Environ Health Perspect
DOI: 10.1289/EHP614

Note to readers with disabilities: *EHP* strives to ensure that all journal content is accessible to all readers. However, some figures and Supplemental Material published in *EHP* articles may not conform to 508 standards due to the complexity of the information being presented. If you need assistance accessing journal content, please contact ehponline@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.

Supplemental Material

The Joint Effect of Prenatal Exposure to Metal Mixtures on Neurodevelopmental Outcomes at 20-40 Months of Age: Evidence from Rural Bangladesh

Linda Valeri, Maitreyi M. Mazumdar, Jennifer F. Bobb, Birgit Claus Henn, Ema Rodrigues, Omar I.A. Sharif, Molly L. Kile, Quazi Quamruzzaman, Sakila Afroz, Mostafa Golam, Citra Amarasiriwardena, David C. Bellinger, David C. Christiani, Brent A. Coull, and Robert O. Wright

Table of Contents

1 Bayesian Kernel Machine Regression
1.1 Overview
1.2 Prior specification
1.3 Implementation
2 Sensitivity analyses for BKMR
3 BKMR with Hierarchical Variable Selection
Table S1. Descriptive characteristics of mother-infant pairs in the neurodevelopment study and mother-infant pairs in the reproductive health study
Table S2. Results from multivariable regression models of cognitive composite score stratified by clinic and for all sample.

Table S3. Results from multivariable regression models of linguistic composite score stratified by clinic and for all sample.

Figure S1. Joint effect of the mixture on Cognitive Composite Score estimated by BKMR in all sample

Figure S2. Joint effect of the mixture on Linguistic Composite Score estimated by BKMR in all sample

Figure S3. Joint effect of the mixture on Linguistic Composite Score in Pabna clinic estimated by BKMR

Figure S4. Joint effect of the mixture on Linguistic Composite Score in Sirajdikhan clinic estimated by BKMR

Additional Files

Supplemental Code and Data ZIP File

bkmr_ehp.zip

bkmr

Description: The folder contains functions and read me files to run BKRM)

- “bkmr.Rproj” indicates that BKMR is an R package
- “R” (folder contains main r functions to run BKMR and to produce plots)
- “Vignettes” (folder contains example for use of BKMR)
- “man” (folder contains files describing the objects in the BKMR package)

code_BKMR_analyses_EHP.R

Description: code used to conduct the analyses presented in the paper: “The joint effect of prenatal exposure to metal mixtures on neurodevelopmental outcomes at 20-40 months of age: evidence from rural Bangladesh”, by Valeri et al. The code calls the BKMR procedure files)

Supplemental Material: The joint effect of prenatal exposure to metal mixtures on neurodevelopmental outcomes at 24 months: evidence from rural Bangladesh

Linda Valeri, Maitreyi M. Mazumdar, Jennifer F. Bobb, Birgit Claus Henn, Ema Rodrigues, Omar I.A. Sharif, Molly L. Kile, Quazi Quamruzzaman, Sakila Afroz, Mostafa Golam, Citra Amarasiriwardena, David C. Bellinger, David C. Christiani, Brent A. Coull, Robert O. Wright

1 Bayesian Kernel Machine Regression

1.1 Overview

For each subject $i = 1, ..., n$, we assume

$$Y_i = h(x_i) + z_i^T \beta, + \epsilon_i$$

where Y_i is a neurodevelopment endpoint (CS, MCS, or LCS), $x_i = (As_i, Mn_i, Pb_i)$ denotes the mixture exposure composed by respectively arsenic, manganese, and lead cord blood log-transformed and centered concentrations. $z_i = (z_{i1}, ..., z_{iP})^T$ contains a set of potential confounders, and $\epsilon_i \sim N(0, \sigma^2)$. In the context of environmental mixtures $h(\cdot)$ typically characterizes an exposure-response function that may incorporate non-linearity and/or interaction among the mixture components. In such a setting, it can be difficult to specify a set of basis functions to represent $h(\cdot)$ or to fit the resulting model that has a high-dimensional parameter space; we therefore employ a kernel machine representation (Cristianini and Shawe-Taylor, 2000).

Operationally, Liu and others (2007) showed that the model can be expressed as the mixed model

$$y_i \sim N(h_i + z^T \beta, \sigma^2) \quad independent; \;\; i = 1, ..., n,$$
$$h = (h_1, ..., h_n)^T \sim N(0, \tau K),$$

where K, referred to as the kernel matrix, has (i, j)-element $K(x_i, x_j)$.
In the present study we employ the Gaussian kernel, which flexibly captures a wide range of underlying functional forms for $h(\cdot)$, although the methods are applicable to a broad choice of kernels. To provide some intuition for BKMR using the Gaussian kernel, consider the effect on neurodevelopment of the metals exposure for the ith person, given by $h_i = h(x_i)$. Under the gaussian kernel, we assume $cor(h_i, h_j) = exp\{-(1/\rho)\sum_{m=1}^{3}(z_{im} - z_{jm})^2\}$, which implies that two subjects with similar exposures (x_i "close" to x_j) will have more similar neurodevelopment outcomes (h_i will be close to h_j). Note that the ρ parameter regulates the smoothness of the dose-response function.

1.2 Prior specification

Here we specify prior distributions for the parameters of the Bayesian kernel machine regression (BKMR) model described in the previous section. We assumed $\beta \sim 1$ (flat prior) and $\sigma^{-2} \sim Gamma(a_\sigma, b_\sigma)$, where we set the shape parameter a_σ and scale parameter b_σ to each be 0.001. It is convenient to parameterize BKMR by $\lambda = \tau\sigma^{-2}$, and we assumed a Gamma prior distribution for the variance component of λ having mean and variance each set to 100 (Let a_λ and b_λ denote corresponding shape and rate parameters). For the distribution of the smoothness parameter ρ we assumed $\rho \sim Unif(a, b)$ with $a = 0$ and $b = 100$. Further details on prior specification can be found in Bobb et al. (2015).

1.3 Implementation

For details on the Markov Chain Monte Carlo (mcmc) sampler used to run BKMR, please see Bobb et al. (2015). All of the parameters were sampled using Gibbs steps, except for $\lambda = \tau\sigma^{-2}$ and ρ, which were sampled using the Metropolis-Hastings (M-H) algorithm. For the M-H steps, we used a random walk proposal distribution centered about the current parameter value. We tuned the variance of the proposal distribution to achieve a good acceptance rate (about 20%).
The mcmc sampler was run for 10,000 iterations. Convergence of the fit was assessed by inspection of the chain trace plot for the parameters involved in the estimation excluding a burn-in of dimension 5,000. Inferences on mixture effects were obtained by computing for each mcmc sample the posterior estimates of cognitive scores and posterior credible intervals at different levels of the mixture components. In particular, we estimated:

(1) The cumulative effect of the mixture by estimating the change in the predicted cognitive scores for departures of all the components of the mixture from their median level;

(2) The effect of an IQR change of each metal on neurodevelopment and potential interactions among the metals by estimating the change in the predicted cognitive scores for a change in the component of interest from its 25th to 75 percentile, while setting the other metals at the median, the 25th, or the 75th percentile levels;

(3) The dose response relationship of each mixture component and potential interactions among the metals by estimating the predicted cognitive scores for each level of the component of interest, setting the other metals at the median, the 25th, or the 75th percentile levels.

2 Sensitivity analyses for BKMR

The results of the BKMR fit can be sensitive to the choice of ρ parameter, which controls the smoothness of the exposure response function. The parameter ρ can take values $[0, Inf)$. Our strategy was to consider the class of uniform prior distributions for $\rho \sim Unif(a, b)$, where we varied the value of b. We considered lower degree of smoothness ($b = 50$), and higher degree of smoothness ($b = 200$) with respect to what was specified in the primary analyses. Findings were not sensitive to the choice of this smoothing parameter.

3 BKMR with Hierarchical Variable Selection

In situations where pollutant concentrations in the mixture are highly correlated, the above formulation may fail because the data may not be able to distinguish among these correlated pollutants. Bobb et al. (2015) therefore also propose a hierarchical variable selection approach, which incorporates knowledge of the structure of the mixture into the model. Mixture components can at times be partitioned into groups of elements. These groups may be defined by high correlations or by external knowledge such as timing of exposure or the source of each component or other common feature. We here assume that group membership is pre-specified by the investigator. Once group membership is defined, BKMR carries out a hierarchical variable selection strategy that first estimates the probability that each group of pollutants should be included in the model, and then assesses whether there is evidence in the data that one of the pollutants in the group drives the group effect.
Suppose the pollutants can be partitioned, using prior knowledge, into groups $z_1, ..., z_m$. For example, a wealth of information about air pollution sources is typically known, allowing for the pollution constituents to be grouped S_g $g = (1, ..., G)$ such that within-group correlation is high while across-group correlation is moderate to low. We then define an indicator variable δ_{S_g} distributed as

$$\delta_{S_g}|\omega_g \sim Multinomial(\omega_g, \pi_{S_g}), g = 1, ..., G$$

$$\omega_g \sim Bernoulli(\pi)$$

where δ_{S_g} is the vector of indicator variables and π_{S_g} is the corresponding vector of prior probabilities for the pollutants z_m in group S_g. This approach allows at most a single pollutant from a group (of highly correlated pollutants) to enter into the model at a time. Although this assumes that two pollutants from the same group do not have independent or interactive effects on the health outcome, in the setting of high within-group correlation, such effects would not be identifiable by any model.

In our study, we are interested in understanding whether the findings are robust to adjustment to child exposures to heavy metals and we therefore define two groups. Prenatal arsenic, manganese and lead cord blood concentrations form one group, and 20-40 month exposure to arsenic and manganese (measured in water) and lead (measured in blood), as considered in Rodriguez et al. (2016), form a second group. Findings were not sensitive to the further adjustment for 20-40 months exposure to heavy metals and prenata exposure was found to be the most important window of vulnerability for neurodevelopment at 20-40 months.

Supplemental Tables

Table S1 Descriptive characteristics of mother-infant pairs in the neurodevelopment study and mother-infant pairs in the reproductive health study

	Repro Study (n=1608)[a]			**Neurodevelopment Study (n=825)[b]**		
	n (%)	Mean ± SD	Range	n (%)	Mean ± SD	Range
Prenatal exposure measures (GM ± GSD)[c]						
Cord blood As* (µg/dl)	1093	0.56±2.34	0.06-23.4	818	3.27 ±2.38	0.07-27.71
Cord blood Mn* (µg/dl)	1093	6.40±2.12	1.23-303.1	818	5.36±2.28	1.24-303.1
Cord blood Pb* (µg/dl)	1093	3.19±2.35	0.36-83.5	818	7.22±2.44	0.27-79.18
Child characteristics						
Birth weight (g)	1184	2.84±0.41	0.80-4.80	823	2.85 ±0.40	1.02-4.8
Gestational age at birth (weeks)	1180	37.99±2.0	22-42	824	39.54±1.55	30-43
Head circumference at birth (cm)	1184	32.67±1.33	24-48	825	32.72±1.36	24-48
Female sex	583 (49.3)			405 (49.2)		
Maternal characteristics						
Age at enrollment (years)	1608	22.89±4.18	18-41	825	22.99±4.23	18-41
Education: ≥secondary	1369 (85.1)			701 (84.9)		

Any smokers in household: yes	680 (42.3)			351 (42.6)		
Protein intake LOW*	584 (36.3)			207 (25.1)		
Protein intake MEDIUM*	696 (43.2)			427 (51.7)		
Protein intake HIGH*	328 (20.3)			191 (23.1)		

a – Numbers may not sum to total sample size (n=1608) for some characteristics due to missing data (of the 1613 mothers, 5 had twins and they were excluded from the reproductive study)

b – Numbers may not sum to total sample size (n=827) for some characteristics due to missing data

c – Geometric mean ± geometric standard deviation reported for blood metals concentrations.

IQR in Neurodevelopmental study: As = (0.4, 1.0), Mn = (4.3, 5.6), Pb = (1.6, 6.5)

*Individuals in Reproductive study differed from Neurodevelopment study, $p<0.05$.

Table S2. Results from multivariable regression models of cognitive composite score stratified by clinic and for all sample.

	Dependent variable:		
	Cognitive Composite Score		
	Sirajdikhan	Pabna	All
As	−0.017	0.073	0.082*
	(0.034)	(0.079)	(0.047)
Mn	0.025	−0.206**	−0.088**
	(0.065)	(0.094)	(0.038)
Pb	−0.075*	0.024	−0.084
	(0.045)	(0.088)	(0.070)
Female	−0.122**	−0.043	−0.093
	(0.060)	(0.101)	(0.059)
Testing Age	0.159***	0.210***	0.189***
	(0.014)	(0.018)	(0.011)
Testing Age^2	−0.014***	−0.016***	−0.015***
	(0.003)	(0.003)	(0.002)
Mother Age	−0.046	0.178*	0.090
	(0.065)	(0.103)	(0.061)
Mother Age^2	0.001	−0.003	−0.002
	(0.001)	(0.002)	(0.001)
Mother Raven Score	0.018	−0.060	−0.002
	(0.030)	(0.071)	(0.033)
Mother Raven Score^2	0.017	−0.117***	−0.035*
	(0.018)	(0.042)	(0.019)
Home Score	0.122***	−0.106	−0.070*
	(0.045)	(0.069)	(0.039)
Home Score^2	−0.166***	−0.044	−0.047**
	(0.033)	(0.035)	(0.022)
Secondary Education	0.067	0.462***	0.375***
	(0.110)	(0.134)	(0.089)
Smoking Environment	0.074	−0.125	−0.026
	(0.063)	(0.101)	(0.060)
Protein Intake	−0.033	0.307***	0.121**
	(0.062)	(0.100)	(0.058)
Pabna Clinic			−0.415***
			(0.118)
Constant	0.938	−3.181**	−1.243
	(0.818)	(1.267)	(0.767)
Observations	403	389	792
R^2	0.420	0.369	0.349
Adjusted R^2	0.397	0.344	0.335
Residual Std. Error	0.584 (df = 387)	0.964 (df = 373)	0.817 (df = 775)
F Statistic	18.648*** (df = 15; 387)	14.543*** (df = 15; 373)	25.939*** (df = 16; 775)

Note: *p<0.1; **p<0.05; ***p<0.01

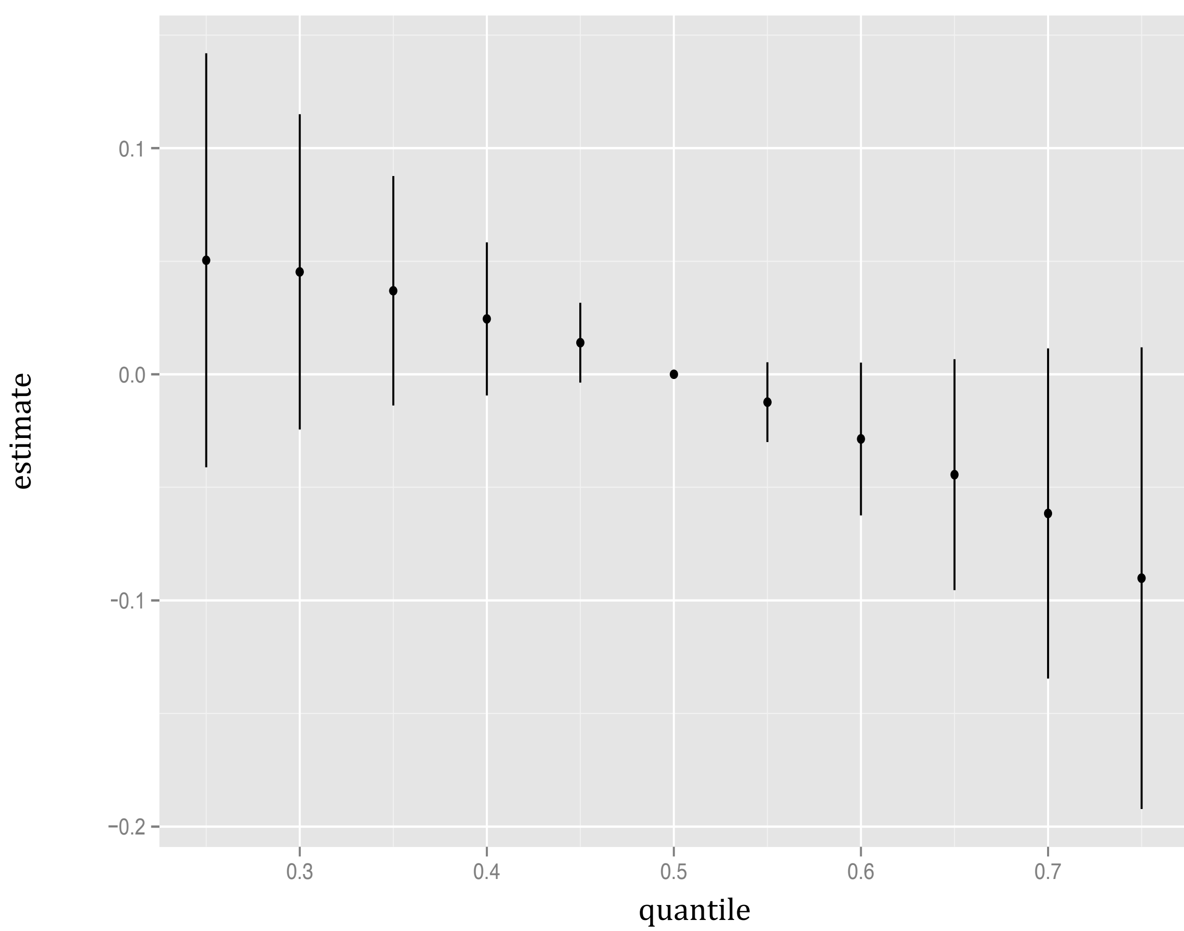
Table S3. Results from multivariable regression models of linguistic composite score stratified by clinic and for all sample.

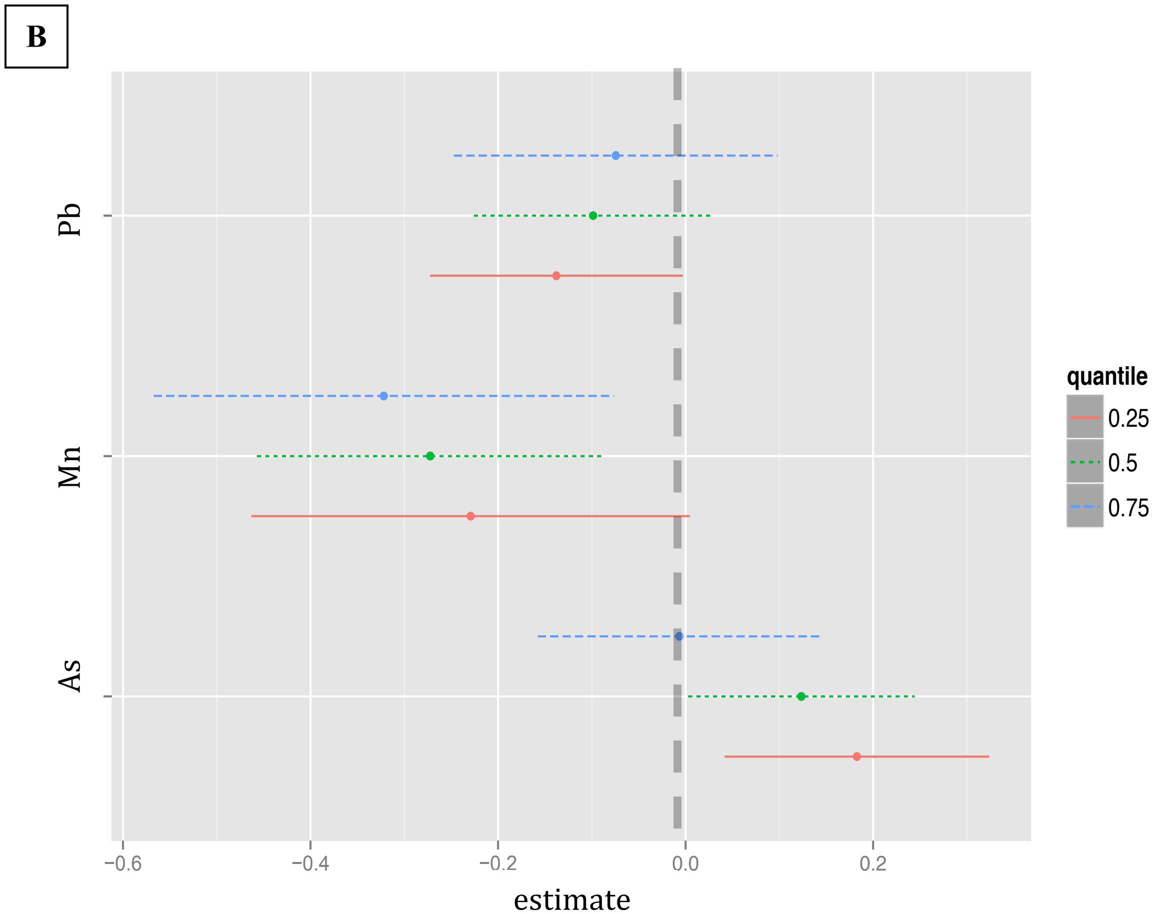
	Dependent variable:		
	Linguistic Composite Score		
	Sirajdikhan	Pabna	All
As	−0.080*	0.035	0.045
	(0.047)	(0.081)	(0.049)
Mn	0.034	−0.047	−0.029
	(0.038)	(0.096)	(0.039)
Pb	−0.028	−0.028	−0.090
	(0.051)	(0.066)	(0.073)
Female	−0.011	0.214**	0.081
	(0.068)	(0.103)	(0.061)
Testing Age	0.116***	0.208***	0.171***
	(0.016)	(0.019)	(0.012)
Testing Age^2	0.0003	−0.001	−0.001
	(0.001)	(0.002)	(0.001)
Mother Age	−0.028	0.059	0.025
	(0.073)	(0.105)	(0.064)
Mother Age^2	0.001	−0.001	−0.0001
	(0.002)	(0.002)	(0.001)
Mother Raven Score	0.074**	−0.018	0.035
	(0.034)	(0.073)	(0.034)
Mother Raven Score^2	0.010	−0.026	−0.004
	(0.020)	(0.043)	(0.020)
Home Score	0.159***	−0.063	0.015
	(0.051)	(0.070)	(0.041)
Home Score^2	−0.067*	−0.068*	−0.037
	(0.038)	(0.035)	(0.023)
Secondary Education	0.047	0.384***	0.312***
	(0.124)	(0.137)	(0.092)
Smoking Environment	0.193***	−0.023	0.066
	(0.071)	(0.103)	(0.063)
Protein Intake	0.136*	0.331***	0.193***
	(0.071)	(0.102)	(0.060)
Pabna Clinic			−0.307**
			(0.123)
Constant	0.927	−1.058	0.077
	(0.961)	(1.137)	(0.749)
Observations	403	388	791
R^2	0.304	0.420	0.374
Adjusted R^2	0.277	0.397	0.361
Residual Std. Error	0.686 (df = 387)	0.864 (df = 372)	0.799 (df = 774)
F Statistic	11.275*** (df = 15; 387)	17.953*** (df = 15; 372)	28.917*** (df = 16; 774)
Note:			*p<0.1; **p<0.05; ***p<0.01

Figure S1. Joint effect of the mixture on Cognitive Composite Score estimated by BKMR in all sample

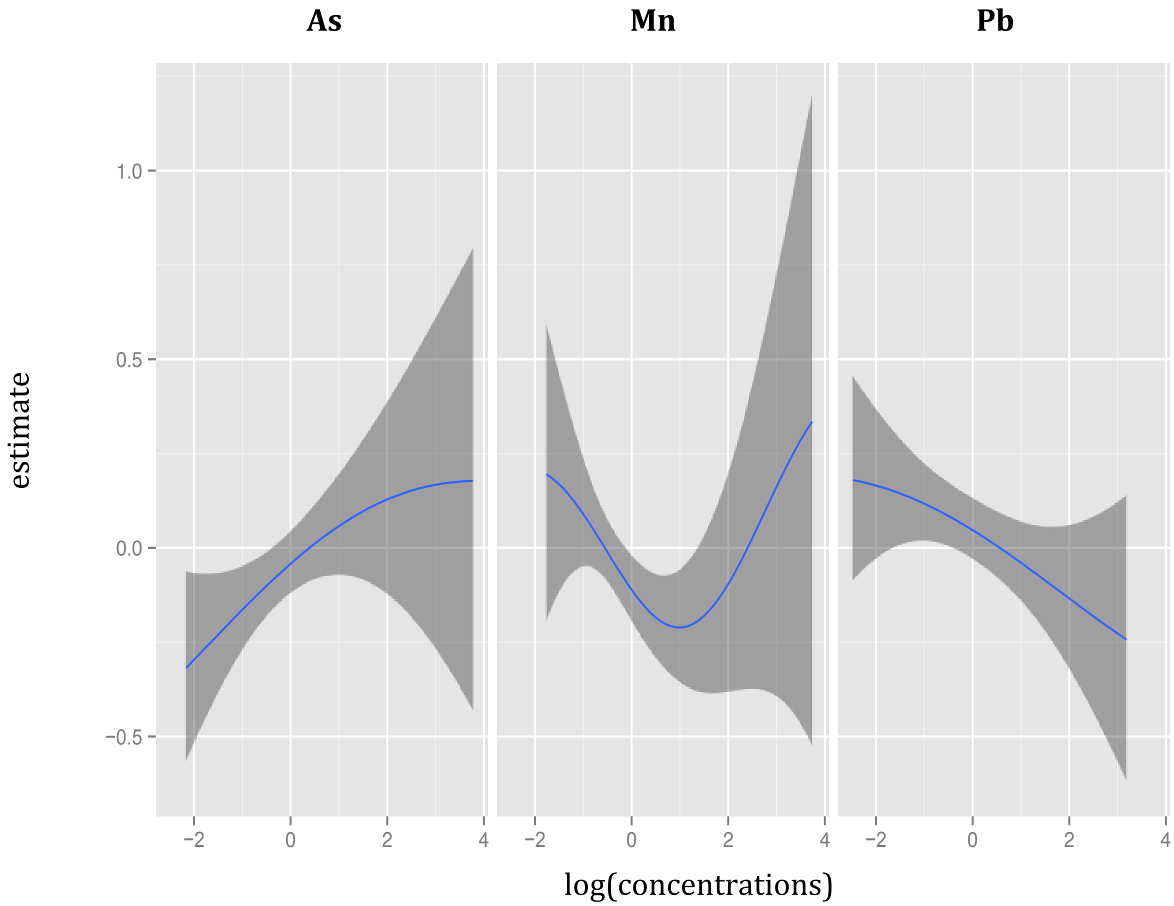
Model adjusted for clinic, child gender, maternal IQ, maternal education, maternal protein intake, smoking environment, age at testing and maternal age. **(A)** Overall effect of the mixture (estimates and 95% credible intervals). This plot compares the risk when all exposures are at a particular quantile to when all are at the 50th percentile. **(B)** Single pollutant association (estimates and 95% credible intervals). This plot compares the risk when a single pollutant is at the 75th versus 25th percentile, when all of the other exposures are fixed at either 25th, 50th or 75th percentile. **(C)** Univariate exposure-response functions and 95% confidence bands for each of the other pollutants fixed at the median. **(D)** Bivariate exposure-response functions for each of the other pollutants fixed at the median.

A




B
Pb
Mn
As
−0.6
−0.4
−0.2
0.0
0.2
estimate
quantile
0.25
0.5
0.75

C



D

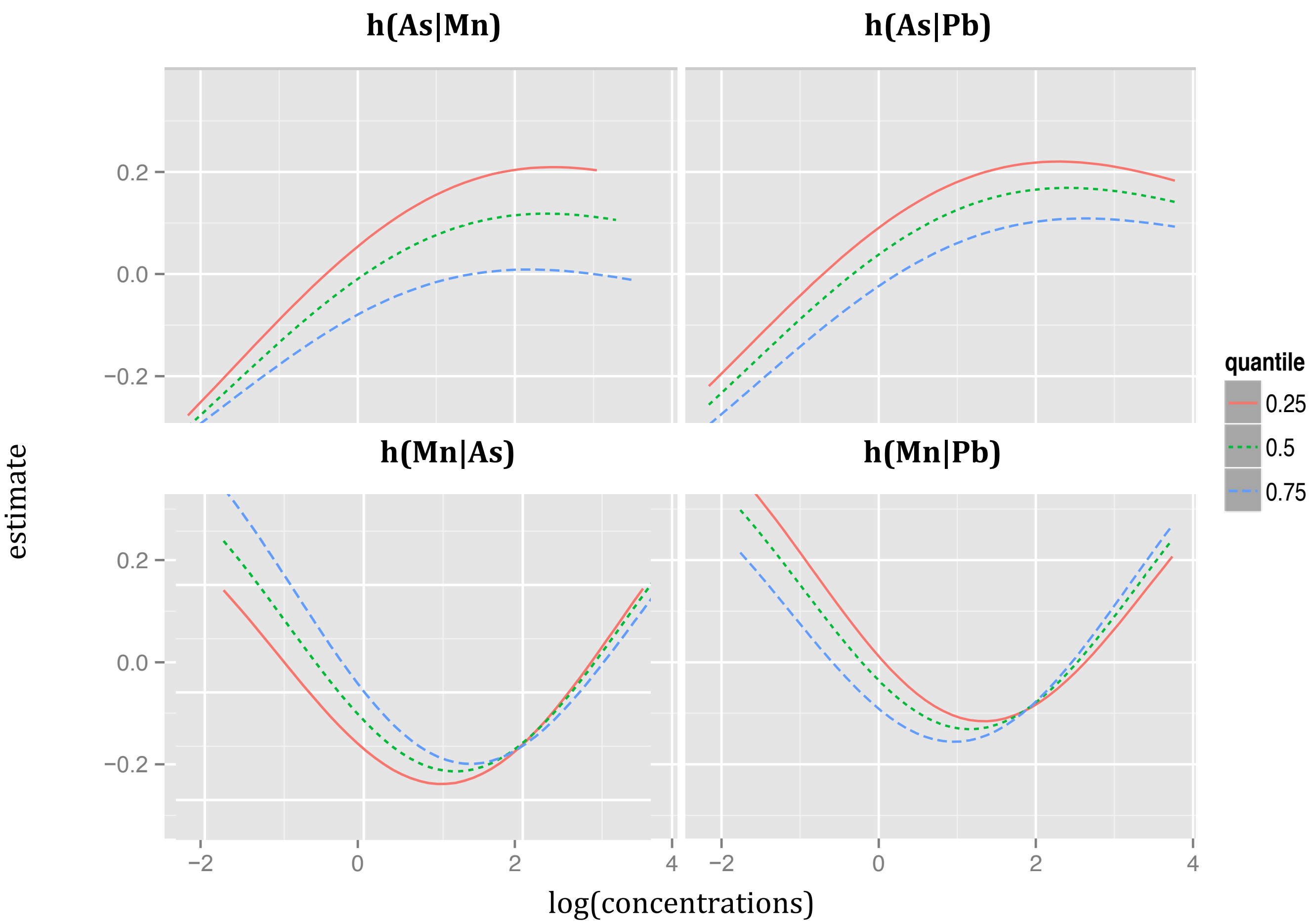
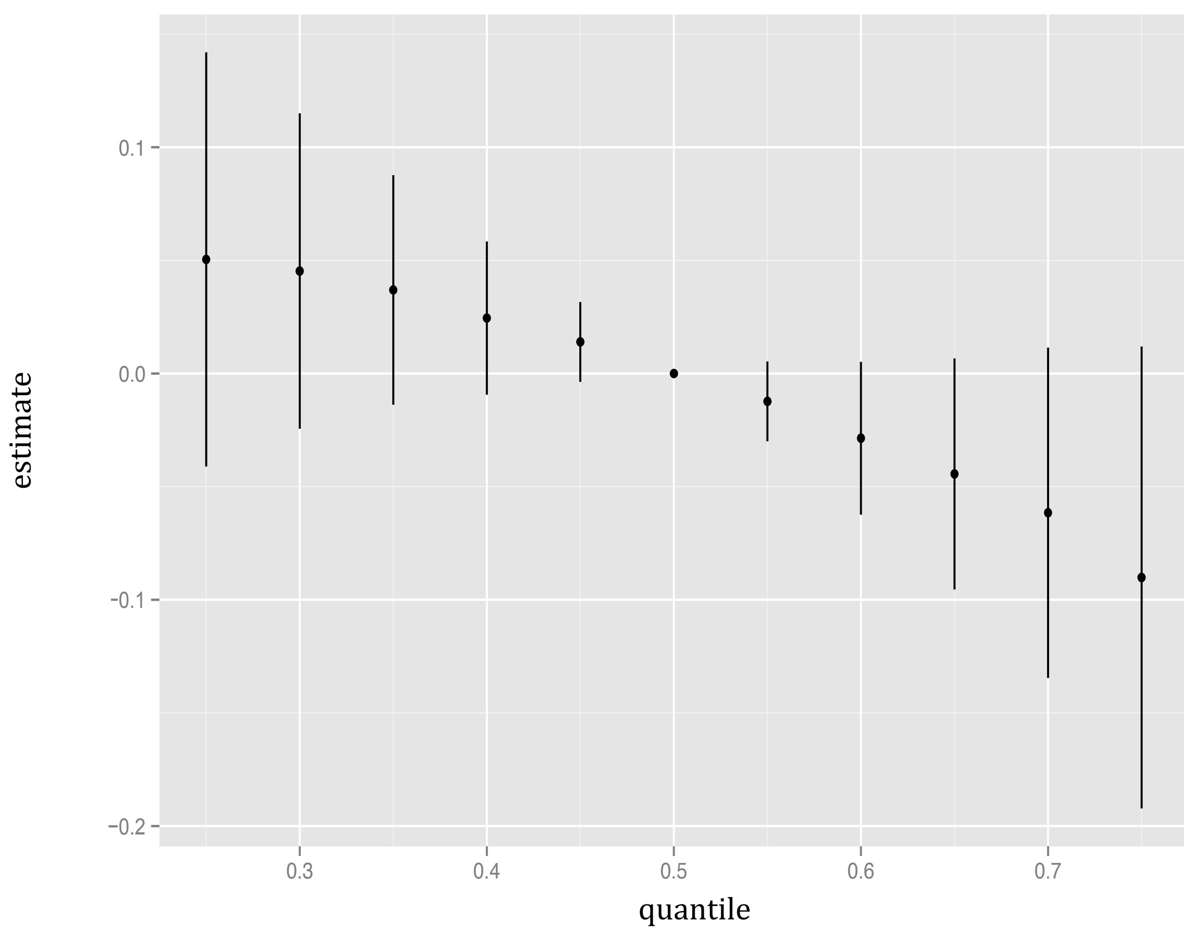

h(As|Mn)
h(As|Pb)
h(Mn|As)
h(Mn|Pb)
0.2
0.0
−0.2
−2
0
2
4
estimate
log(concentrations)
quantile
0.25
0.5
0.75

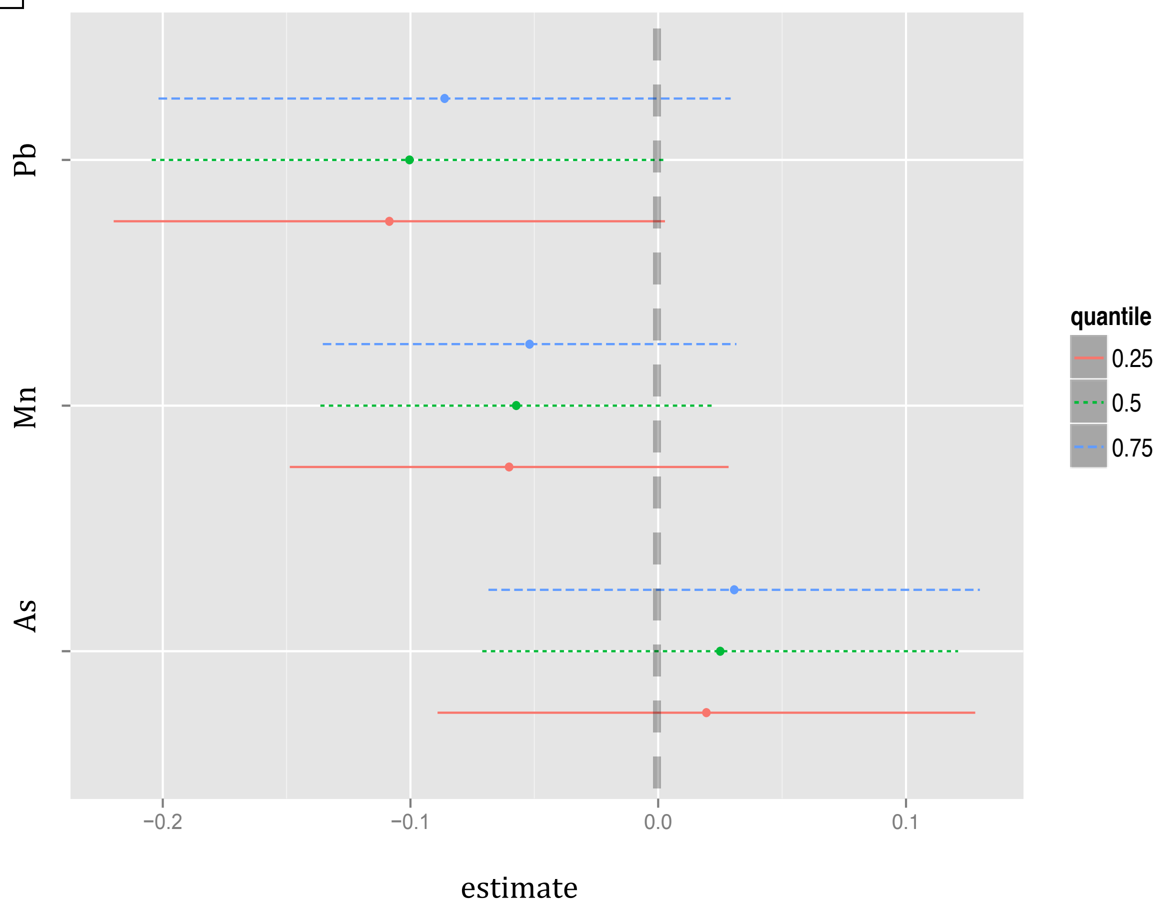
Figure S2. Joint effect of the mixture on Linguistic Composite Score estimated by BKMR in all sample

Model adjusted for clinic, child gender, maternal IQ, maternal education, maternal protein intake, smoking environment, age at testing and maternal age. **(A)** Overall effect of the mixture (estimates and 95% credible intervals). This plot compares the risk when all exposures are at a particular quantile to when all are at the 50th percentile. **(B)** Single pollutant association (estimates and 95% credible intervals). This plot compares the risk when a single pollutant is at the 75th versus 25th percentile, when all of the other exposures are fixed at either 25th, 50th or 75th percentile. **(C)** Univariate exposure-response functions and 95% confidence bands for each of the other pollutants fixed at the median. **(D)** Bivariate exposure-response functions for each of the other pollutants fixed at the median.

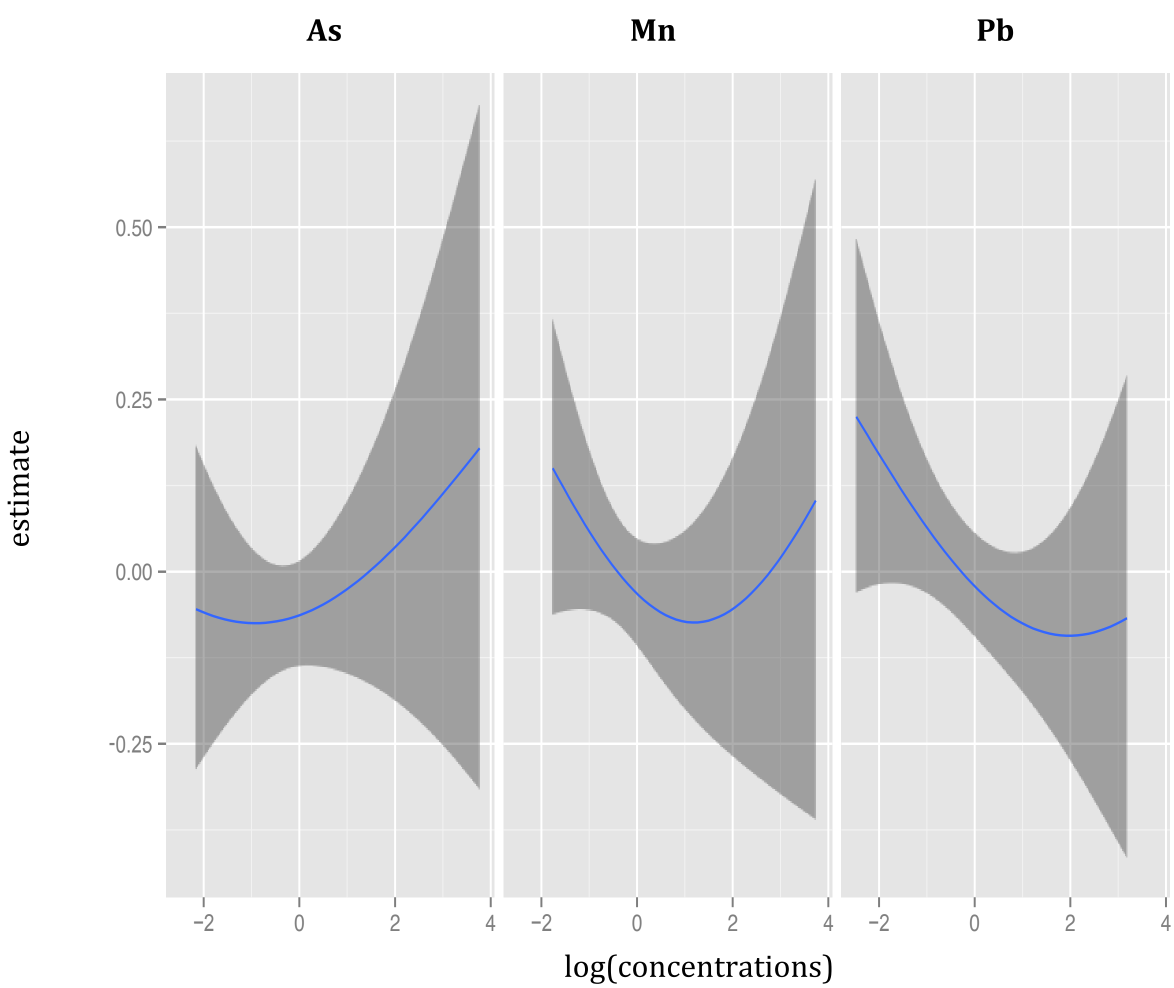
A



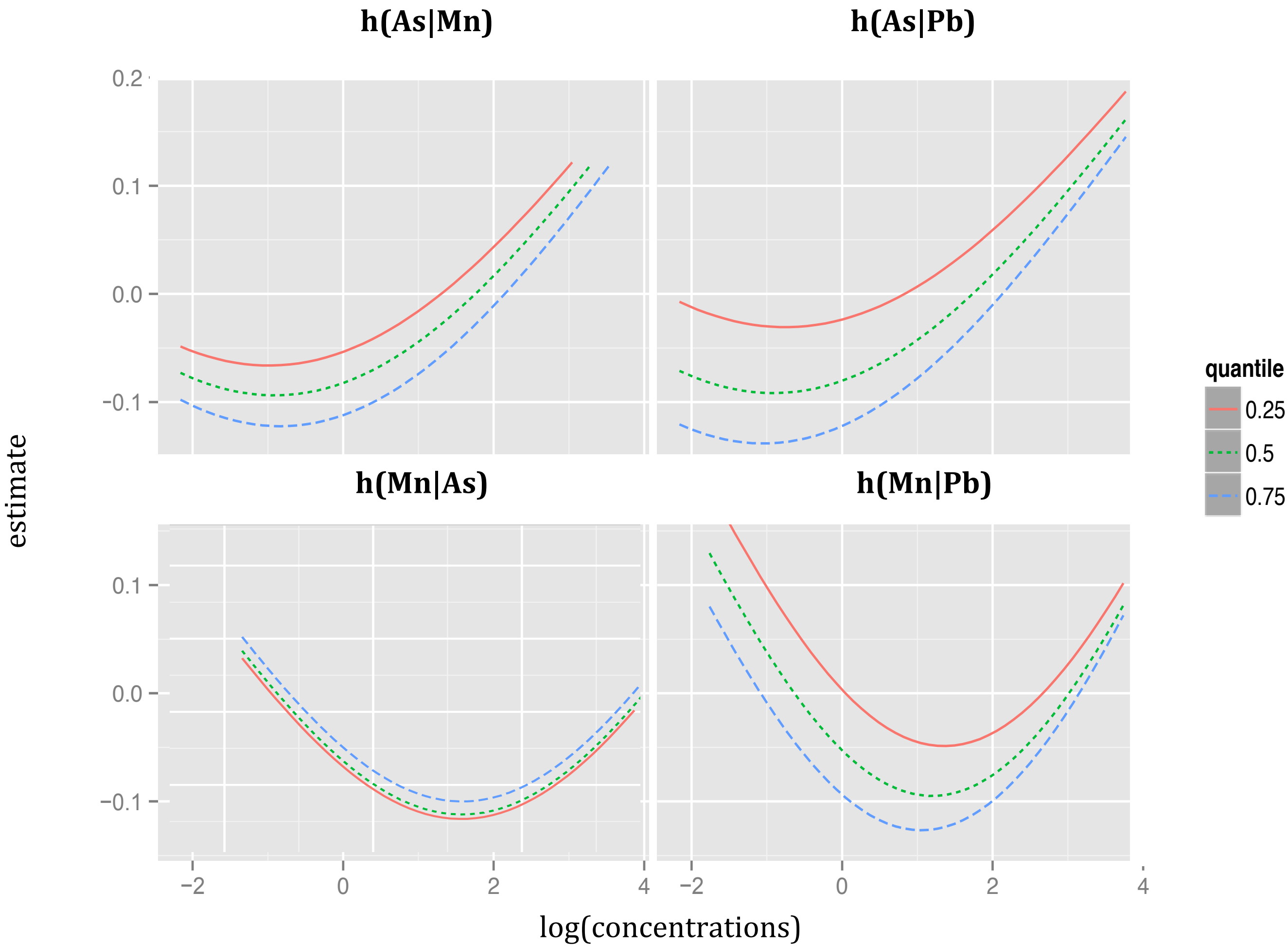
B



C



D

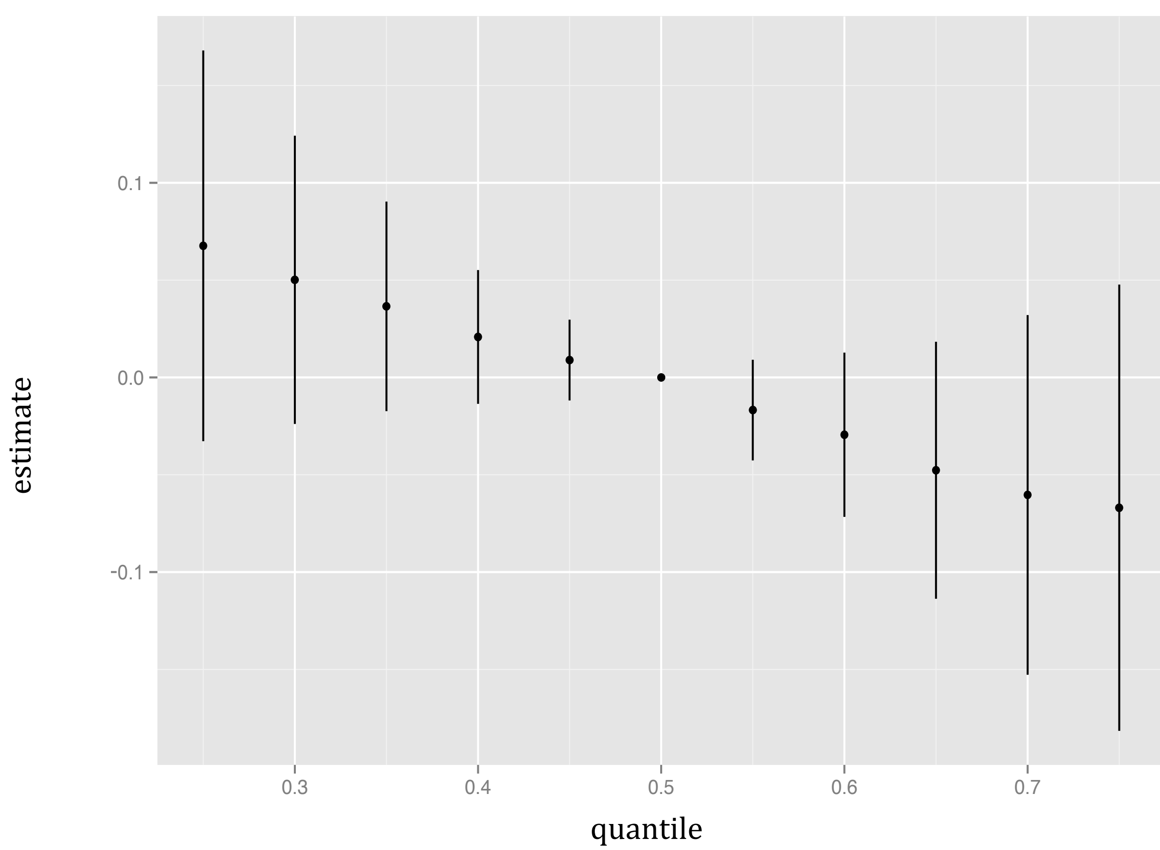


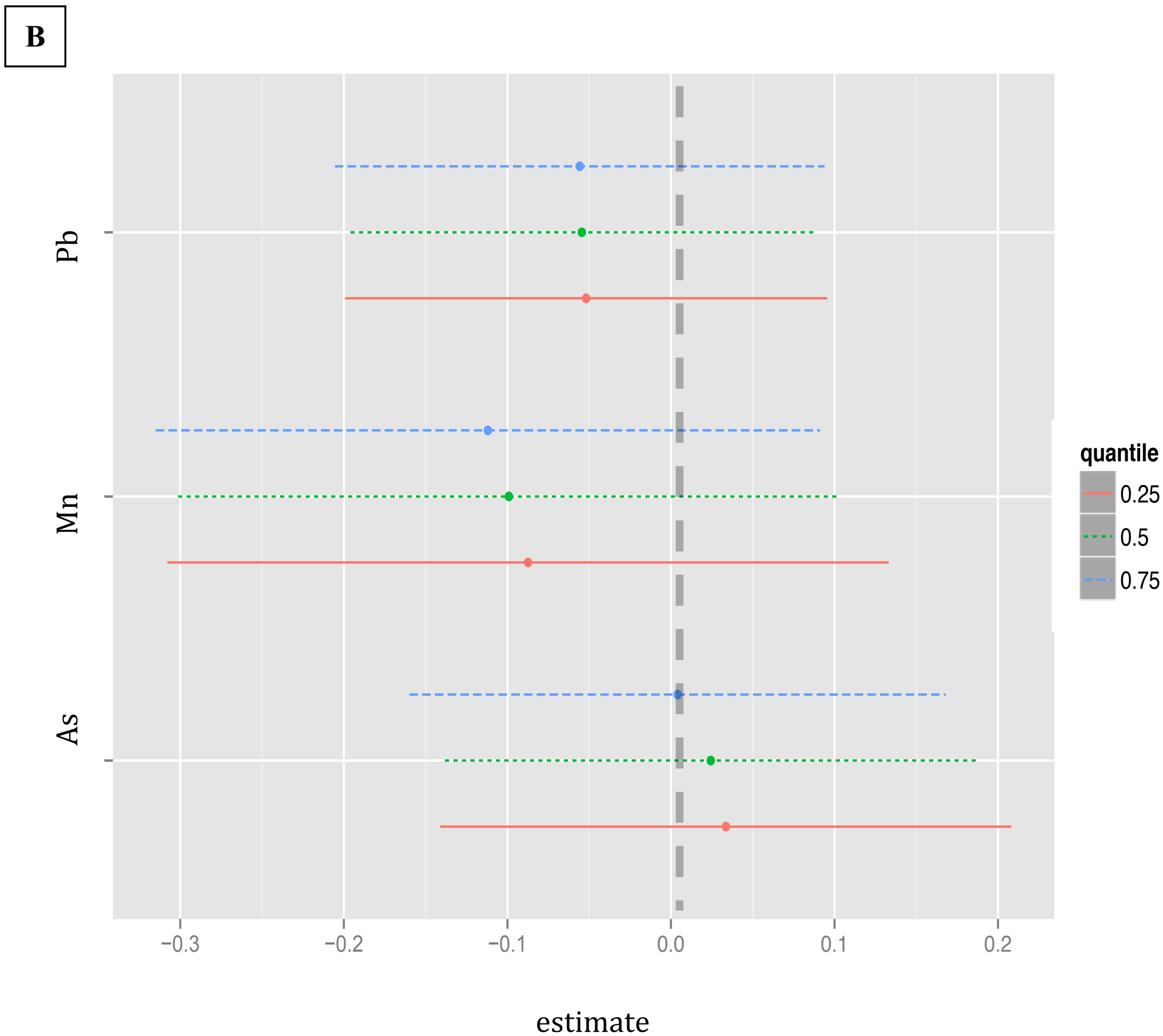
h(As|Mn)
h(As|Pb)
h(Mn|As)
h(Mn|Pb)
estimate
log(concentrations)
quantile
0.25
0.5
0.75
0.2
0.1
0.0
-0.1
-2
0
2
4

Figure S3. Joint effect of the mixture on Linguistic Composite Score in Pabna clinic estimated by BKMR

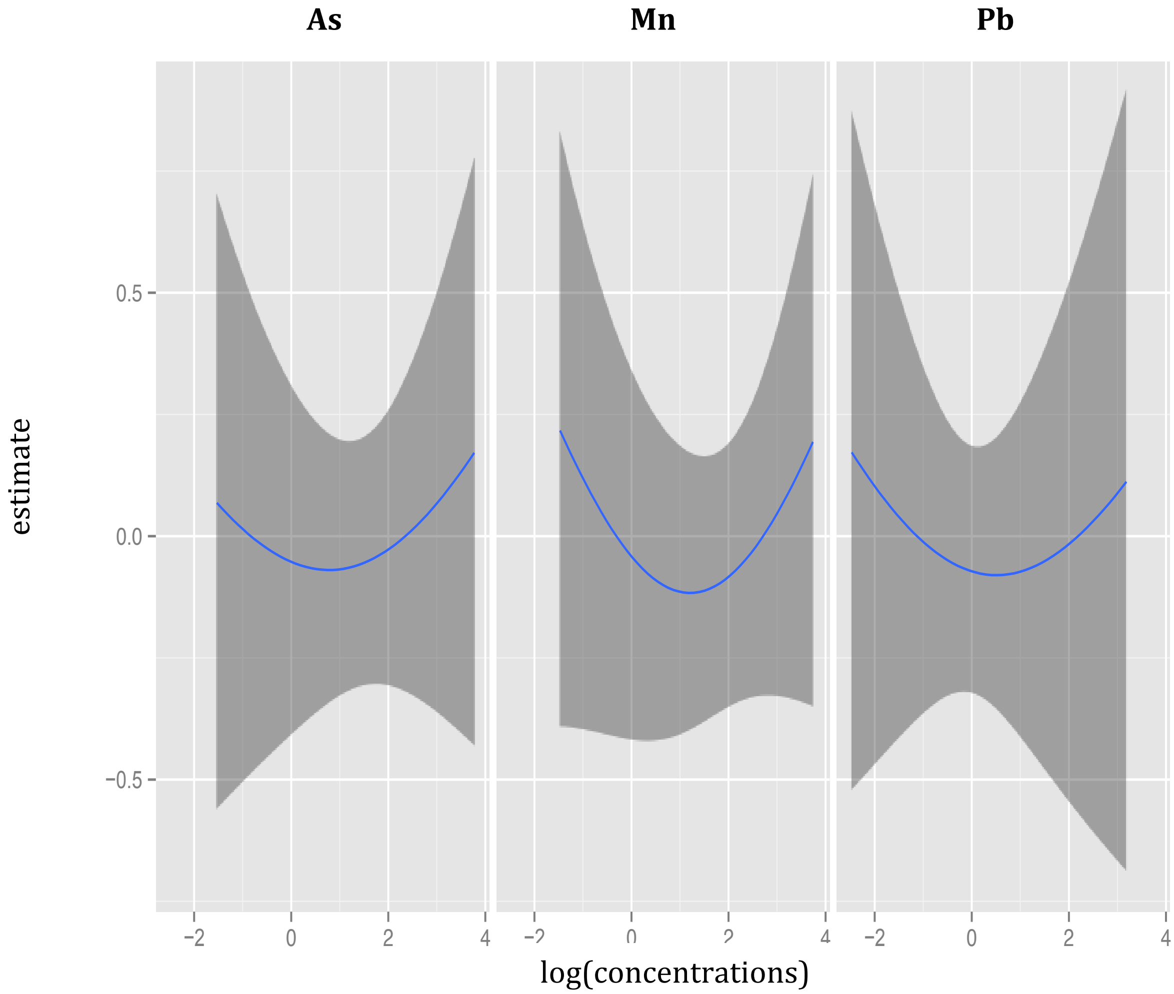
Model adjusted for clinic, child gender, maternal IQ, maternal education, maternal protein intake, smoking environment, age at testing and maternal age. **(A)** Overall effect of the mixture (estimates and 95% credible intervals). This plot compares the risk when all exposures are at a particular quantile to when all are at the 50th percentile. **(B)** Single pollutant association (estimates and 95% credible intervals). This plot compares the risk when a single pollutant is at the 75th versus 25th percentile, when all of the other exposures are fixed at either 25th, 50th or 75th percentile. **(C)** Univariate exposure-response functions and 95% confidence bands for each of the other pollutants fixed at the median. **(D)** Bivariate exposure-response functions for each of the other pollutants fixed at the median.

A




B
Pb
Mn
As
−0.3
−0.2
−0.1
0.0
0.1
0.2
estimate
quantile
0.25
0.5
0.75

C



D

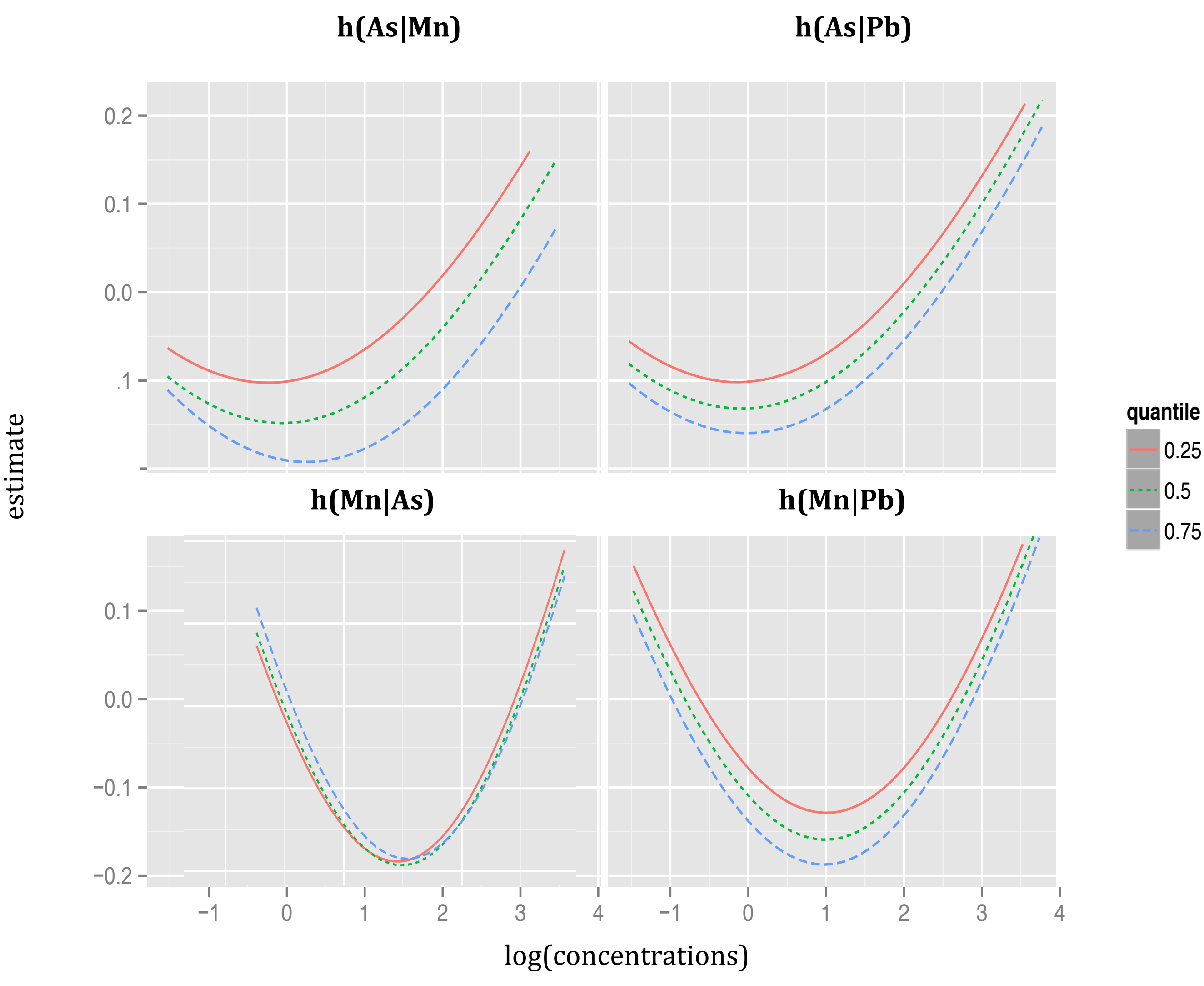
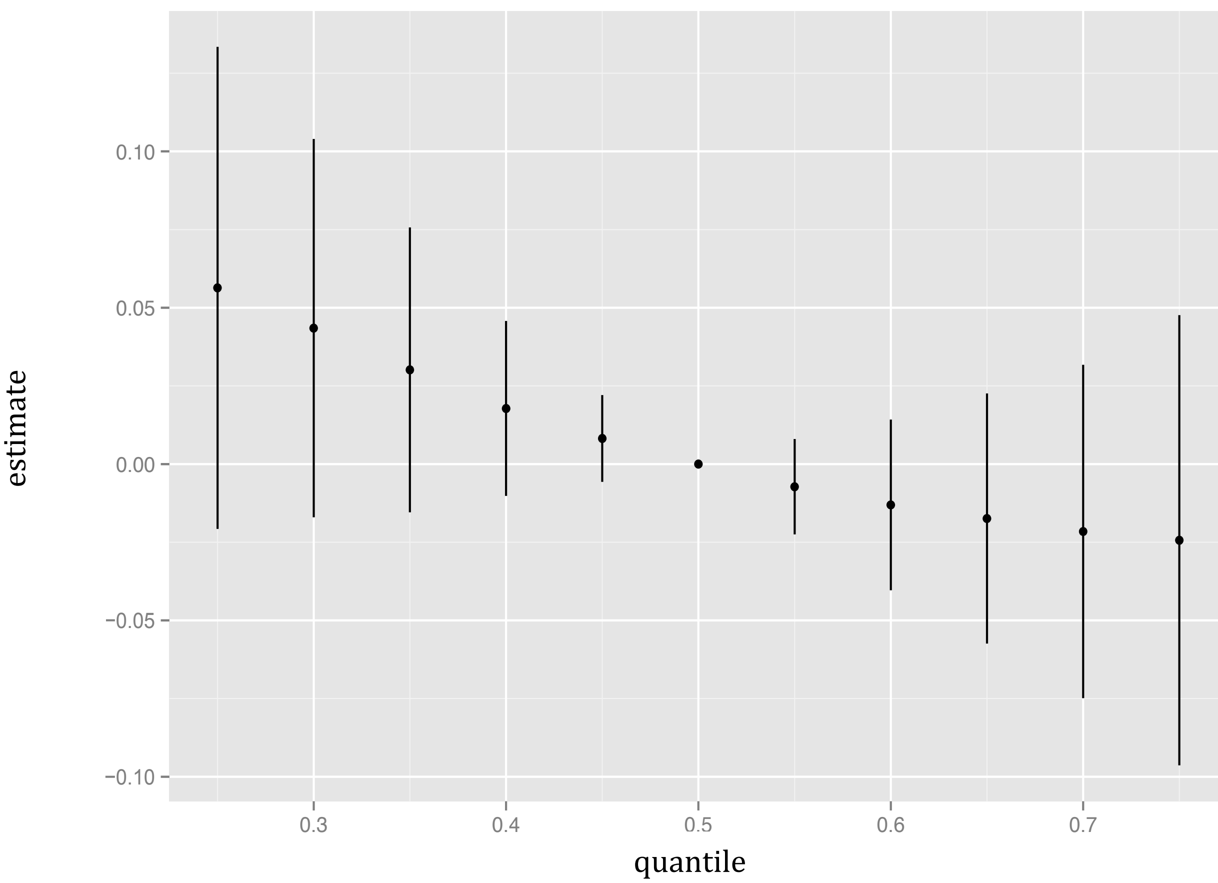

h(As|Mn)
h(As|Pb)
h(Mn|As)
h(Mn|Pb)
0.2
0.1
0.0
.1
-
-0.1
-0.2
−1
0
1
2
3
4
estimate
log(concentrations)
quantile
0.25
0.5
0.75

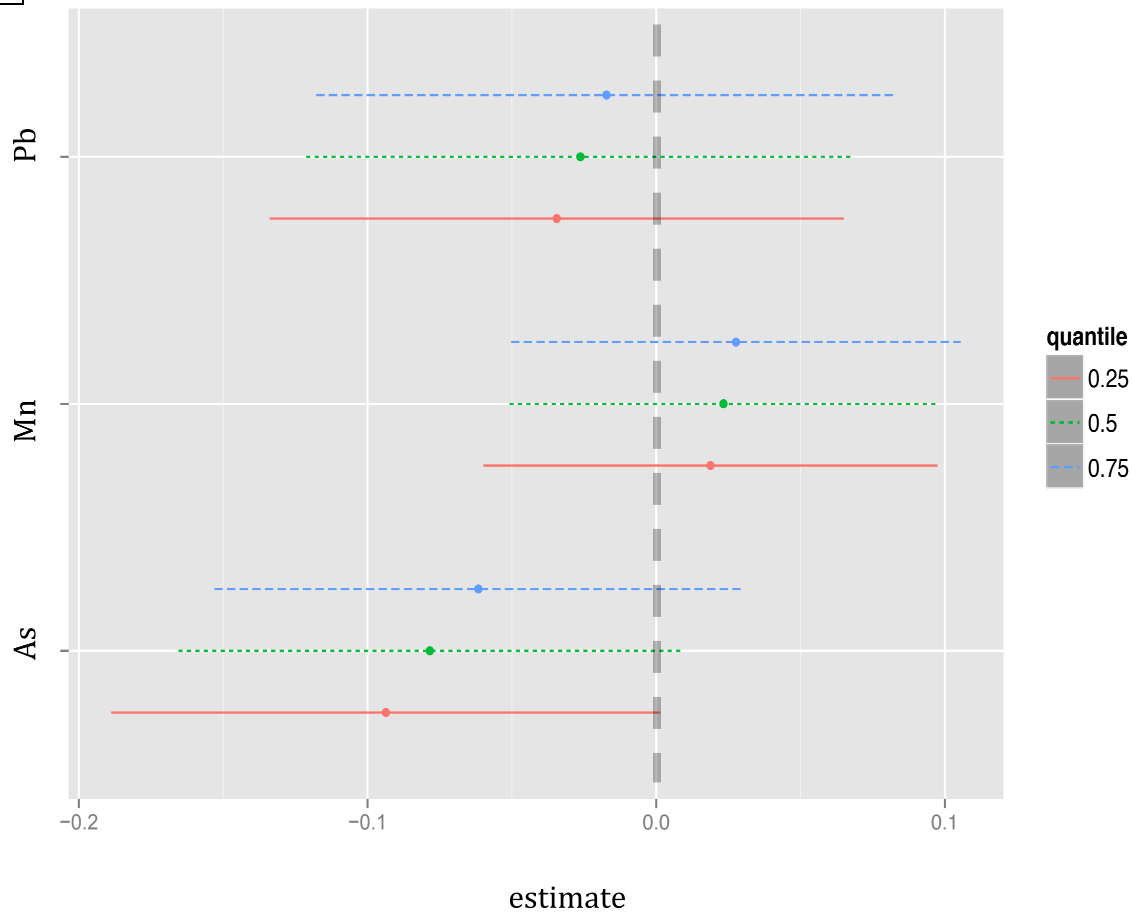
Figure S4. Joint effect of the mixture on Linguistic Composite Score in Sirajdikhan clinic estimated by BKMR

Model adjusted for clinic, child gender, maternal IQ, maternal education, maternal protein intake, smoking environment, age at testing and maternal age. **(A)** Overall effect of the mixture (estimates and 95% credible intervals). This plot compares the risk when all exposures are at a particular quantile to when all are at the 50th percentile. **(B)** Single pollutant association (estimates and 95% credible intervals). This plot compares the risk when a single pollutant is at the 75th versus 25th percentile, when all of the other exposures are fixed at either 25th, 50th or 75th percentile. **(C)** Univariate exposure-response functions and 95% confidence bands for each of the other pollutants fixed at the median. **(D)** Bivariate exposure-response functions for each of the other pollutants fixed at the median.

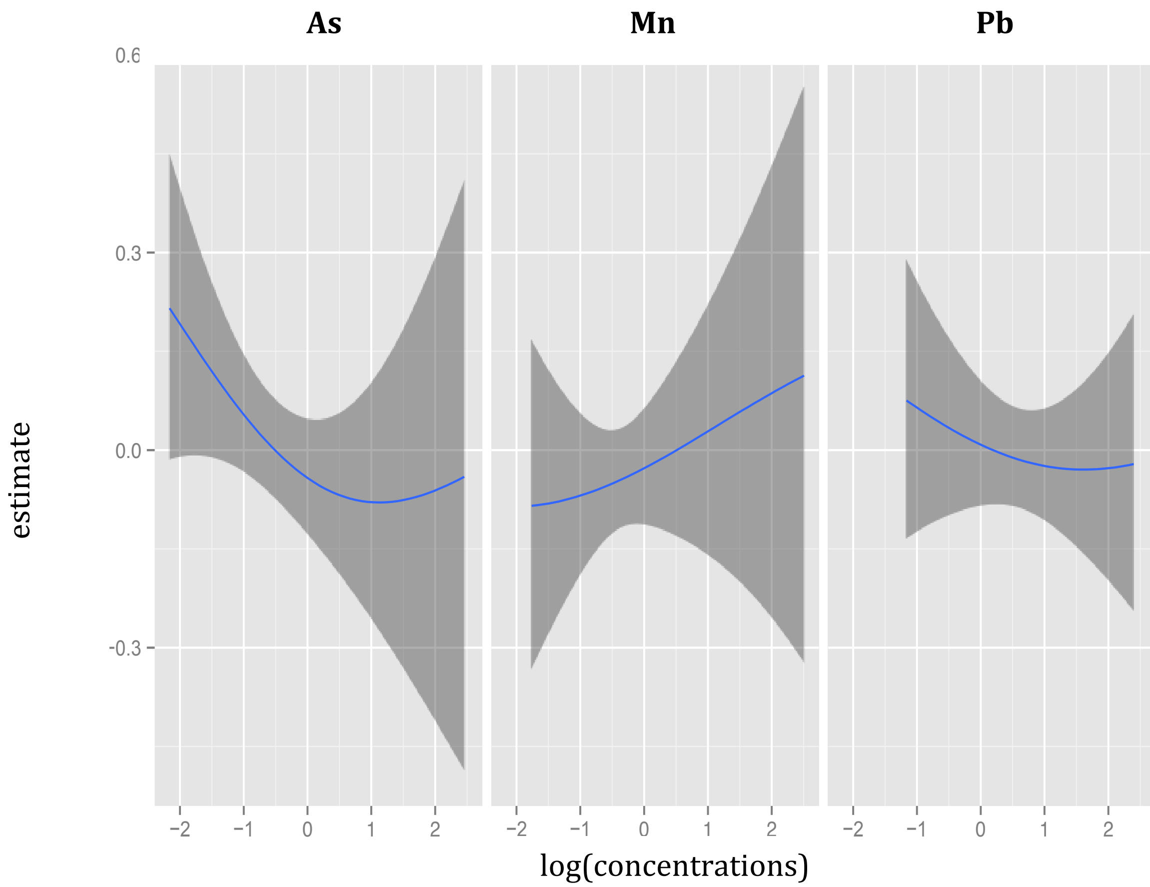
A



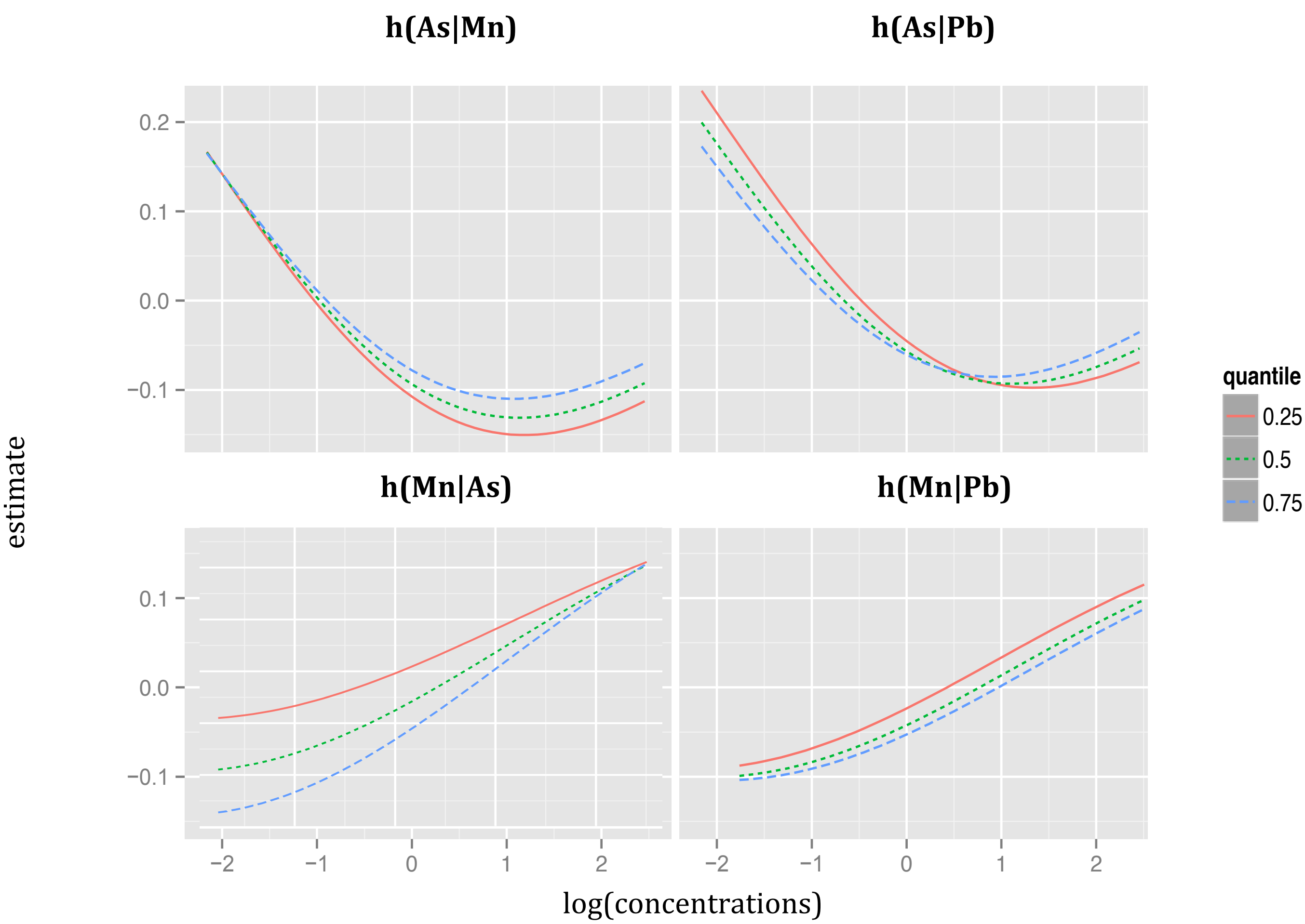
B



C



D


h(As|Mn)
h(As|Pb)
h(Mn|As)
h(Mn|Pb)
estimate
log(concentrations)
quantile
0.25
0.5
0.75
0.2
0.1
0.0
−0.1
−2
−1
0
1
2