ORIGINAL ARTICLE

Modeling the health effects of time-varying complex environmental mixtures: Mean field variational Bayes for lagged kernel machine regression

Shelley H. Liu[1] | **Jennifer F. Bobb**[2] | **Birgit Claus Henn**[3] | **Lourdes Schnaas**[4] | **Martha Tellez-Rojo**[5] | **Chris Gennings**[1] | **Manish Arora**[1] | **Robert O. Wright**[1] | **Brent A. Coull**[6] | **Matt P. Wand**[7]

SUPPLEMENTARY MATERIALS

R code: The R code for fitting MFVB for lagged kernel machine regression is provided in our Github account. Code for a simulation study can be found here (https://github.com/shelleyhliu/VB-LKMR-Simulations), and code for a case study can be found here (https://github.com/shelleyhliu/VB-LKMR-CaseStudy).

Supplementary Materials Part A: Derivations

$$\Sigma_h^{-1} = \begin{bmatrix} \frac{G_{1,11}}{\tau_1^2} + \frac{1}{\omega_1^2} & \frac{G_{1,12}}{\tau_1^2} & \cdots & \frac{G_{1,1n}}{\tau_1^2} & \frac{-1}{\omega_1^2} & 0 & \cdots & 0 & \cdots \\ \frac{G_{1,12}}{\tau_1^2} & \frac{G_{1,22}}{\tau_1^2} + \frac{1}{\omega_1^2} & \cdots & \frac{G_{1,2n}}{\tau_1^2} & 0 & \frac{-1}{\omega_1^2} & 0 & \cdots & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots & \cdots \\ \frac{G_{1,1n}}{\tau_1^2} & \frac{G_{1,2n}}{\tau_1^2} & \cdots & \frac{G_{1,nn}}{\tau_1^2} + \frac{1}{\omega_1^2} & 0 & 0 & 0 & \frac{-1}{\omega_1^2} & \cdots \\ \frac{-1}{\omega_1^2} & 0 & \cdots & 0 & \frac{G_{2,11}}{\tau_2^2} + \frac{1}{\omega_1^2} + \frac{1}{\omega_2^2} & \frac{G_{2,12}}{\tau_2^2} & \cdots & \frac{G_{2,1n}}{\tau_2^2} & \ddots \\ 0 & \frac{-1}{\omega_1^2} & \ddots & \vdots & \frac{G_{2,12}}{\tau_2^2} & \frac{G_{2,22}}{\tau_2^2} + \frac{1}{\omega_1^2} + \frac{1}{\omega_2^2} & \cdots & \frac{G_{2,2n}}{\tau_2^2} & \cdots \\ \vdots & \vdots & \ddots & 0 & \vdots & \vdots & \vdots & \vdots & \cdots \\ 0 & 0 & \cdots & \frac{-1}{\omega_1^2} & \frac{G_{2,1n}}{\tau_2^2} & \frac{G_{2,2n}}{\tau_2^2} & \cdots & \frac{G_{2,nn}}{\tau_2^2} + \frac{1}{\omega_1^2} + \frac{1}{\omega_2^2} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

SUPPLEMENTARY FIGURE S1 Form of covariance matrix Σ_h^{-1}

Form of $log p(\underline{\boldsymbol{y}}; q)$.

$$log p(\underline{\boldsymbol{y}}; q) = \frac{-n}{2} log(2\pi) - \frac{1}{2} log(|\mu_{q(\Sigma^{-1}_{\tau^2,\omega^2})}|) - \frac{1}{2}\mu^T_{q(h)}\mu_{q(\Sigma^{-1}_{\tau^2,\omega^2})}\mu_{q(h)} + \left\{\mu_{q(\frac{1}{\sigma^2})} W^T W + \mu_{q(\Sigma^{-1}_{\tau^2,\omega^2})}\right\}^{-1} \tag{1}$$

$$+\frac{1}{2}\left\{\left\{\mu_{q(h)} - \left\{\mu_{q(\frac{1}{\sigma^2})} W^T W + \mu_{q(\Sigma^{-1}_{\tau^2,\omega^2})}\right\}^{-1} \mu_{q(\frac{1}{\sigma^2})}) W^T (Y - X\mu_q\beta)\right\}^T \left\{\mu_{q(\frac{1}{\sigma^2})} W^T W + \mu_{q(\Sigma^{-1}_{\tau^2,\omega^2})}\right\}^{-1} \left\{\mu_{q(h)} - \left\{\mu_{q(\frac{1}{\sigma^2})} W^T W + \mu_{q(\Sigma^{-1}_{\tau^2,\omega^2})}\right\}^{-1} \mu_{q(\frac{1}{\sigma^2})}) W^T (Y - X\mu_q\beta)\right\}\right\} \tag{2}$$

$$\frac{p}{2} log(2\pi) + log(|(\mu_{q(\frac{1}{\sigma^2})} X^T X)^{-1}|)^{\frac{1}{2}}) + 2\mu_{q(\frac{1}{\sigma^2})}[\mu_q(\beta) - (X^T X)^{-1} X^T (Y - W\mu_{q(h)}]^T (X^T X)^{-1} [\mu_q(\beta) - (X^T X)^{-1} X^T (Y - W\mu_{q(h)}] \tag{3}$$

$$a log(\gamma) - log(\Gamma(a)) + log(\frac{n}{2} + a) - (\frac{n}{2} + a) log\left\{\frac{(Y - W\mu_{q(h)} - X\mu_q\beta)^T (Y - W\mu_{q(h)} - X\mu_q\beta)}{2} + \gamma\right\} + \frac{NT}{2} log(\frac{\mu_q(\lambda_1^2)}{2}) \tag{4}$$

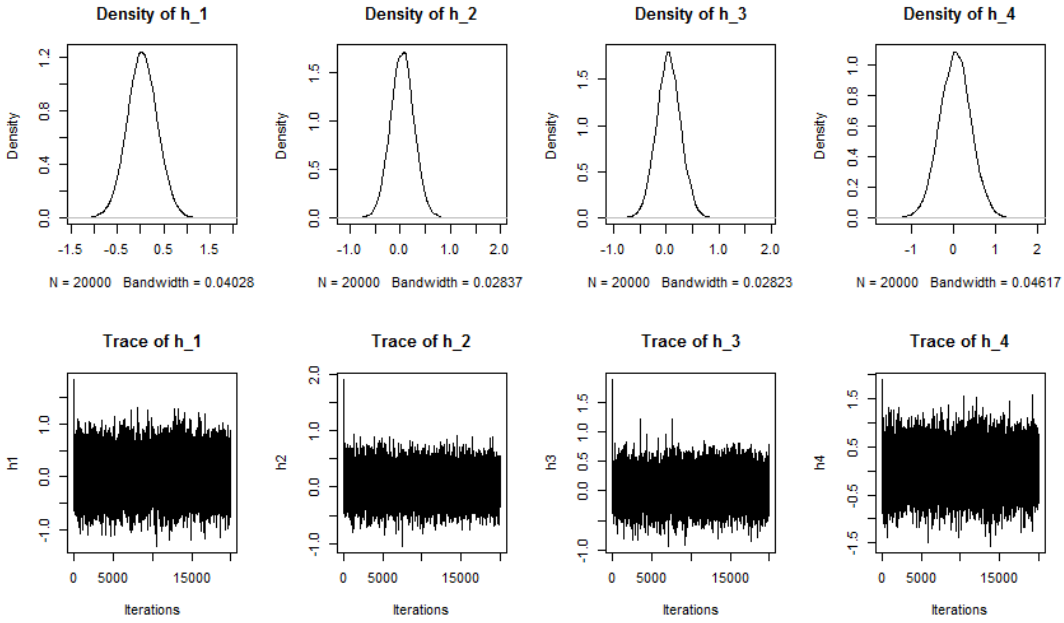
$$-T log(\Gamma((N+1)/2)) + \frac{T}{2} log(\pi) + N \sum_{t=1}^{T} log(\mu^2_{q(\tau_t^2)}) - \frac{\mu_q(\lambda_1^2)}{2} \sum_{t=1}^{T} \mu_{q(\tau_t^2)} + \frac{\mu_q(\lambda_1^2)}{2} \sum_{t=1}^{T} \frac{\left\{\frac{1}{\mu_{q(\tau_t^2)}} - (\frac{\mu_q(\lambda_1^2)}{\mu_{q(||h_t||^2_{Q_t})}})^{1/2}\right\}^2}{(\frac{\mu_q(\lambda_1^2)}{\mu_{q(||h_t||^2_{Q_t})}})(\frac{1}{\mu_{q(\tau_t^2)}})} + \frac{T-1}{2} log(\frac{\mu_q(\lambda_2^2)}{2}) \tag{5}$$

$$+\frac{T-1}{2} log(\pi) + \sum_{t=1}^{T-1} \frac{log(\mu_{q(\omega_t^2)})}{2} - \frac{\mu_q(\lambda_2^2)}{2} \sum_{t=1}^{T-1} \mu_{q(\omega_t^2)} + \frac{\mu_q(\lambda_2^2)}{2} \sum_{t=1}^{T-1} \frac{\left\{\frac{1}{\mu_{q(\omega_t^2)}} - (\frac{\mu_q(\lambda_1^2)}{\mu_{q(\sum_{n=1}^{N}(h_{t+1,n} - h_{t,n})^2)}})^{1/2}\right\}^2}{(\frac{\mu_q(\lambda_1^2)}{\mu_{q(\sum_{n=1}^{N}(h_{t+1,n} - h_{t,n})^2)}})(\frac{1}{\mu_{q(\omega_t^2)}})} \tag{6}$$

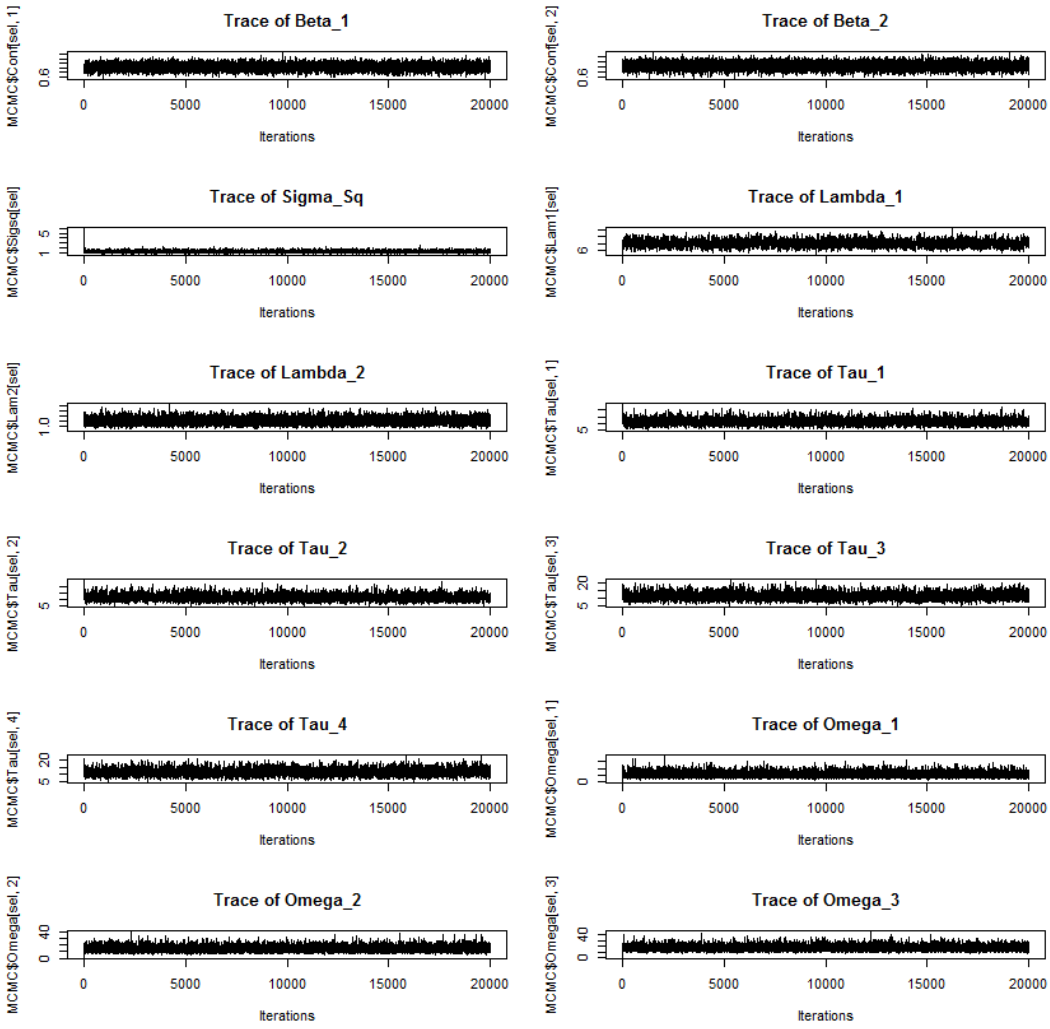
$$+r_2 log(\delta_2) - log(\Gamma(r_2)) + log(\Gamma(T - 1 + r_2)) - (T - 1 + r_2) log\left\{\sum_{t=1}^{T-1} \frac{\mu_{q(\omega_t^2)}}{2} + \delta_2\right\} - (T-1) log(\mu_q(\lambda_2^2)) + \mu_q(\lambda_2^2) \sum_{t=1}^{T-1} \frac{\mu_{q(\omega_t^2)}}{2} \tag{7}$$

$$+r_1 log(\delta_1) - log(\Gamma(r_1)) + log(\Gamma(\frac{T(N+1)}{2} + r_1)) - (\frac{T(N+1)}{2} + r_1) log\left\{\sum_{t=1}^{T} \frac{\mu_{q(\tau_t^2)}}{2} + \delta_1\right\} - \frac{T(N+1)}{2} log(\mu_q(\lambda_1^2)) + \mu_q(\lambda_1^2) \sum_{t=1}^{T} \frac{\mu_{q(\tau_t^2)}}{2} \tag{8}$$

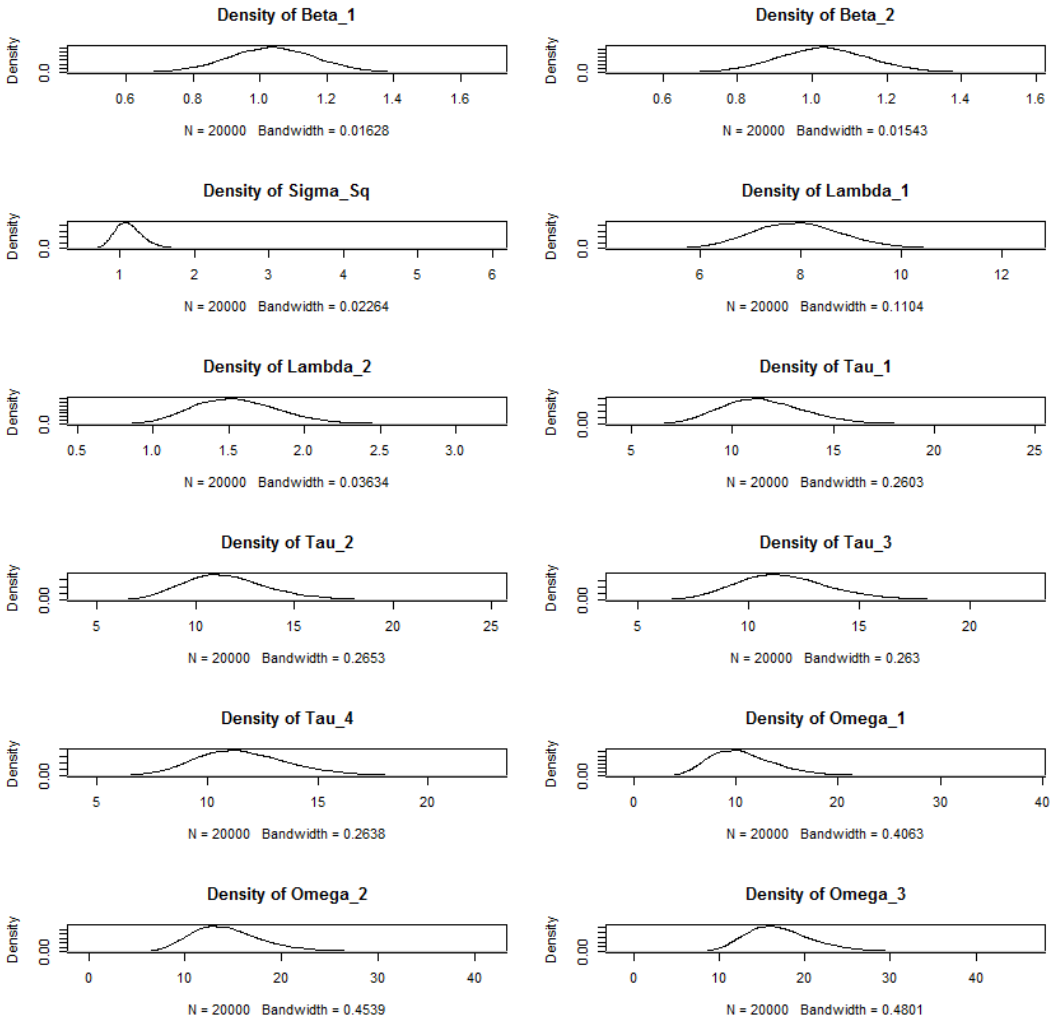
Supplementary Materials Part B: MCMC Diagnostics



SUPPLEMENTARY FIGURE S2a Marginal density plots and trace plots for estimation of $\boldsymbol{h}$ for a random representative dataset from the simulation study using MCMC-LKMR. We present the mean of $\boldsymbol{h}_t$, t= 1,2,3,4 for each iteration of the Gibbs sampler.



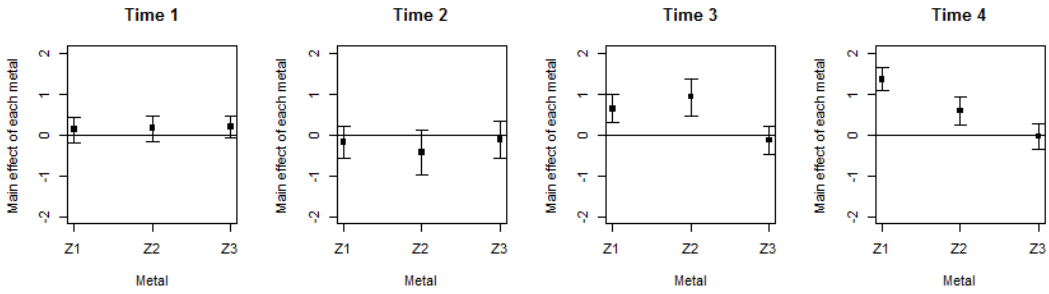
SUPPLEMENTARY FIGURE S2b Trace plots for estimation of $\beta, \sigma^2, \boldsymbol{\lambda}^2, \boldsymbol{\omega}^2, \boldsymbol{\tau}^2$ for a random representative dataset from the simulation study using MCMC-LKMR.



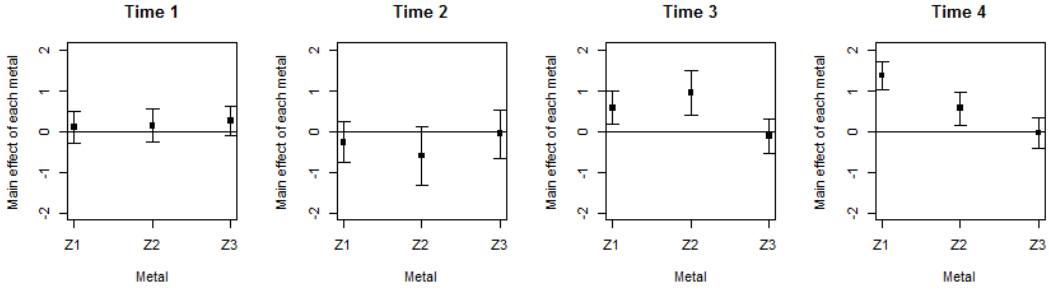
SUPPLEMENTARY FIGURE S2c Marginal density plots for estimation of $\beta, \sigma^2, \boldsymbol{\lambda}^2, \boldsymbol{\omega}^2, \boldsymbol{\tau}^2$ for a random representative dataset from the simulation study using MCMC-LKMR.

Supplementary Materials Part C: MFVB vs. MCMC Simulations

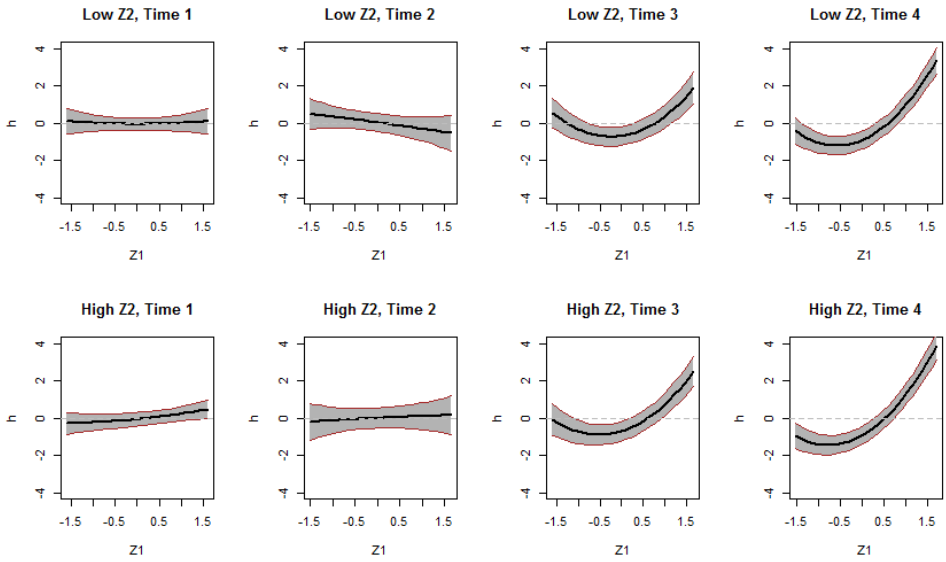
Because we note that σ^2 is estimated to be smaller under VB than under MCMC, we expanded the simulation study of N=300 to explore this impact on the ensuing inference. In particular, we create plots which are also used in the Application section to understand the relative importance of each time-varying mixture component, as well as the interaction between two mixture components across time windows, comparing inference under MFVB-LKMR and MCMC-LKMR. In each of these plots, we see that the confidence bars or confidence bands are slightly narrower for VB than for MCMC. However, under this simulating example, we see that this slight narrowing does not lead to a change in inference on the estimated exposure-response relationship across the two estimation methods.



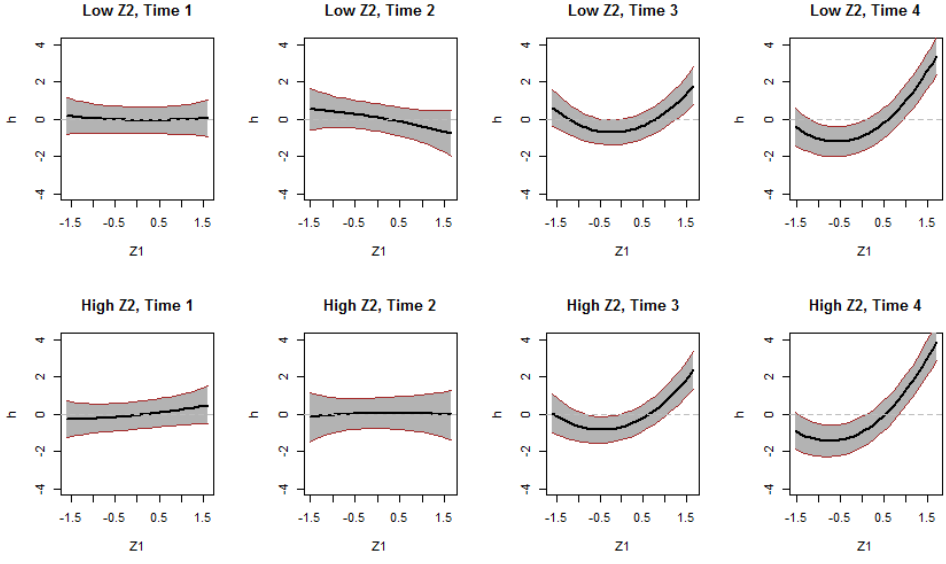
SUPPLEMENTARY FIGURE S3a Average MFVB-LKMR estimated relative importance of each exposure at four time windows for the simulation study. Relative importance is quantified by the difference in the estimated exposure response function at two given levels of exposure to a single metal, holding all other metals at median exposures. For each dataset, the relative importance was estimated at the same grid of points, which generally correspond to the 75th vs. 25th percentile of exposure, and the 95% interval for relative importance is defined as the predicted relative importance +/- 1.96 times the square root of the corresponding variance.



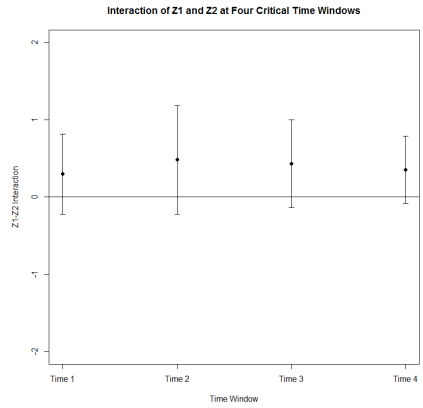
SUPPLEMENTARY FIGURE S3b Average MCMC-LKMR estimated relative importance of each exposure at four time windows for the simulation study. Relative importance is quantified by the difference in the estimated exposure response function at two given levels of exposure to a single metal, hodling all other metals at median exposures. For each dataset, the relative importance was estimated at the same grid of points, which generally correspond to the 75th vs. 25th percentile of exposure, and the 95% interval for relative importance is defined as the predicted relative importance +/- 1.96 times the square root of the corresponding variance.



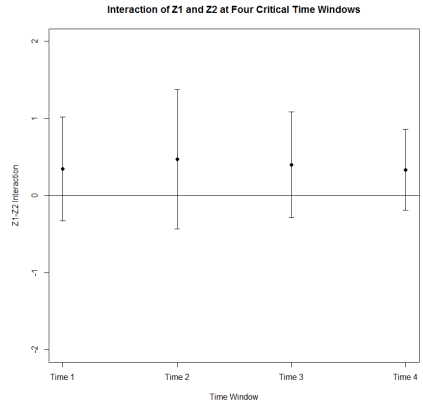
SUPPLEMENTARY FIGURE S4a Plot of the average cross-section of the MFVB-LKMR estimated exposure-response surface for Z1, at low and high fixed values of Z2, which roughly correspond to 25th (top panel) and 75th (bottom panel) of Z2 exposure, holding Z3 at its median value, for a simulation study.



SUPPLEMENTARY FIGURE S4b Plot of the average cross-section of the MCMC-LKMR estimated exposure-response surface for Z1, at low and high fixed values of Z2, which roughly correspond to 25th (top panel) and 75th (bottom panel) of Z2 exposure, holding Z3 at its median value, for a simulation study.



SUPPLEMENTARY FIGURE S5a Plot of the average MFVB-LKMR estimated interaction effect between Z1 and Z2, holding Z3 at its median exposure, for a simulation study. Interaction was defined as: First, we estimated the exposure-response effect for high versus low Z1 exposures, at high Z2 levels and median Z3. Next, we estimated the exposure-response effect for high versus low Z2 exposures, at low Z2 levels and median Z3. The difference between these two estimated exposure-response effects quantifies the Z1-Z2 interaction. The high and low levels of exposure were fixed, and roughly correspond to the 75th and 25th percentiles of exposure.



SUPPLEMENTARY FIGURE S5b Plot of the average MCMC-LKMR estimated interaction effect between Z1 and Z2, holding Z3 at its median exposure, for a simulation study. Interaction was defined as: First, we estimated the exposure-response effect for high versus low Z1 exposures, at high Z2 levels and median Z3. Next, we estimated the exposure-response effect for high versus low Z2 exposures, at low Z2 levels and median Z3. The difference between these two estimated exposure-response effects quantifies the Z1-Z2 interaction. The high and low levels of exposure were fixed, and roughly correspond to the 75th and 25th percentiles of exposure.