WEB MATERIAL

Bayesian G-Computation for Estimating Impacts of Interventions on Exposure Mixtures: Demonstration With Metals From Coal-Fired Power Plants and Birth Weight

Alexander P. Keil, Jessie P. Buckley, and Amy E. Kalkbrenner

TABLE OF CONTENTS

WEB APPENDIX 1: FURTHER DETAILS ON BAYESIAN G-COMPUTATION 2

WEB APPENDIX 2: SENSITIVITY ANALYSES 3

ASSESSING INTERVENTION EXPOSURE LEVELS RELATIVE TO THE JOINT EXPOSURE RANGE 3

Web Figure 1 4

Web Figure 2 5

WEB APPENDIX 3: JOINT EFFECTS IN THE JOINT EXPOSURE RANGE 5

Web Figure 3 7

WEB APPENDIX 4: SIMULATION STUDY 7

Web Table 1 10

Web Table 2 11

Web Table 3 14

Web Table 4 15

Web Table 5 16

WEB APPENDIX 1

Further Details on Bayesian g-Computation

Recall that the distributions of the model parameters from our underlying statistical model are explored via Markov chain Monte Carlo (MCMC). For the *m*th iteration of the MCMC algorithm, the individual level predictions, denoted by $\hat{\mu}_{i,m}(X_{int,i}, T_{int,k})$ are given by **Equation 1** in the main text, but with the *m* specific values of the parameters substituted into the equation. Under standard causal identification conditions and correct model assumptions, these predictions form the basis of the estimated population average birth weight under a given intervention, denoted by $\hat{E}[Y^{x_{int},t_{int,k}}]$, is given by.

$$\hat{E}_m\left[Y_m^{x_{int},t_{int,k}}\right] = \qquad 1/N \sum_{i=1}^{N}\left[\hat{\mu}_{i,m}(X_{int,i}, T_{int,k})\right]$$

Under causal identification conditions given in the manuscript, correct model specification, and standard conditions of the validity of MCMC, the simulated values of $\hat{E}_m\left[Y_m^{x_{int},t_{int,k}}\right]$ constitute posterior draws from the distribution of the post-intervention, population average birth weight. The mean change in birth weight following the intervention was obtained by taking the difference $\hat{E}_m\left[Y_m^{x_{int},t_{int,k}}\right] - \hat{E}_m\left[Y_m^{x_{obs},t_{obs,k}}\right]$ (where "obs" denotes "observed). That is, the posterior distribution of our estimate of the effect of decommissioning is obtained by subtracting the population average of the predictions under the natural course from the population average of the predictions under the intervention.

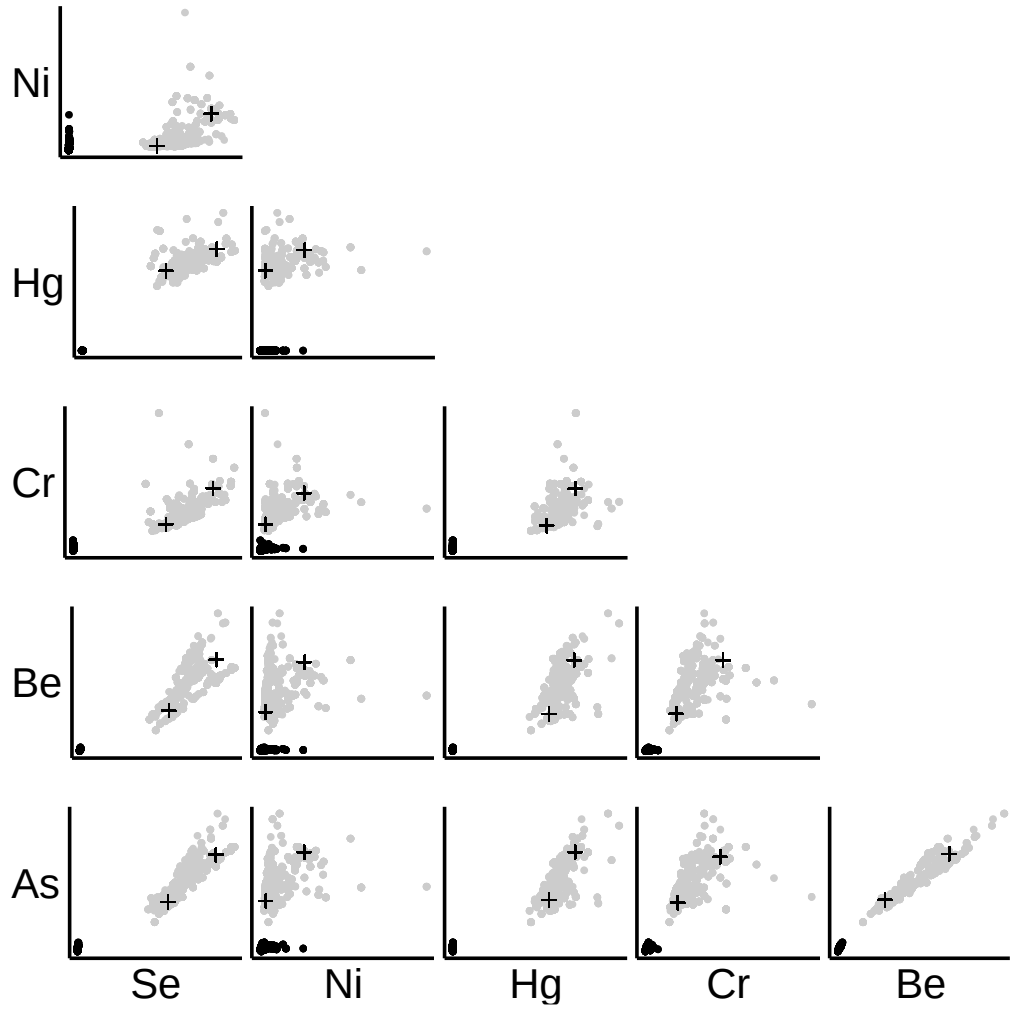
WEB APPENDIX 2
Sensitivity Analyses

We assessed the sensitivity of our inference to the following characteristics of our statistical model: the use of quadratic terms for maternal age, the priors underlying the hierarchical structure, the strength of model selection priors, and forcing certain variables into the model. First, we fit a model using main terms only (except for including a quadratic term for maternal age), to evaluate the impact of including many product terms in the model. Second, we assessed the impact of strengthening or weakening model selection priors by increasing or decreasing the prior precision on selection probabilities. We assessed sensitivity to use of quadratic terms for maternal age by fitting a model that included a smoothing spline for maternal age. We assessed sensitivity to our choice of prior for model coefficients by fitting a model that used a 3 degree of freedom t-distribution as the prior. We also assessed sensitivity to forcing certain terms into the model by fitting two separate models that forced inclusion of all main effects of potential confounders or all main effects of exposures and potential confounders. We fit three Bayesian models that did not include model averaging and selection, to evaluate the impact of Bayesian model averaging for allowing potentially highly flexible models in which the first model included hierarchical priors given in **Table 1** of the manuscript, the second relaxed priors on the hierarchy by using uniform priors on the shared variance terms of model coefficients, and the third model utilized independent normally distributed priors on all model coefficients. Lastly, we used non-Bayesian g-computation with identical model specification and bootstrap confidence intervals to assess any impacts of using the posterior mean to define point estimates, rather than the maximum likelihood value.

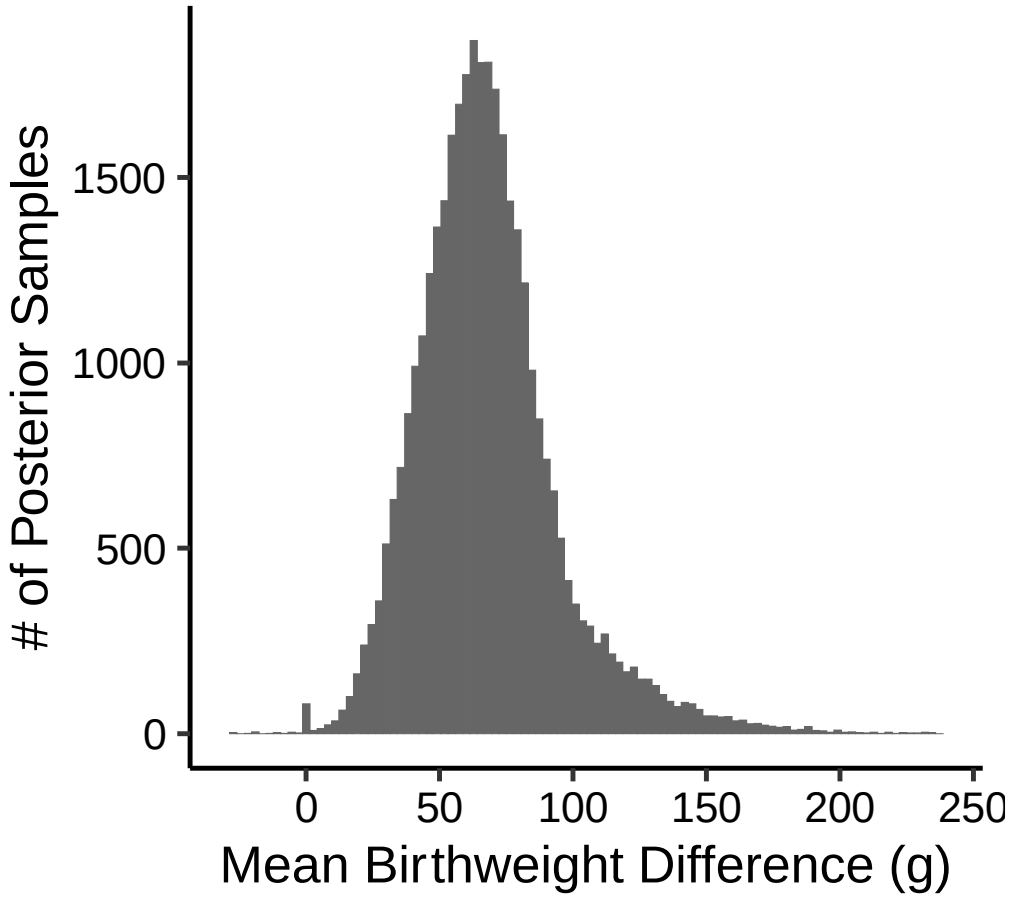
Assessing intervention exposure levels relative to the joint exposure range

The intervention values of exposure given in the manuscript are guaranteed to be outside of the range of exposures for at least some individuals, given that the intervention was operationalized by multiplying the observed exposures by a constant < 1.0. We assessed overlap between the observed exposures and the intervention levels via a bivariate scatter plot (**Web Figure 1**). Intervention values of Nickel were broadly

within range of the observed exposures, while all other exposures displayed some separation between the distributions of observed and intervention values.



Web Figure 1. Distribution of observed and intervention values of exposures (gray dots = observed/natural course; black dots = "decommissioning" intervention; black crosses = 10th and 90th percentiles of observed exposures). Under the natural course, the mean (SD) of exposures in units of 10ng/m^3 were: As=1.88 (0.46), Be=0.89 (0.26), Cr=1.06 (0.41), Hg=16.31 (1.84), Ni=20.56 (18.87), Se=5.13 (0.87); Under decommissioning, those means (SDs) would be: As=0.15 (0.04), Be=0.03 (0.01), Cr=0.10 (0.04), Hg=0.03 (0.00), Ni=5.44 (5.00), Se=0.06 (0.01).



Web Figure 2. Posterior distribution of the mean birth weight difference across live births in Milwaukee, WI (2011-2013) comparing the natural course to the hypothetical intervention to decommission 3 coal-fired power plants in Milwaukee County. The spike at zero represents the impact of Bayesian model averaging, which admits effect sizes of exactly zero.

WEB APPENDIX 3

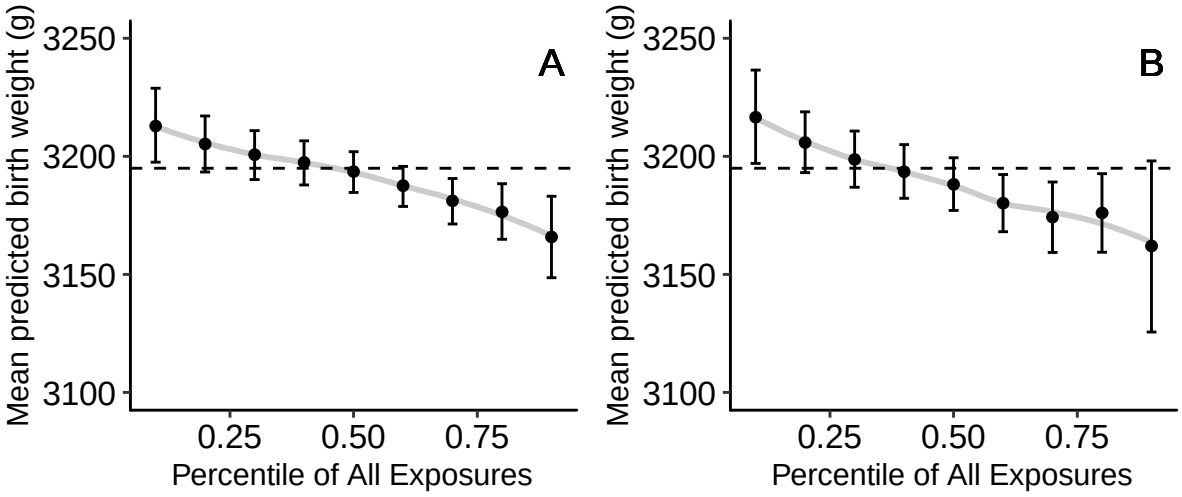
Joint Effects in the Joint Exposure Range

To address sensitivity of our (qualitative) results, we performed an additional analysis that restricted hypothetical interventions to exposure points within the range of the observed data. We note that this was not our primary approach because, while such an effect characterizes joint effects that are, in principle,

less prone to errors due to model misspecification, it does not map directly onto realistic interventions in a concrete way.

We operationalized this analysis by comparisons of 9 hypothetical interventions in which we set all exposure values in the population to a specific quantile of observed exposures. These interventions were at the 10^{th}, 20^{th}, …, 90^{th} percentile. Because these interventions were based on the same statistical model as the primary analysis, they could be evaluated by creating exposure/covariate (design) matrices for each intervention value and then making predictions using those design matrices and the posterior draws of model parameters obtained in the main analysis. For computational efficiency, we thinned the original 36,000 post burn-in draws by taking every 36^{th} draw and summarized over the resulting 1,000 posterior samples. We repeated this analysis for models A (main analysis, Bayesian model averaging) and J (no model averaging) from the main analysis.

The posterior distributions of predicted birth weight at each of the proposed interventions demonstrated that higher values of the joint exposure level coincided with lower predicted population average birth weight (**Web Figure 3**). At these hypothetical intervention levels, both Bayesian models produced similar inference with respect to the posterior mean of the population average birth weight. Credible intervals were wider for the model without Bayesian model averaging, but the differences in precision between these two models were markedly smaller than the analysis from the manuscript assessing the potential impact of coal plant decommissioning. This last result demonstrates that, when model extrapolation is reduced, our results appear more robust to the choice of which statistical model is used.



Web Figure 3. Posterior means and 95% credible intervals for assessing interventions to jointly set all exposures to observed quantiles contrasting two Bayesian statistical models. Panel A represents results using our Bayesian model averaging approach (model A) and Panel B represents results using our hierarchical Bayesian model with no selection (model J). Population mean given by dashed lines.

WEB APPENDIX 4

Simulation Study

We performed a small simulation study to examine whether Bayesian g-computation can be used to estimate effects of interventions in cases in which the expected levels of exposures in at least one level of the intervention are outside the support of the data, possibly in strata of covariates. This condition is sometimes referred to as "stochastic non-positivity" in which the intervention values have positive probability in the study sample. Gill and Robins refer to this assumption as "identifiability" (1) and note that it is interpreted as the condition that "when there was an opportunity to apply the [intervention], that opportunity was at least sometimes taken."

We first review notation for point-exposure and time-fixed outcome data such as the example given in the main text. We denote multivariate exposure as $\boldsymbol{X}$, multivariate covariate as $\boldsymbol{Z}$, and a univariate outcome as Y. Under the counterfactual in which we set exposure to some "intervention" or "plan" value of $\boldsymbol{g}$, the potential outcomes are denoted as $Y^{\boldsymbol{g}}$, which refer to the outcomes we would observe, if exposure had been set to $\boldsymbol{X}=\boldsymbol{g}$. For discrete $\boldsymbol{X}$, $\boldsymbol{Z}$ and Y, positivity or identifiability can be stated as $\Pr(\boldsymbol{X}=\boldsymbol{g}, \boldsymbol{Z}=\boldsymbol{z}) > 0$. This condition ensures that the conditional potential outcome probability $\Pr(Y^{\boldsymbol{g}} = y \mid \boldsymbol{X}=\boldsymbol{g}, \boldsymbol{Z}=\boldsymbol{z})$ is well defined in the sense that the scenario on which we condition is within the realm of possibility (which

ensures a unique value to the conditional probability). The extension to continuous $\boldsymbol{X}$, $\boldsymbol{Z}$ and Y given by Gill and Robins is defined by Gill and Robins to mean that the joint distribution of ($\boldsymbol{X}$, $\boldsymbol{Z}$) are inside the "support" of the joint distribution of ($\boldsymbol{X}$, $\boldsymbol{Z}$), where "support" in the continuous case refers to all values comprising the "close" neighborhood of observed covariates and exposures. Informally, this means that if we propose an intervention that sets exposures X_1 and X_2 to some values x_1 and x_2, then there should be some observations in the data where X_1 is "close" to x_1 and X_2 is "close" to x_2.

We next note that the assumption of "identifiability" may be relaxed by parametric modeling. "Identifiability" is non-parametric in the sense that it refers to scenarios in which the estimator of the distribution of potential outcomes is estimable from the data without assuming a parametric model. In practice when any member of $\boldsymbol{X}$, $\boldsymbol{Z}$ or Y are continuous, some sort of parametric or semi-parametric model is used, which imposes a constraint on the possible joint distribution of ($\boldsymbol{X}$, $\boldsymbol{Z}$, Y). In the point-exposure, discrete time-fixed outcome setting, this implies that $\Pr(Y^{\boldsymbol{g}} = y \mid \boldsymbol{X}=\boldsymbol{g}, \boldsymbol{Z}=\boldsymbol{z})$ is indexed by some model with finite dimensional parameters, such that the conditional distribution of potential outcomes is given by $\Pr(Y^{\boldsymbol{g}} = y \mid \boldsymbol{X}=\boldsymbol{g}, \boldsymbol{Z}=\boldsymbol{z}, \boldsymbol{\beta})$. We note here that this quantity is generally unique and identified when the finite dimensional parameter $\boldsymbol{\beta}$ is known (e.g. as would be the case with generalized linear models with known values of model parameters). In practice we almost never know $\boldsymbol{\beta}$, so we can substitute $\Pr(Y^{\boldsymbol{g}} = y \mid \boldsymbol{X}=\boldsymbol{g}, \boldsymbol{Z}=\boldsymbol{z}, \widehat{\boldsymbol{\beta}})$ for the potential outcome distribution using an estimate of the model parameters denoted by $\widehat{\boldsymbol{\beta}}$. Thus, rather than assuming positivity or non-parametric identifiability, we can assume that the parametric model is correctly specified in the sense that the true distribution $\Pr(Y^{\boldsymbol{g}} = y \mid \boldsymbol{X}=\boldsymbol{g}, \boldsymbol{Z}=\boldsymbol{z})$ is contained in the model $\Pr(Y^{\boldsymbol{g}} = y \mid \boldsymbol{X}=\boldsymbol{g}, \boldsymbol{Z}=\boldsymbol{z}, \boldsymbol{\beta})$ and, as is well known, consistent estimates of the potential outcome distribution can be obtained if we have a consistent estimator $\widehat{\boldsymbol{\beta}}$ (and all other necessary causal assumptions hold). While model specification can often be assessed empirically through goodness-of-fit measures, we note that the use here, a substitute for a positivity assumption, is less amenable to empirical checks. Namely, an estimate of $\boldsymbol{\beta}$ may appear to yield a model that fits the data well, but this provides no guarantee that accurate "local fit" provides an accurate global fit to regions of the joint distribution ($\boldsymbol{X}$, $\boldsymbol{Z}$) that are not observed in one's particular sample of data.

We illustrate that the model specification assumption can be substituted for positivity/identifiability in parametric models, including Bayesian linear models which were used in the main text. We simulated data according to the following structure. $\boldsymbol{Z}$ comprised three independent standard normal variables. $\boldsymbol{X}$ comprised two normally distributed variables (X_1, X_2) each with unit variance and means given by $15 + \boldsymbol{Z\alpha}$, where $\boldsymbol{\alpha}$ was the row vector (1,1,1). This induced a Spearman correlation between (X_1, X_2) approximately 0.8 so as to emulate correlated variables found in our example from the main text. A

univariate outcome Y was generated as a normally distributed variable with unit variance and mean given by $(X_1, \boldsymbol{Z}, \boldsymbol{Z}*\boldsymbol{Z}, X_1*\boldsymbol{Z})\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ was the row vector (1, 2, -2, 0, -0.3, 0.0, 0.5, 1.0, 1.0, 1.0) and "*" implies element-wise/row-wise multiplication. Thus, the assumed model implies that X_1 has a causal effect on Y, but X_2 does not, and on the additive scale the effect of X_1 depends on the confounders $\boldsymbol{Z}$, and the conditional regression function between $\boldsymbol{Z}$ and Y is a quadratic function. Data were simulated in sample sizes of 100 and 1000.

We analyzed the data using a Bayesian, hierarchical linear model in which we estimated the posterior distributions of linear model parameters for a correctly specified regression model given as the function $E(Y \mid X_1, X_2, \boldsymbol{Z}, \boldsymbol{Z}*\boldsymbol{Z}, X_1*\boldsymbol{Z}, X_2*\boldsymbol{Z}, \boldsymbol{\beta_c})$, where the terms for X_2 and $X_2*\boldsymbol{Z}$ were extraneous in the sense that the true coefficients corresponding to these terms are all zero (but they do not induced bias by including in the model). Thus, this analysis represents our default approach to modeling described in the main text, where we model deviations from the additive scale through product terms and quadratic basis functions of independent variables with the expectation that some of these coefficients are close to zero. We also estimated the parameters of a mis-specified model given by $E(Y \mid X_1, X_2, \boldsymbol{Z}, \boldsymbol{\beta_m})$, which includes only main terms of exposures and confounders, and roughly represents a default linear/additive model in which we assume that all product terms and non-linear bases are zero.

For analysis of both correctly and incorrectly specified models, we imposed a hierarchical model by assuming that the coefficients for the main exposure effects derived from a common normal distribution with priors $\boldsymbol{\beta_X} \sim Normal(\mu_1, \tau_1^2)$ and all other coefficients arose from a separate common normal distribution with priors $\boldsymbol{\beta_{ZX}} \sim Normal(\mu_2, \tau_2^2)$. We used conjugate, vague hyper priors on the prior distribution of $(\mu_1, \mu_2, \tau_1, \tau_2)$ to leverage Gibbs sampling for estimation, which allowed relatively quick analysis of the data and facilitated full Bayesian inference.

Following the main text, we used the estimated parameters from the linear model to estimate the effects of a hypothetical intervention to reduce $\boldsymbol{X}$ from its observed values $\boldsymbol{x} = (x_1, x_2)$ to (1.0, 1.0) for all individuals, which estimates the potential mean difference given by $E(Y^{x} - Y^{(1,1)})$. We compared the estimated mean potential outcomes and the estimated mean difference to the expected mean difference for each dataset in terms of bias (mean difference between the point estimate of the mean potential outcome and the true mean potential outcome), Monte Carlo Standard deviation (standard deviation of the point estimates) and root-mean squared error (the average squared bias). The true values of the population average effect parameters are given as: $E[Y^{(15,15)}] = 14.55$ and $E[Y^{(15,15)}] = 0.55$. Results are given and discussed in the main text.

Web Table 1. Bayesian prior distributions for main analysis and selected sensitivity analyses for the Bayesian linear model of birth weight as a function of metals and confounders

	Model[a]									
Term[b]	A,E[c]	B	C	D	F	G,H[d]	I	J	K	L
β_0	$N(0,100)$	$N(0,100)$	$N(0,100)$	$N(0,100)$	$N(0,100)$	$N(0,100)$	$N(0,100)$	$N(0,100)$	$N(0,100)$	$N(0,100)$
$\beta_1 - \beta_{14}$	$N(\mu_1, \tau_1^2)$	$N(\mu_1, \tau_1^2)$	$N(\mu_1, \tau_1^2)$	$N(\mu_1, \tau_1^2)$	$T(3, \mu_1, \tau_1^2)$	$N(\mu_1, \tau_1^2)$	$N(0,1)$	$N(\mu_1, \tau_1^2)$	$N(\mu_1, \tau_1^2)$	$N(0,1)$
$\beta_{15} - \beta_{83}$	$N(\mu_2, \tau_2^2)$		$N(\mu_2, \tau_2^2)$	$N(\mu_2, \tau_2^2)$	$T(3, \mu_2, \tau_2^2)$	$N(\mu_2, \tau_2^2)$		$N(\mu_2, \tau_2^2)$	$N(\mu_2, \tau_2^2)$	$N(0,1)$
$\delta_1 - \delta_{14}$	$B(1,1)$	$B(1,1)$	$B(1,1)$	$B(1,1)$	$B(1,1)$					
$\delta_{15} - \delta_{83}$	$B(9,1)$		$B(1,1)$	$B(90,10)$	$B(9,1)$	$B(9,1)$				
μ_1, μ_2	$N(0,1)$	$N(0,1)$	$N(0,1)$	$N(0,1)$	$N(0,1)$	$N(0,1)$		$N(0,1)$	$N(0,1)$	
τ_1	$C^+(0,2)$	$C^+(0,2)$	$C^+(0,2)$	$C^+(0,2)$	$C^+(0,2)$	$C^+(0,2)$		$IG(.001,.001)$	$U(-\infty,\infty)$	
τ_2	$C^+(0,1)$		$C^+(0,1)$	$C^+(0,1)$	$C^+(0,1)$	$C^+(0,1)$		$IG(.001,.001)$	$U(-\infty,\infty)$	
σ	$C^+(0,2)$	$C^+(0,2)$	$C^+(0,2)$	$C^+(0,2)$	$C^+(0,2)$	$C^+(0,2)$	$IG(.001,.001)$	$IG(.001,.001)$	$IG(.001,.001)$	$IG(.001,.001)$

Prior distributions denoted are as follows: $\boldsymbol{N}$ = normal(mean, variance), $\boldsymbol{B}$=beta(a,b), $\boldsymbol{C^+}$= Cauchy(location, scale) truncated below at 0, T(d.f., location, squared-scale) = non-central t distribution with 3 degrees of freedom, $\boldsymbol{IG}$ = inverse gamma(a,b), *U*=uniform(min, max)

[a]Corresponding to model label in Table 3

[b]$\boldsymbol{\beta}$: model coefficients; $\boldsymbol{\delta}$: selection indicators; $\boldsymbol{\mu}$: model coefficient prior mean; $\boldsymbol{\tau^2}$: model coefficient prior variance; $\boldsymbol{\sigma^2}$: model error term prior variance

[c]Model E includes the smoothing spline terms for maternal age in the same group as the main term coefficients 1-14.

[d]Model G includes selection on main exposure variables, whereas model H has no selection on the main term coefficients 1-14.

Web Table 2. Model coefficient (mean difference in birth weight (g) per unit change in independent variable) for all terms in the Bayesian model-averaging g-computation approach

Term	Mean[a] (95% CrI[b])	PIP[c]	GPIP[d]
Intercept	3179.4 (3161.9, 3196.7)		
cen_magebirth (maternal age at birth, standardized)	42.3 (32.9, 51.5)	1.00	1.00
cfemale (child female sex)	-116.5 (-130.9, -102.0)	1.00	1.00
mwhite (maternal white race, non-Hispanicity)	266.0 (244.2, 287.4)	1.00	1.00
mhisp (maternal Hispanicity)	25.3 (0.0, 77.6)	0.59	0.60
mother (maternal non-black, non-white race)	206.6 (182.3, 229.1)	1.00	1.00
msmk (maternal smoking during pregnancy)	-191.8 (-211.8, -171.9)	1.00	1.00
married (maternal marital status)	12.8 (0.0, 36.2)	0.60	0.60
cen_hg (mercury compounds, standardized)	-0.7 (-10.2, 0.0)	0.11	0.11
cen_se (selenium compounds, standardized)	-1.4 (-17.1, 0.0)	0.16	0.17
cen_be (beryllium compounds, standardized)	-6.2 (-23.2, 0.0)	0.48	0.48
cen_as (arsenic compounds, standardized)	-4.0 (-21.3, 11.5)	0.39	0.40
cen_cr (hexavalent chromium compounds, standardized)	-7.9 (-20.2, 0.0)	0.64	0.68
cen_ni (nickel compounds, standardized)	1.2 (0.0, 13.4)	0.15	0.22
cen_magebirth*cen_magebirth	-25.1 (-31.3, -18.8)	1.00	
cen_magebirth*cfemale	0.0 (0.0, 0.0)	0.00	
cen_magebirth*mwhite	0.0 (0.0, 0.0)	0.00	
cen_magebirth*mhisp	0.0 (0.0, 0.0)	0.00	
cen_magebirth*mother	0.2 (0.0, 0.0)	0.01	
cen_magebirth*msmk	-54.1 (-75.6, -31.9)	0.99	
cen_magebirth*mmarried	0.0 (0.0, 0.0)	0.00	
cen_hg*cen_magebirth	0.0 (0.0, 0.0)	0.00	
cen_magebirth*cen_se	0.0 (0.0, 0.0)	0.00	
cen_be*cen_magebirth	0.0 (0.0, 0.0)	0.00	
cen_as*cen_magebirth	0.0 (0.0, 0.0)	0.00	
cen_cr*cen_magebirth	0.0 (0.0, 0.0)	0.00	
cen_magebirth*cen_ni	0.0 (0.0, 0.0)	0.00	
cen_hg*cfemale	0.0 (0.0, 0.0)	0.00	
cen_hg*mwhite	0.0 (0.0, 0.0)	0.00	
cen_hg*mhisp	0.0 (0.0, 0.0)	0.00	
cen_hg*mother	0.0 (0.0, 0.0)	0.00	
cen_hg*msmk	0.0 (0.0, 0.0)	0.00	
cen_hg*mmarried	0.0 (0.0, 0.0)	0.00	
cen_hg*cen_hg	0.0 (0.0, 0.0)	0.00	
cen_hg*cen_se	0.0 (0.0, 0.0)	0.00	
cen_be*cen_hg	0.0 (0.0, 0.0)	0.00	

cen_as*cen_hg	0.0 (0.0, 0.0)	0.00
cen_cr*cen_hg	0.0 (0.0, 0.0)	0.00
cen_hg*cen_ni	0.0 (0.0, 0.0)	0.00
cen_se*cfemale	0.0 (0.0, 0.0)	0.00
cen_se*mwhite	0.0 (0.0, 0.0)	0.00
cen_se*mhisp	0.2 (0.0, 0.0)	0.01
cen_se*mother	0.1 (0.0, 0.0)	0.00
cen_se*msmk	0.0 (0.0, 0.0)	0.00
cen_se*mmarried	0.0 (0.0, 0.0)	0.00
cen_se*cen_se	0.0 (0.0, 0.0)	0.00
cen_be*cen_se	0.0 (0.0, 0.0)	0.00
cen_as*cen_se	0.0 (0.0, 0.0)	0.00
cen_cr*cen_se	0.0 (0.0, 0.0)	0.00
cen_ni*cen_se	0.0 (0.0, 0.0)	0.00
cen_be*cfemale	0.0 (0.0, 0.0)	0.00
cen_be*mwhite	0.0 (0.0, 0.0)	0.00
cen_be*mhisp	0.1 (0.0, 0.0)	0.00
cen_be*mother	0.0 (0.0, 0.0)	0.00
cen_be*msmk	0.0 (0.0, 0.0)	0.00
cen_be*mmarried	0.0 (0.0, 0.0)	0.00
cen_be*cen_be	0.0 (0.0, 0.0)	0.00
cen_as*cen_be	0.0 (0.0, 0.0)	0.00
cen_be*cen_cr	0.0 (0.0, 0.0)	0.00
cen_be*cen_ni	0.0 (0.0, 0.0)	0.00
cen_as*cfemale	0.0 (0.0, 0.0)	0.00
cen_as*mwhite	0.0 (0.0, 0.0)	0.00
cen_as*mhisp	0.3 (0.0, 0.0)	0.01
cen_as*mother	0.0 (0.0, 0.0)	0.00
cen_as*msmk	0.0 (0.0, 0.0)	0.00
cen_as*mmarried	0.0 (0.0, 0.0)	0.00
cen_as*cen_as	0.0 (0.0, 0.0)	0.00
cen_as*cen_cr	0.0 (0.0, 0.0)	0.00
cen_as*cen_ni	0.0 (0.0, 0.0)	0.00
cen_cr*cfemale	0.0 (0.0, 0.0)	0.00
cen_cr*mwhite	-1.4 (-30.0, 0.0)	0.04
cen_cr*mhisp	0.0 (0.0, 0.0)	0.00
cen_cr*mother	0.0 (0.0, 0.0)	0.00
cen_cr*msmk	0.0 (0.0, 0.0)	0.00
cen_cr*mmarried	0.0 (0.0, 0.0)	0.00
cen_cr*cen_cr	0.0 (0.0, 0.0)	0.00

cen_cr*cen_ni	0.0 (0.0, 0.0)	0.00
cen_ni*cfemale	0.0 (0.0, 0.0)	0.00
cen_ni*mwhite	-0.3 (0.0, 0.0)	0.01
cen_ni*mhisp	0.0 (0.0, 0.0)	0.00
cen_ni*mother	1.6 (0.0, 24.7)	0.07
cen_ni*msmk	0.0 (0.0, 0.0)	0.00
cen_ni*mmarried	0.0 (0.0, 0.0)	0.00
cen_ni*cen_ni	0.0 (0.0, 0.0)	0.00

[a]Mean = posterior mean of $\delta * \beta$ terms from model (1) in the manuscript
[b]CrI = credible intervals based on quantiles of posterior (Note: these may not contain the posterior mean for terms where the PIP is very low)
[c]PIP = posterior inclusion probability; posterior mean of δ from model (1) in the manuscript
[d]GPIP = groupwise posterior inclusion probability for any main or product term including the referenced variable
Bayesian Model-Averaging estimates given as the posterior mean and 95% CrI over 36000 MCMC iterations across 8 independent chains. Variables included in the model included standardized (to have mean=0, SD=1) variables of each metal (continuous: cen_as, cen_be, cen_cr, cen_hg, cen_sb, cen_ni), standardized maternal age at birth (continuous: cen_magebirth), maternal race (4 categories: mwhite= white, non-Hispanic, mother = any other race, non-Hispanic, mhisp = white, Hispanic and black, non-Hispanic as referent), maternal smoking during pregnancy (yes/no: msmk), maternal marital status (yes/no: mmarried), and child's female sex (yes/no: cfemale)

Web Table 3. Adjusted Change in Birth Weight (g) with Airborne Metals from Single-Pollutant (non-Bayesian) Models, Only Metal Terms Shown

Term	MD[a] (95% CI[b])
Mercury Compounds (Hg)	-13.0 (-20.4, -5.6)
Selenium Compounds (Se)	-16.7 (-24.7, -8.7)
Beryllium Compounds (Be)	-16.2 (-23.7, -8.7)
Arsenic Compounds (As)	-16.8 (-24.4, -9.2)
Hexavalent Chromium (Cr)	-16.2 (-23.9, -8.5)
Nickel Compounds (Ni)	-0.54 (-9.0, 7.9)

Variables included in each model included standardized (to have mean=0, SD=1) variables of each metal, standardized maternal age at birth (continuous), maternal race (4 categories: white, non-Hispanic, white and Hispanic, black and non-Hispanic, any other race and non-Hispanic), maternal smoking during pregnancy (yes/no), maternal marital status (yes/no), and child's female sex (yes/no)

[a]MD = mean difference (in birth weight)

[b]CI = confidence intervals

Web Table 4. Adjusted Change in Birth Weight (g) with Airborne Metals from Multi-Pollutant (non-Bayesian) Model, Only Metal Terms Shown

Term	MD[a] (95% CI[b])
Mercury Compounds (Hg)	-5.4 (-16, 4.7)
Selenium Compounds (Se)	-13 (-33, 7)
Beryllium Compounds (Be)	-19 (-50, 12)
Arsenic Compounds (As)	18 (-26, 62)
Hexavalent Chromium (Cr)	-13 (-22,-3.8)
Nickel Compounds (Ni)	7.8 (-2.3, 18)

Variables included in each model included standardized (to have mean=0, SD=1) variables of each metal, standardized maternal age at birth (continuous), maternal race (4 categories: white, non-Hispanic, white and Hispanic, black and non-Hispanic, any other race and non-Hispanic), maternal smoking during pregnancy (yes/no), maternal marital status (yes/no), and child's female sex (yes/no)

[a]MD = mean difference (in birth weight)

[b]CI = confidence intervals

Web Table 5. Model coefficient (mean difference in birth weight (g) per unit change in independent variable) for all terms using standard Bayesian hierarchical modeling without Bayesian model averaging

Term	Mean[a] (95% CrI[b])
Intercept	3175.1 (3155.5, 3194.7)
cen_magebirth (maternal age at birth, standardized)	33.6 (23.1, 44.3)
cfemale (child female sex)	-116.1 (-130.5, -101.5)
mwhite (maternal white, non-Hispanic race/ethnicity)	263.3 (241.1, 285.6)
mhisp (maternal Hispanic ethnicity)	42.6 (2.9, 82.3)
mother (maternal non-black, non-white race)	194.7 (170.0, 219.3)
msmk (maternal smoking during pregnancy)	-187.1 (-207.1, -167.2)
married (maternal marital status)	22.4 (5.7, 39.2)
cen_hg (mercury compounds, standardized)	-6.7 (-21.1, 7.8)
cen_se (selenium compounds, standardized)	-13.0 (-37.5, 11.3)
cen_be (beryllium compounds, standardized)	-10.5 (-51.0, 29.8)
cen_as (arsenic compounds, standardized)	3.2 (-49.3, 55.5)
cen_cr (hexavalent chromium compounds, standardized)	-14.2 (-32.3, 4.3)
cen_ni (nickel compounds, standardized)	12.0 (-8.0, 32.1)
cen_magebirth*cen_magebirth	-24.8 (-31.2, -18.4)
cen_magebirth*cfemale	-0.4 (-9.4, 8.4)
cen_magebirth*mwhite	-2.9 (-13.8, 6.6)
cen_magebirth*mhisp	0.9 (-10.3, 12.5)
cen_magebirth*mother	6.3 (-3.2, 18.1)
cen_magebirth*msmk	-11.0 (-26.8, 0.2)
cen_magebirth*mmarried	0.0 (-9.6, 9.4)
cen_hg*cen_magebirth	-0.1 (-7.3, 7.0)
cen_magebirth*cen_se	2.1 (-6.3, 10.8)
cen_be*cen_magebirth	-2.7 (-11.8, 5.6)
cen_as*cen_magebirth	-1.4 (-11.5, 8.0)
cen_cr*cen_magebirth	3.1 (-3.6, 10.1)
cen_magebirth*cen_ni	1.7 (-5.1, 8.4)
cen_hg*cfemale	-0.1 (-9.6, 9.1)
cen_hg*mwhite	2.3 (-7.4, 13.3)
cen_hg*mhisp	1.8 (-9.2, 13.8)
cen_hg*mother	-0.2 (-10.7, 9.8)
cen_hg*msmk	1.3 (-8.7, 12.1)
cen_hg*mmarried	-0.1 (-10.0, 9.5)
cen_hg*cen_hg	2.5 (-2.4, 7.5)
cen_hg*cen_se	0.1 (-8.8, 9.2)

cen_be*cen_hg	-0.6 (-10.1, 8.0)
cen_as*cen_hg	0.5 (-9.3, 10.5)
cen_cr*cen_hg	2.0 (-6.7, 11.5)
cen_hg*cen_ni	1.0 (-7.3, 9.7)
cen_se*cfemale	2.2 (-7.5, 12.6)
cen_se*mwhite	0.6 (-9.8, 11.6)
cen_se*mhisp	2.6 (-8.4, 15.2)
cen_se*mother	5.5 (-4.4, 17.8)
cen_se*msmk	0.3 (-10.1, 11.3)
cen_se*mmarried	2.9 (-6.8, 13.9)
cen_se*cen_se	-1.2 (-10.0, 7.5)
cen_be*cen_se	-4.7 (-17.0, 5.3)
cen_as*cen_se	-3.4 (-15.2, 6.6)
cen_cr*cen_se	1.0 (-7.6, 10.0)
cen_ni*cen_se	2.1 (-7.7, 12.4)
cen_be*cfemale	0.4 (-9.5, 10.4)
cen_be*mwhite	3.1 (-6.7, 14.6)
cen_be*mhisp	2.1 (-9.1, 14.4)
cen_be*mother	-1.1 (-12.4, 9.1)
cen_be*msmk	-1.2 (-12.2, 8.9)
cen_be*mmarried	0.1 (-10.2, 10.1)
cen_be*cen_be	2.9 (-5.8, 13.6)
cen_as*cen_be	0.7 (-9.2, 11.4)
cen_be*cen_cr	-2.4 (-13.2, 7.0)
cen_be*cen_ni	1.2 (-8.9, 11.5)
cen_as*cfemale	0.5 (-9.9, 10.8)
cen_as*mwhite	0.5 (-10.0, 11.4)
cen_as*mhisp	2.6 (-8.5, 15.1)
cen_as*mother	1.1 (-9.8, 11.8)
cen_as*msmk	-1.8 (-13.3, 8.5)
cen_as*mmarried	0.0 (-10.8, 10.3)
cen_as*cen_as	-0.7 (-10.9, 9.2)
cen_as*cen_cr	0.3 (-9.8, 10.5)
cen_as*cen_ni	1.5 (-7.9, 10.9)
cen_cr*cfemale	1.2 (-7.9, 10.5)
cen_cr*mwhite	-4.9 (-17.4, 5.0)
cen_cr*mhisp	1.5 (-9.9, 13.5)
cen_cr*mother	3.7 (-6.1, 14.7)

cen_cr*msmk	-1.3 (-11.8, 8.5)
cen_cr*mmarried	0.6 (-9.1, 10.3)
cen_cr*cen_cr	0.9 (-3.8, 5.4)
cen_cr*cen_ni	3.1 (-6.7, 14.2)
cen_ni*cfemale	1.7 (-7.3, 11.1)
cen_ni*mwhite	-5.4 (-17.4, 4.4)
cen_ni*mhisp	1.4 (-9.7, 13.1)
cen_ni*mother	6.8 (-2.8, 18.6)
cen_ni*msmk	-1.8 (-12.6, 8.2)
cen_ni*mmarried	0.8 (-8.6, 10.3)
cen_ni*cen_ni	-1.9 (-5.8, 1.9)

[a]Mean = posterior mean of $\beta * \delta$ terms from model (1) in the manuscript (with $\delta = 1$ for all terms in this model)

[b]CrI = credible intervals based on quantiles of posterior (Note: these may not contain the posterior mean for terms where the PIP is very low)

[c]PIP = posterior inclusion probability posterior mean of δ from model (1) in the manuscript

Estimates given as the posterior mean and 95% CrI over 36000 MCMC iterations across 8 independent chains. Variables included in the model included standardized (to have mean=0, SD=1) variables of each metal (continuous: cen_as, cen_be, cen_cr, cen_hg, cen_sb, cen_ni), standardized maternal age at birth (continuous: cen_magebirth), maternal race (4 categories: mwhite= white, non-Hispanic, mother = any other race, non-Hispanic, mhisp = white, Hispanic and black, non-Hispanic as referent), maternal smoking during pregnancy (yes/no: msmk), maternal marital status (yes/no: mmarried), and child's female sex (yes/no: cfemale)