.F., Choisy, O., Hoareau, C., and Berlin, M. (1995a). The Seychelles Study of Fetal Methylmercury Exposure and Child Development. *Neurotoxicol 16*:583-596.

Marsh, D.O., Turner, M.D., Smith, J.C., Perez, V.M.H., Allen, P., and Richdale N (1995b). Fetal MeHg study in a Peruvian fish eating population. *Neurotoxicol 16*:717-726.

McKeown-Eyssen G.E., Ruedy J., and Neims, A. (1983). Methylmercury exposure in Northern Quebec. II. Neurological findings in children. Am. J. Epidemiol. 118:470-479.

Myers, G.J., Davidson ,P.W., Shamlaye, C.F., Axtell, C.D., Cernichiari, E, Choisy, O., Choi, A., Cox ,C., and Clarkson, T.W. (1997). Effects of prenatal methylmercury exposure from a high fish diet on developmental milestones in the Seychelles child development study. *Neurotoxicol 18*:819-830.

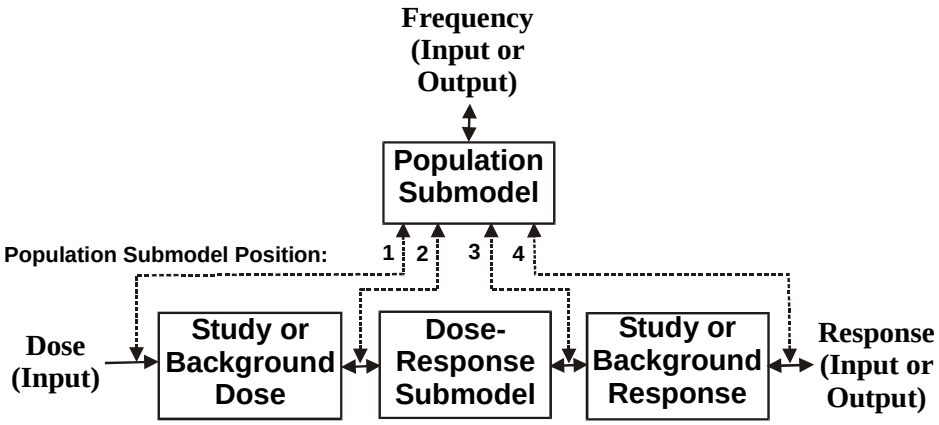
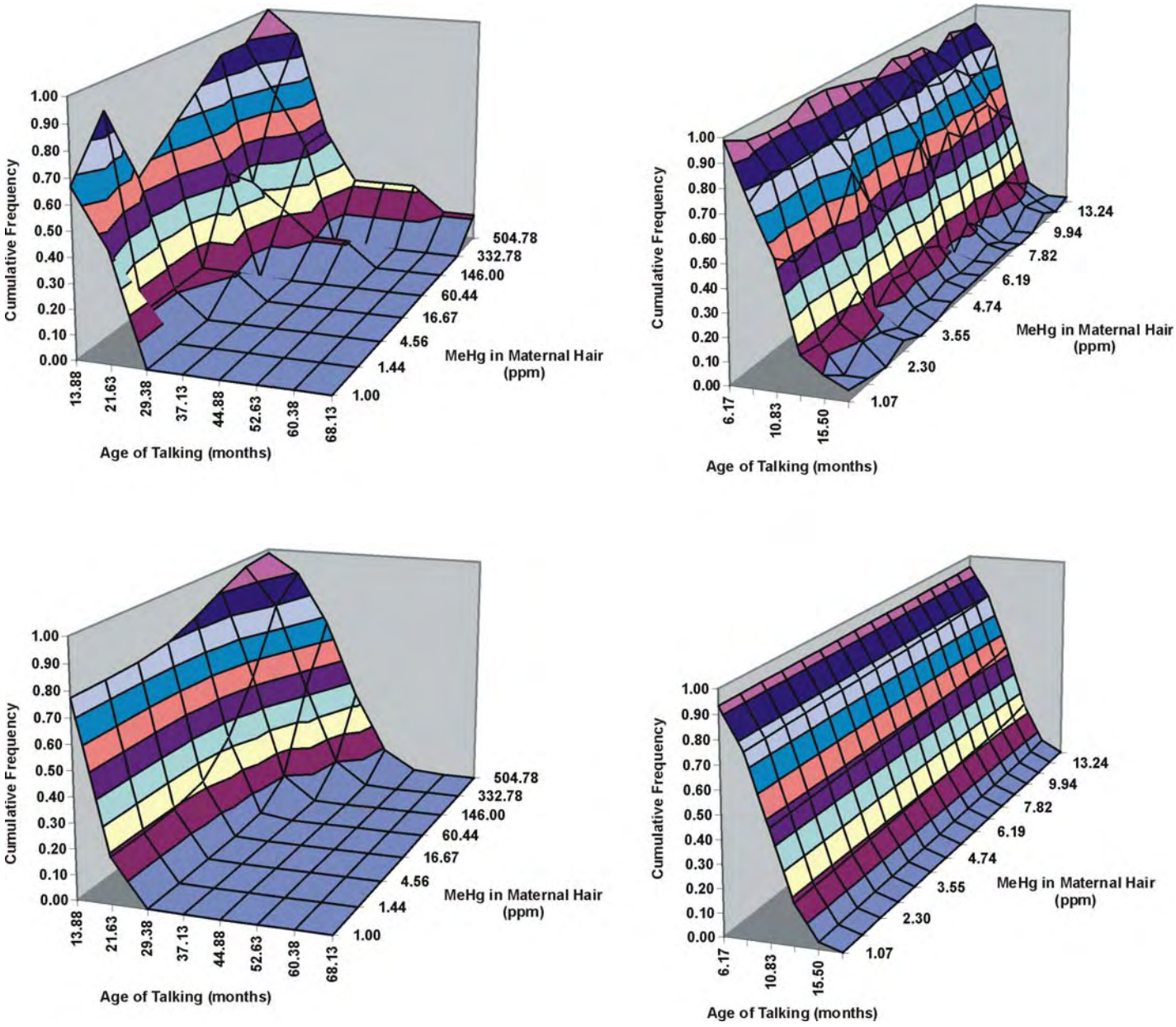


Figure 1: **Model Assembly from Four Components**

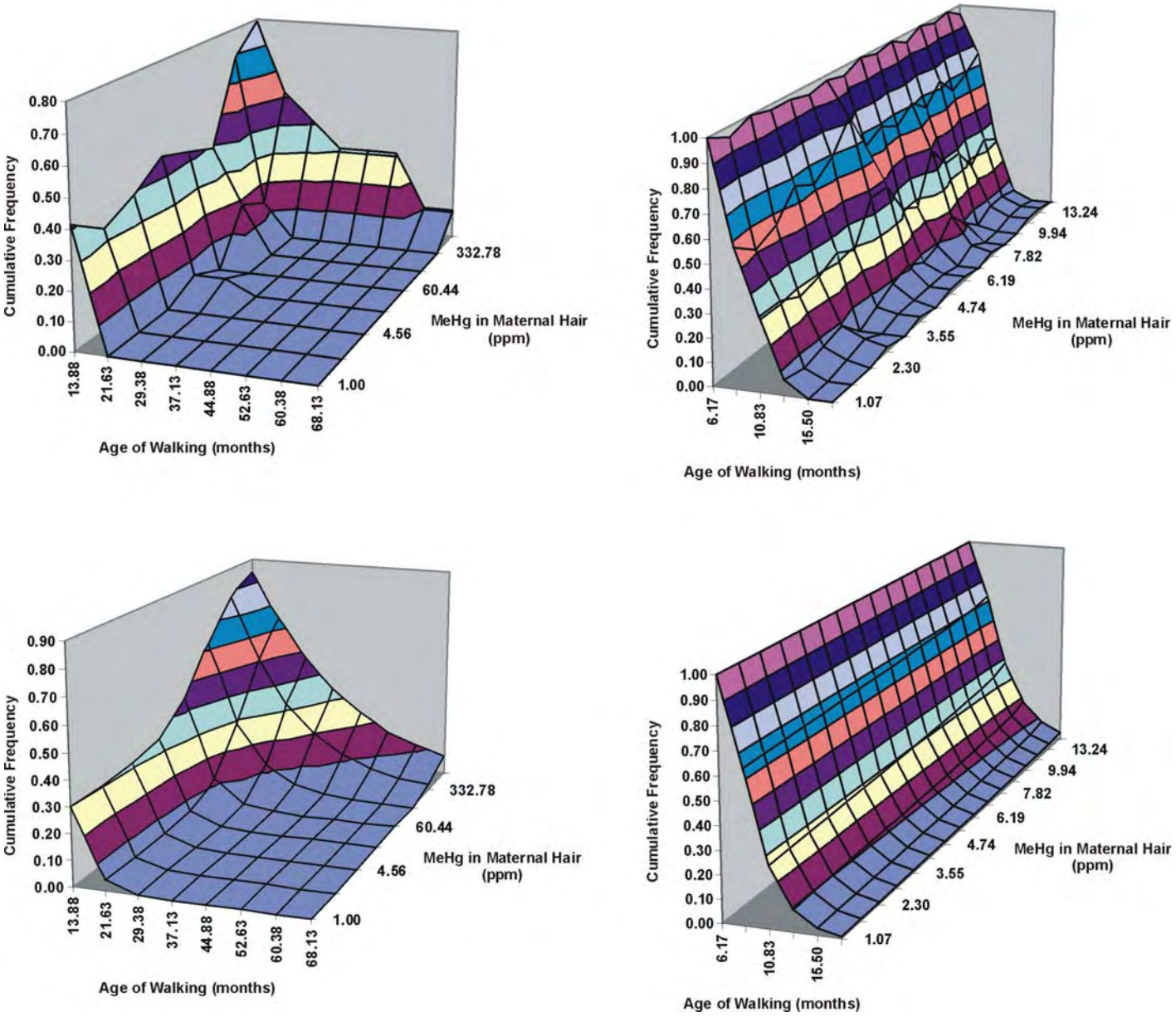
General structure of the models used to integrate the results from the Iraqi and Seychelles studies. The central component is the dose-response function that relates dose to the magnitude of an individual outcome (i.e. AOT, AOW, or CM). The background and study functions add parameters to account for dose-independent influences that are study-independent or study-dependent. The population submodel converts the individual model into a population model by introducing a statistical distribution at one of four positions in the individual model.

Figure 2: Age of Talking, Data and a Model



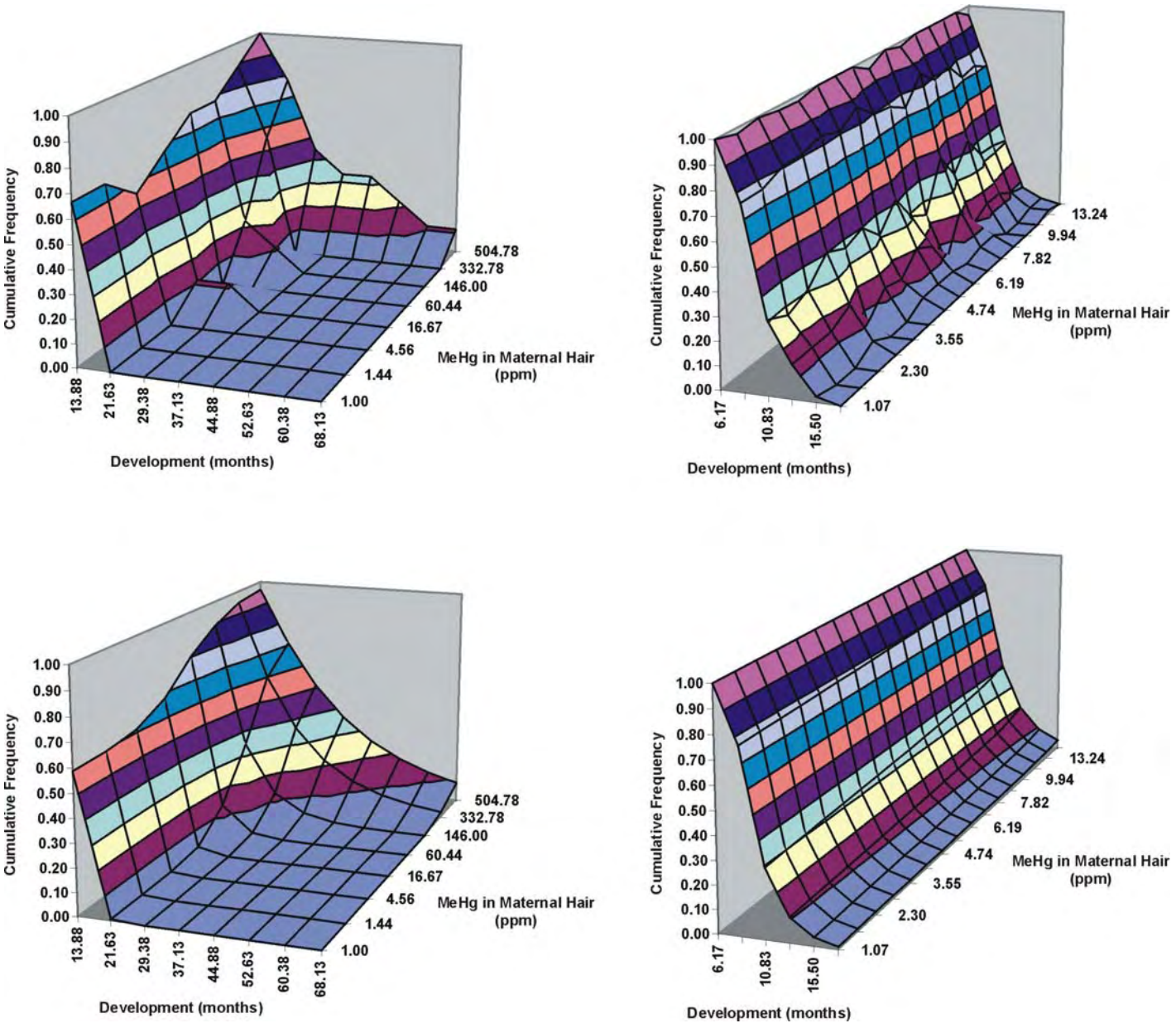
The charts at the top reflect the cumulative incidence tables constructed from the raw data (Age of Talking) from the Iraqi (left) and Seychelles (right). The charts at the bottom reflect a common model fit to both data sets. The Z-axis reflects the percent of the population in each dose group (Y-axis) with a AOT above the X-axis value. The X-axis values are chosen to represent the range of value encountered in the studies, and therefore do not necessarily generate incidences of 0 or 1 at all dose groups

Figure 3: Age of Walking, Data and a Model



The charts at the top reflect the cumulative incidence tables constructed from the raw data (Age of Walking) from the Iraqi (left) and Seychelles (right). The charts at the bottom reflect a common model fit to both data sets. The Z-axis reflects the percent of the population in each dose group (Y-axis) with a AOW above the X-axis value. The X-axis values are chosen to represent the range of value encountered in the studies, and therefore do not necessarily generate incidences of 0 or 1 at all dose groups.

Figure 4: Combined measure, Data and a Model



The charts at the top reflect the cumulative incidence tables constructed from the raw data (Combined Measure) from the Iraqi (left) and Seychelles (right). The charts at the bottom reflect a common model fit to both data sets. The Z-axis reflects the percent of the population in each dose group (Y-axis) with a Development score above the X-axis value. The X-axis values are chosen to represent the range of value encountered in the studies, and therefore do not necessarily generate incidences of 0 or 1 at all dose groups

Table 1: Functions Used to Construct Models of Methylmercury Effects

Submodel	*Functions*
Dose vs Individual Response	Linear, Hockey Stick, Mass Action, First Order, Exponential, Logistic
Population Variability	Normal, Lognormal, Weibull, Logistic
Dose Independent Factors	None, Background Dose, Background Effect, Background Dose and Background Effect
Study Factors	None, Study Dose, Study Effect, Study Dose and Study Effect

Table 2: The Top Twenty Models for the Age of Talking Endpoint

Response Submodel	*Population Submodel*	*Position*	*Background Submodel*	*Study Submodel*	*Fit*	*n*	*Weight*	*Map Value*
Linear	Weibull	4	Response	Dose	0.0078	4	1.3951	0.0064
Linear	Weibull	3	Response	Response	0.0078	4	1.3951	0.0128
Exponential	Weibull	4	Response	Dose	0.0074	5	1.3642	0.0191
Exponential	Weibull	3	Response	Response	0.0075	5	1.3480	0.0253
Exponential	Weibull	3	dose and response	Dose	0.0071	6	1.3250	0.0314
Hockey Stick	Weibull	3	dose and response	Dose	0.0072	6	1.3165	0.0374
Exponential	Weibull	3	Dose	Dose	0.0078	5	1.3094	0.0434
Hockey Stick	Weibull	4	Response	Response	0.0078	5	1.3072	0.0494
Hockey Stick	Weibull	4	Response	Dose	0.0078	5	1.3072	0.0554
Linear	Weibull	4	Response	dose and response	0.0078	5	1.3072	0.0614
Hockey Stick	Weibull	4	Dose	Response	0.0078	5	1.3072	0.0674
Linear	Weibull	4	dose and response	Response	0.0078	5	1.3072	0.0734
First Order	Weibull	3	Response	Response	0.0078	5	1.3043	0.0794
Linear	Weibull	4	dose and response	Dose	0.0078	5	1.3035	0.0854
Mass Action	Weibull	3	Response	Response	0.0078	5	1.3012	0.0914
Mass Action	Weibull	4	Response	Dose	0.0079	5	1.2966	0.0974
First Order	Weibull	4	Response	Dose	0.0079	5	1.2944	0.1033
Mass Action	Weibull	4	Dose	Response	0.0079	5	1.2870	0.1092
Exponential	Weibull	4	Response	dose and response	0.0074	6	1.2844	0.1151
Hockey Stick	Weibull	3	Dose	Dose	0.0080	5	1.2823	0.1210

Table 3: Sample Output for Maternal Hair MeHg (ppm) vs. Child AOT (months)

Dose	*Population*	*Study*	*Uncertainty*		
(ppm in Hair)	*Frequency*		*Average*	*Median*	*0.95*
1	0.5	Seychelles	10.42	10.46	10.52
1	0.95	Seychelles	15.33	15.06	16.23
10	0.5	Seychelles	10.74	10.79	10.87
10	0.95	Seychelles	15.77	15.38	16.70
100	0.5	Seychelles	13.88	13.91	15.11
100	0.95	Seychelles	20.15	19.63	23.28
10 vs. 1	0.5	Seychelles	0.32	0.31	0.47
10 vs. 1	0.95	Seychelles	0.45	0.43	0.73
1	0.5	Iraq	16.82	16.93	17.91
1	0.95	Iraq	23.55	23.66	27.36
10	0.5	Iraq	17.13	17.25	18.00
10	0.95	Iraq	23.98	24.11	27.76
100	0.5	Iraq	20.23	20.39	20.95
100	0.95	Iraq	28.28	28.67	32.23
10 vs. 1	0.5	Iraq	0.31	0.31	0.42
10 vs. 1	0.95	Iraq	0.44	0.43	0.63

The average, median, and 95th percentiles for predicted AOT is given for various combinations of dose, population frequency, study population, and likelihood. The values for the doses "10 vs 1" represent the net difference in expected AOT with maternal concentrations of methylmercury in hair at 10 ppm vs 1 ppm.

Table 4: Sample Output for Maternal Hair MeHg (ppm) vs. Child AOW (months)

Dose	*Population*	*Study*	*Uncertainty*		
(ppm in Hair)	*Frequency*		*Average*	*Median*	*0.95*
1	0.5	Seychelles	9.95	9.96	10.23
1	0.95	Seychelles	14.43	14.33	15.10
10	0.5	Seychelles	10.32	10.38	10.58
10	0.95	Seychelles	15.21	14.94	16.33
100	0.5	Seychelles	14.11	14.35	15.17
100	0.95	Seychelles	22.99	21.45	28.64
10 vs. 1	0.5	Seychelles	0.38	0.39	0.48
10 vs. 1	0.95	Seychelles	0.78	0.64	1.29
1	0.5	Iraq	11.09	11.53	12.31
1	0.95	Iraq	16.39	16.10	20.99
10	0.5	Iraq	11.47	11.90	12.72
10	0.95	Iraq	17.17	16.69	22.09
100	0.5	Iraq	15.25	15.41	16.74
100	0.95	Iraq	24.93	23.57	33.13
10 vs. 1	0.5	Iraq	0.38	0.39	0.48
10 vs. 1	0.95	Iraq	0.78	0.64	1.29

The average, median, and 95th percentiles for predicted AOW is given for various combinations of dose, population frequency, study population, and likelihood. The values for the doses "10 vs 1" represent the net difference in expected AOW with maternal concentrations of methylmercury in hair at 10 ppm vs 1 ppm.

Table 5: Sample Output for Maternal Hair MeHg (ppm) vs Child CM (months)

Dose	*Population*	*Study*	*Uncertainty*		
(ppm in Hair)	*Frequency*		*Average*	*Median*	*0.95*
1	0.5	Seychelles	10.23	10.23	10.56
1	0.95	Seychelles	13.37	13.43	14.41
10	0.5	Seychelles	10.56	10.58	10.73
10	0.95	Seychelles	14.04	13.88	15.53
100	0.5	Seychelles	13.73	14.08	15.05
100	0.95	Seychelles	20.27	18.53	26.66
10 vs. 1	0.5	Seychelles	0.33	0.34	0.52
10 vs. 1	0.95	Seychelles	0.67	0.47	1.30
1	0.5	Iraq	14.53	15.14	15.97
1	0.95	Iraq	19.40	18.92	29.37
10	0.5	Iraq	14.86	15.52	15.98
10	0.95	Iraq	20.07	19.58	30.12
100	0.5	Iraq	18.13	18.84	19.78
100	0.95	Iraq	26.36	25.03	37.67
10 vs. 1	0.5	Iraq	0.33	0.34	0.44
10 vs. 1	0.95	Iraq	0.66	0.47	1.16

The average, median, and 95th percentiles for predicted CM is given for various combinations of dose, population frequency, study population, and likelihood. The values for the doses "10 vs 1" represent the net difference in expected CM with maternal concentrations of methylmercury in hair at 10 ppm vs 1 ppm.