Supplemental to "Kernel Machine and Distributed Lag Models for Assessing Windows of Susceptibility to Environmental Mixtures in Children's Health Studies"

Ander Wilson, Hsiao-Hsien Leon Hsu, Yueh-Hsiu Mathilda Chiu,
Robert O. Wright, Rosalind J. Wright, Brent A. Coull

A Additional figures for the preliminary data analysis

Figure S1 shows cross sections of the exposure-response function estimated with BKMR to assess modification. Using BKMR there is no evidence of effect modification.

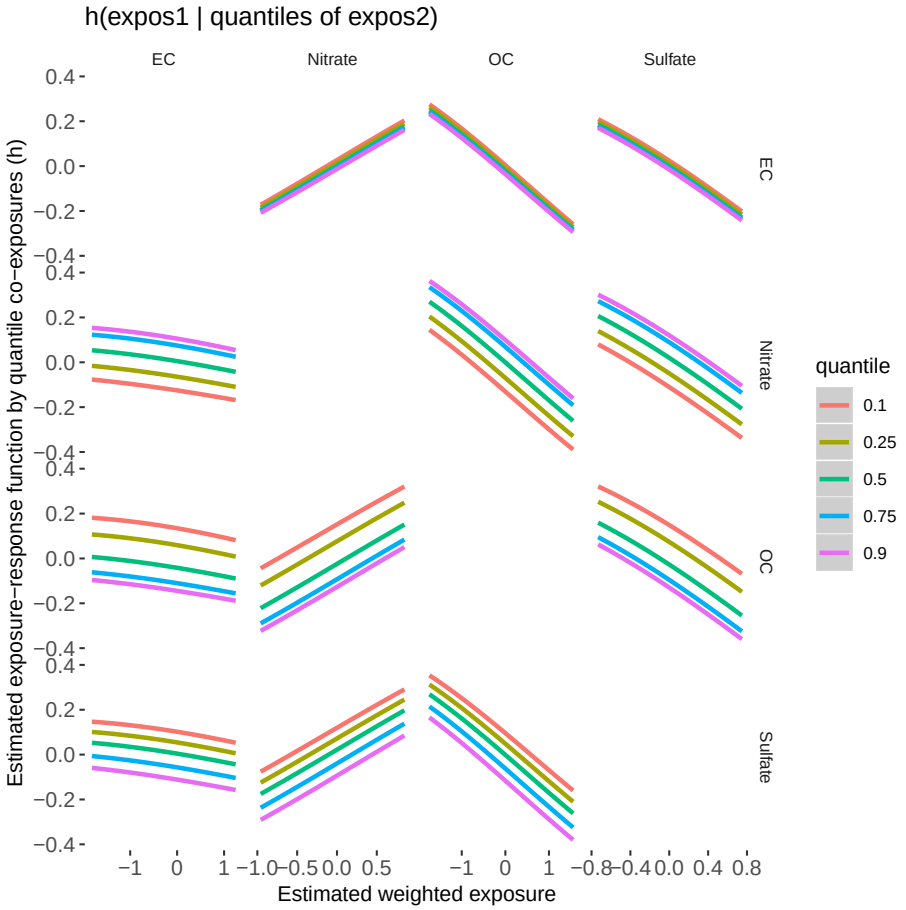

Figure S1: Estimated exposure response function with BKMR. The exposure-response function is shown for each pollutant (columns) at different quantiles of one co-pollutant (rows) and the median of the other two pollutants.

B Additional details on the method

B.1 MCMC approach

To estimate the model parameters, we first integrate out $\mathbf{h}$ from (4) in the main text. This yields $\mathbf{Y} \sim \mathrm{N}(\mathbf{Z}\boldsymbol{\gamma}, \sigma^2\widetilde{\mathbf{K}})$, where $\widetilde{\mathbf{K}} = \mathbf{I}_n + \tau^2\mathbf{K}$. The posterior distribution can be estimated using the decomposition

$$\begin{aligned} p(\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_M^*, \tau^2, \nu_1, \ldots, \nu_M, \boldsymbol{\gamma}, \sigma^{-2}|\mathbf{Y}) \;\; = \;\; & \\ & p(\boldsymbol{\gamma}|\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_M^*, \tau^2, \nu_1, \ldots, \nu_M, \sigma^{-2}, \mathbf{Y}) \\ & p(\sigma^{-2}|\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_M^*, \tau^2, \nu_1, \ldots, \nu_M, \mathbf{Y}) \\ & p(\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_M^*, \tau^2, \nu_1, \ldots, \nu_M|\mathbf{Y}). \end{aligned} \tag{1}$$

In (1), $p(\boldsymbol{\gamma}|\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_M^*, \tau^2, \nu_1, \ldots, \nu_M, \sigma^{-2}, \mathbf{Y})$ and $p(\sigma^{-2}|\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_M^*, \tau^2, \nu_1, \ldots, \tau_M, \mathbf{Y})$ are multivariate normal and gamma distribution. The final term in (1) takes the form

$$\begin{aligned} p(\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_M^*, \nu_1, \ldots, \nu_M, \tau^2|\mathbf{Y}) \;\; \propto \;\; & \\ & (\tau^{-2})^{b_1-1} \left|\widetilde{\mathbf{K}}\right|^{-1/2} \left|\mathbf{Z}^T\widetilde{\mathbf{K}}^{-1}\mathbf{Z}\right|^{-1/2} \exp\left(-b_2\tau^{-2}\right) \\ & \times \Bigg[a_2 + \mathbf{Y}^T\widetilde{\mathbf{K}}^{-1}\mathbf{Y}/2 \\ & \quad + \mathbf{Y}^T\widetilde{\mathbf{K}}^{-1}\mathbf{Z}\left(\mathbf{Z}^T\widetilde{\mathbf{K}}^{-1}\mathbf{Z}\right)^{-1}\mathbf{Z}^T\widetilde{\mathbf{K}}^{-1}\mathbf{Y}/2\Bigg]^{-[(n-p)/2+a_1]} \\ & \times p\left(\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_M^*|\nu_1, \ldots, \nu_M\right) p(\nu_1, \ldots, \nu_M) p(\tau). \end{aligned} \tag{2}$$

Bobb et al. (2015) updated each of the M scalar parameters in the kernel function independently with Metropolis-Hastings. This approach is unappealing for our model. We have $\sum_{m=1}^{M} L_m$ parameters in the kernel function and potentially high correlation among parameters due to the temporal correlation in the exposures. Our MCMC algorithm instead iteratively samples each $\boldsymbol{\theta}_m^*$ using an elliptical slice sampler (Murray et al., 2009) and the kernel of (2). Then we sample τ^{-2} using random walk Metropolis-Hastings and the same integrated kernel in (2). Finally, we use a Gibbs sampler to sample σ^{-2}, $\boldsymbol{\gamma}$, and $\nu_1, \ldots, \nu_M$ from their respective full conditionals. The full conditional for ν_m is a generalized inverse-Gaussian distribution with density function $f(\nu_m; \lambda, \chi, \psi) \propto \nu_m^{\lambda-1} \exp\{-(\chi/\nu_m + \psi\nu_m)/2\}$, where $\lambda = -(L_m - 1)/2$, $\psi = 1$, and $\chi = \kappa_m^{-1}\boldsymbol{\theta}_m^{*T}\boldsymbol{\theta}_m^*$. Algorithm 1 shows the full MCMC approach used to fit the model.

B.2 Posterior inference for $h(\cdot)$

Estimates of $h(\cdot)$ for the observed exposure levels can be obtained by sampling from the conditional distribution of $\mathbf{h}$ from (4) in the main text (Bobb et al., 2015). Specifically, for each MCMC iteration, we sample $\mathbf{h}$ from

$$\mathbf{h}|\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_M^*, \sigma^2, \tau^2, \boldsymbol{\gamma}, \mathbf{Y} \sim \mathrm{N}\left[\tau^2\mathbf{K}\widetilde{\mathbf{K}}^{-1}(\mathbf{Y} - \mathbf{Z}\boldsymbol{\gamma}), \sigma^2\tau^2\mathbf{K}\widetilde{\mathbf{K}}^{-1}\right]. \tag{3}$$

Algorithm 1: MCMC algorithm for BKMR-DLM

Initiate parameters

Iterate:

For m in $1, \ldots, K$ do:

$\boldsymbol{\theta}^*_m \leftarrow$ elliptical slice sampler using $p(\boldsymbol{\theta}^*_1, \ldots, \boldsymbol{\theta}^*_M, \tau^2 | \mathbf{Y})$ as in (16)

$\nu_m \leftarrow$ gen-inv-Gaussian* with $\lambda = -(L_m - 1)/2$, $\psi = 1$, and $\chi = \kappa_m^{-1} \boldsymbol{\theta}^{*T}_m \boldsymbol{\theta}^*_m$.

$\rho_m \leftarrow ||\boldsymbol{\theta}^*_m||^{-2}$ (deterministic)

$\log(\tau^2) \leftarrow$ M-H and with kernel as in (16)

$$\sigma^{-2} \leftarrow \text{gamma}\left[a_1 + (n-p)/2, a_2 + \mathbf{Y}^T \left\{ \widetilde{\mathbf{K}}^{-1} + \widetilde{\mathbf{K}}^{-1}\mathbf{Z}\left(\mathbf{Z}^T\widetilde{\mathbf{K}}^{-1}\mathbf{Z}\right)^{-1}\mathbf{Z}^T\widetilde{\mathbf{K}}^{-1} \right\} \mathbf{Y}/2\right]$$

$$\boldsymbol{\gamma} \leftarrow \text{MVN}\left[\left(\mathbf{Z}^T\widetilde{\mathbf{K}}^{-1}\mathbf{Z}\right)^{-1}\mathbf{Z}^T\widetilde{\mathbf{K}}^{-1}\mathbf{Y}, \left(\mathbf{Z}^T\widetilde{\mathbf{K}}^{-1}\mathbf{Z}\right)^{-1}\right]$$

* Generalized-inv-Gaussian with pdf $f(\rho_m; \lambda, \chi, \psi) \propto \rho_m^{\lambda-1} \exp\{-(\chi/\rho_m + \psi\rho_m)/2\}$

To predict $h(\cdot)$ at new values, we predict $\mathbf{h}_{new}$ by considering the joint distribution

$$\begin{pmatrix} \mathbf{h} \\ \mathbf{h}_{new} \end{pmatrix} \sim \text{N}\left[\mathbf{0}, \sigma^2\tau^2 \begin{pmatrix} \mathbf{K} & \mathbf{K}''^T \\ \mathbf{K}'' & \mathbf{K}' \end{pmatrix}\right]. \tag{4}$$

The subsequent posterior predictive distribution

$$\begin{aligned} \mathbf{h}_{new} | \boldsymbol{\theta}^*_1, \ldots, \boldsymbol{\theta}^*_M, \sigma^2, \tau^2, \boldsymbol{\gamma}, \mathbf{Y} \ \sim& \\ \text{N}\left[\tau^2\mathbf{K}''\widetilde{\mathbf{K}}^{-1}(\mathbf{Y} - \mathbf{Z}\boldsymbol{\gamma}), \sigma^2\tau^2\left\{\mathbf{K}' - \tau^2\mathbf{K}''\widetilde{\mathbf{K}}^{-1}\mathbf{K}''^T\right\}\right]. \end{aligned} \tag{5}$$

We use this distribution to visualize h over a regularly spaced grid in the data analysis.

C Additional results from the simulation study

Figures S2-S5 show the estimated weight functions for the first 100 simulated data sets for simulation scenario A.

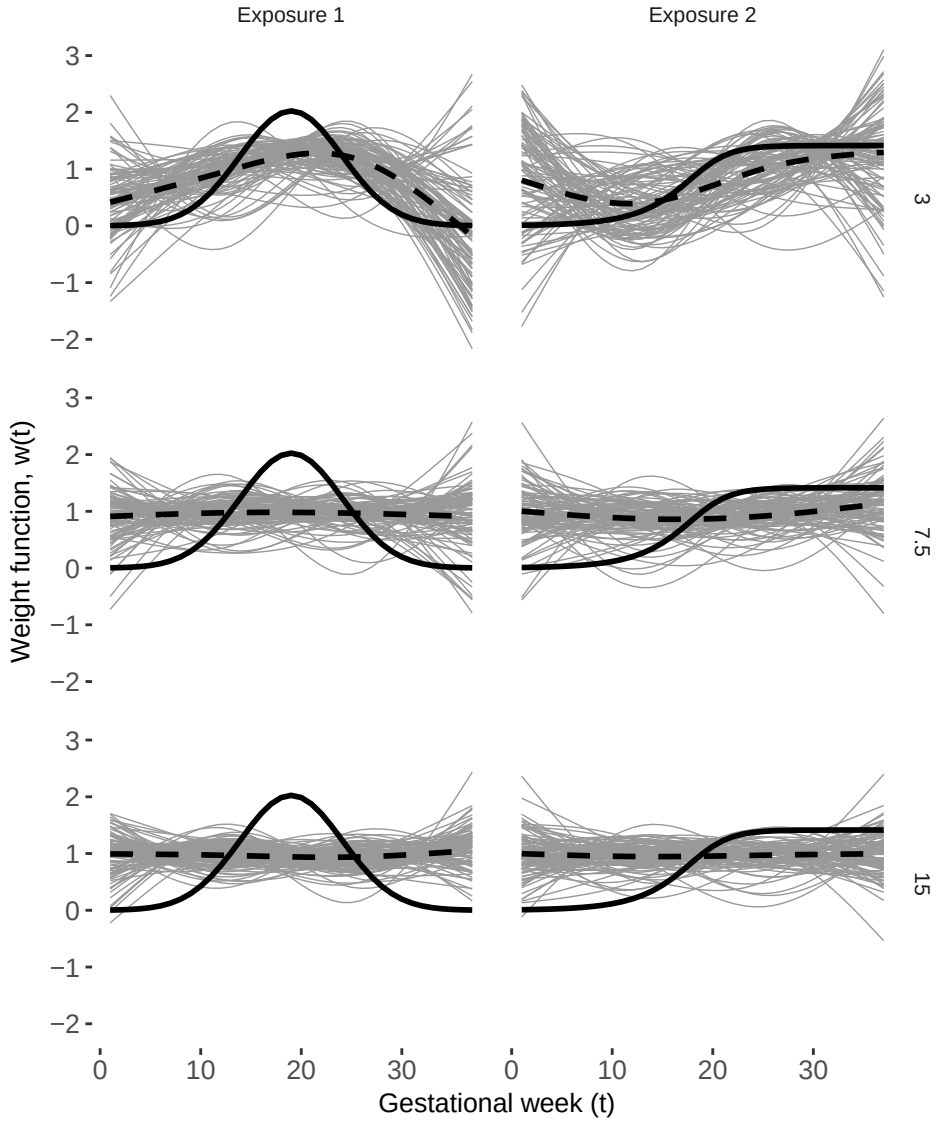

Figure S2: Simulated and estimated weight functions for scenario A with $n = 100$ using BKMR-DLM with a Gaussian kernel. The black line shows the true weight function. The thin grey lines show the estimated weight functions for the first 100 simulated data sets. The dashed black line shows the mean weight function across all simulated data sets. The rows show the residual SD of 3, 7.5, and 15, representing the high, medium, and low signal-to-noise settings.

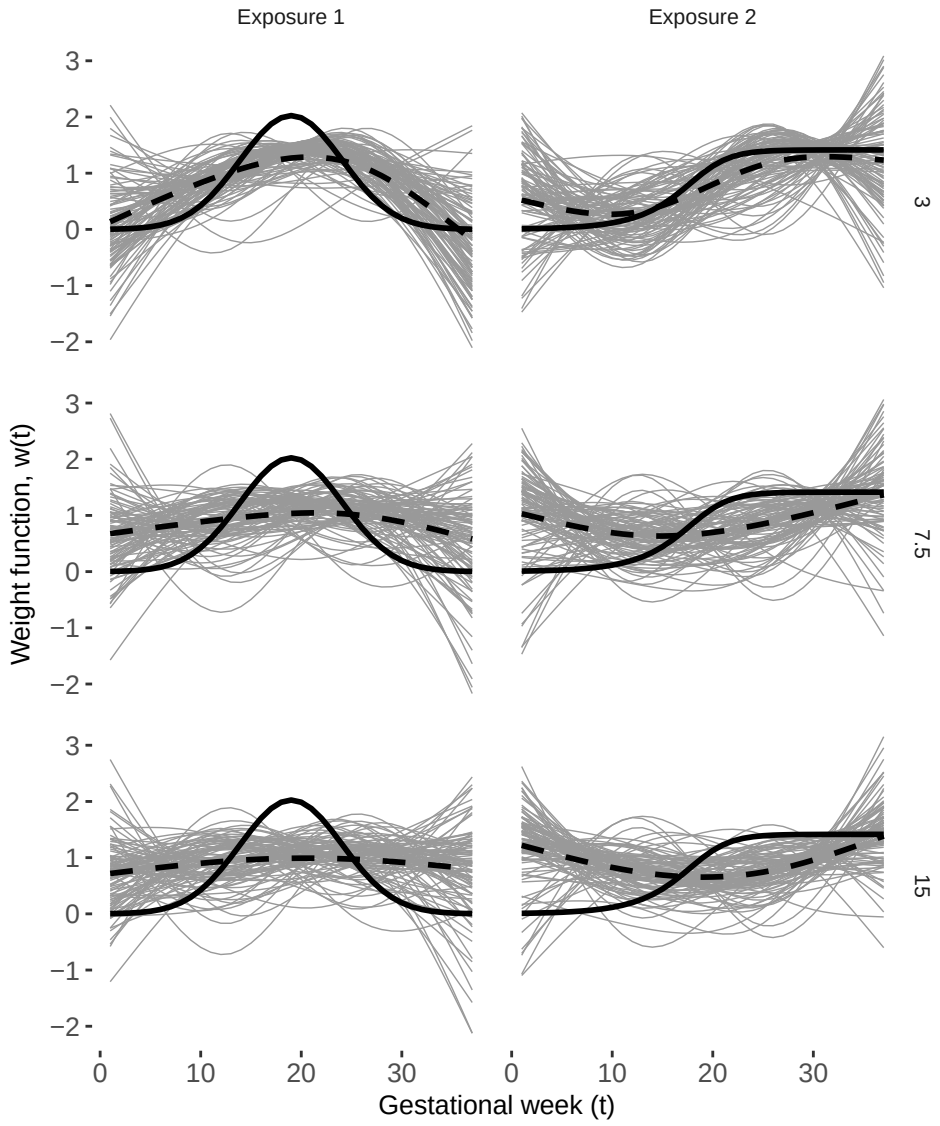

Figure S3: Simulated and estimated weight functions for scenario A with $n = 100$ using BKMR-DLM with a quadratic kernel. The black line shows the true weight function. The thin grey lines show the estimated weight functions for the first 100 simulated data sets. The dashed black line shows the mean weight function across all simulated data sets. The rows show the residual SD of 3, 7.5, and 15, representing the high, medium, and low signal-to-noise settings.

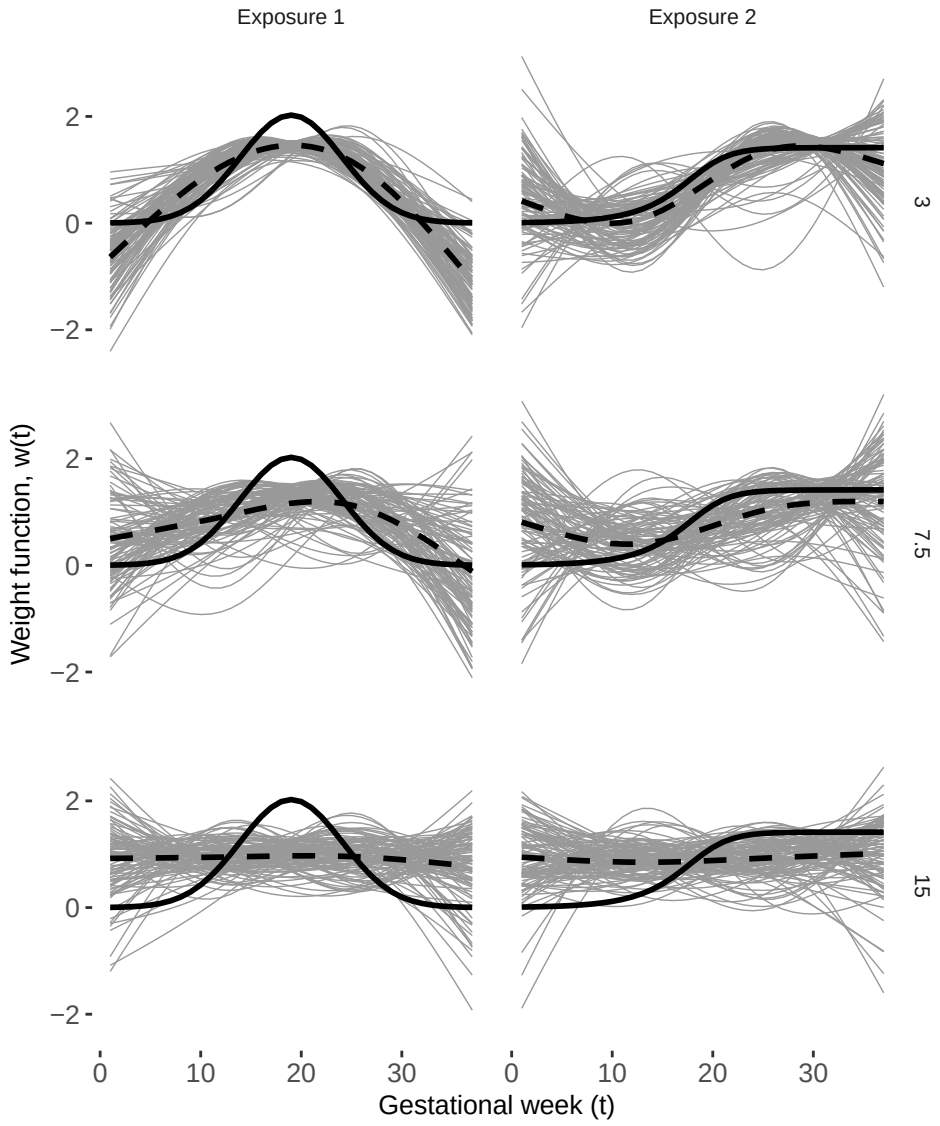

Figure S4: Simulated and estimated weight functions for scenario A with $n = 500$ using BKMR-DLM with a Gaussian kernel. The black line shows the true weight function. The thin grey lines show the estimated weight functions for the first 100 simulated data sets. The dashed black line shows the mean weight function across all simulated data sets. The rows show the residual SD of 3, 7.5, and 15, representing the high, medium, and low signal-to-noise settings.

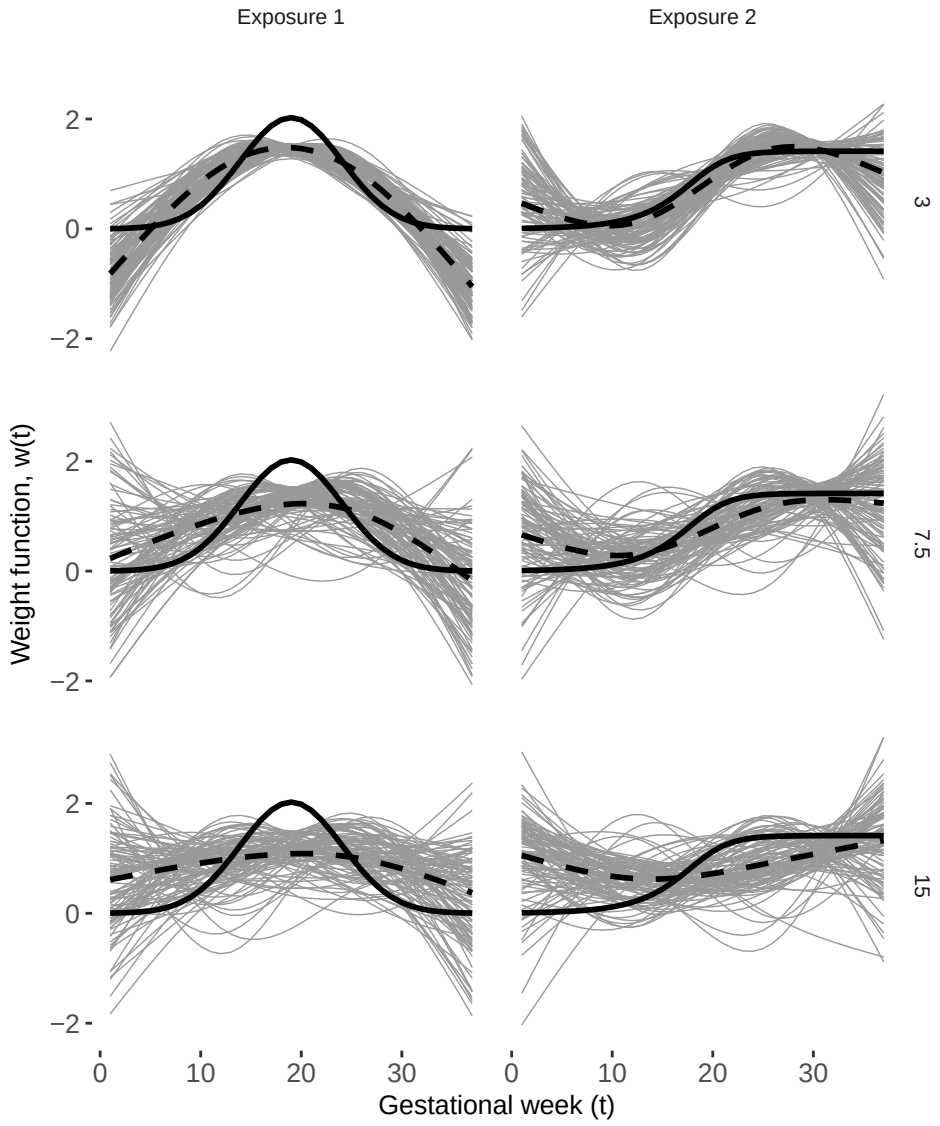

Figure S5: Simulated and estimated weight functions for scenario A with $n = 500$ using BKMR-DLM with a quadratic kernel. The black line shows the true weight function. The thin grey lines show the estimated weight functions for the first 100 simulated data sets. The dashed black line shows the mean weight function across all simulated data sets. The rows show the residual SD of 3, 7.5, and 15, representing the high, medium, and low signal-to-noise settings.

D Alternative simulation scenario C with smoother weight functions

The natural spline basis used to estimate the weight functions with BKMR-DLM are not flexible enough to match the shape of the true weight functions in simulation scenarios A and B presented in the main text. This is evident from Figures S2-S5. This likely limits the performance of BKMR-DLM in those scenarios. To evaluate the impact of this misspecification of the weight function in our model we include an additional simulation scenario that is the same as scenario A in the main text but has smoother weight functions. Specifically, the weight functions are simulated from a natural spline basis with 4 degrees of freedom that can be perfectly matched by the BKMR-DLM models. In each case we projected the original weight function onto a natural spline basis with four degrees of freedom to perfectly match the parameterization of the BKMR-DLM models used in the simulation study. Comparing these scenarios to those in the main text isolates the impact of using an overly smooth basis for the weight function in BKMR-DLM.

Figures S6-S9 illustrate the smoother true weight functions (solid black) and the first 100 estimated weight functions with BKMR-DLM with both a Gaussian and quadratic kernel. The dashed black line is the mean of all simulated data sets. In the cases with stronger signal-to-noise ratio and large sample size the BKMR-DLM models with both Gaussian and quadratic kernel well approximate the true weight function. With smaller sample size and larger signal-to-noise ratio the weight functions are relatively flat. This was also the case for scenario A in the main text.

Table S1 show results from simulation scenario C. Comparing scenario C to scenario A in the main text, BKMR-DLM and BKMR-DLM-quad have lower RMSE on the exposure-response function (h) and weight functions in scenario C where the models can better approximate the true weight functions. This improvement occurs only in the higher signal-to-noise and larger sample size cases. As a result, BKMR-DLM-quad shows greater improvement in estimating the exposure-response function compared to alternative models.

E Alternative simulation scenario D

Scenario D is similar to scenario B in the main text but uses a different exposure-response relationship. Scenario D includes nonlinear main effects and fewer interaction. The data generating mechanism included three active exposures ($PM_{2.5}$, nitrate, and CO). Each model fit to the data included five exposures. The exposure-response function h between the three active exposures and the outcome is the only difference in the two scenarios. For scenario D, the exposure-response function is

$$h_i = 3/[1 + \exp(-2E_1^s)] + 2E_2^s \mathbb{1}_{(E_2^s>0)} - E_1^s E_2^s - 2E_3^s \mathbb{1}_{(E32^s>0)} \tag{6}$$

The weight functions and scaled weighted exposures are as defined in the main text. As in scenario B, we also include in scenario D two additional exposures that are not associated with the outcome. All other details are the same as in scenario B and described in the main text.

Table S2 shows results for scenario D. Overall, the results are similar to the results for scenario B in the main text. BKMR-DLM-quad was best for inference on the exposure-response function with lower RMSE on interval coverage near the nominal level. BKMR

Table S1: Simulation results for scenario C with two active exposures. This is the same as scenario A in the main text but has smoother true weight functions. The table shows (from left to right) RMSE for the exposure response function and 0.95 interval coverage for the exposure-response function, the average pointwise RMSE and pointwise 0.95 coverage for the two weight functions, and the probability that an interaction is detected by comparing the difference in an IQR change in $PM_{2.5}$ at the 75^{th} and 25^{th} percentile of nitrate. A probability of an interaction near 1 indicates the model consistently finds evidence of an interaction and a probability near 0.5 indicates no evidence of interaction.

Model	RMSE h	Coverage h	RMSE $w(t)$	Coverage $w(t)$	Pr(interact)
		$n = 100$, noise: sd$(\epsilon) = 3.0$			
BKMR	1.344	0.606	0.746	NA	0.497
BKMR-DLM	1.200	0.979	0.578	0.912	0.505
BKMR-DLM-quad	0.993	0.931	0.505	0.924	0.657
DLM	1.155	0.866	0.862	0.869	NA
DLNM	1.151	0.833	0.932	0.867	NA
		$n = 100$, noise: sd$(\epsilon) = 7.5$			
BKMR	1.664	0.825	0.746	NA	0.497
BKMR-DLM	2.943	0.988	0.758	0.899	0.519
BKMR-DLM-quad	1.735	0.981	0.711	0.894	0.534
DLM	2.363	0.921	1.124	0.831	NA
DLNM	2.172	0.906	1.122	0.847	NA
		$n = 100$, noise: sd$(\epsilon) = 15.0$			
BKMR	2.281	0.933	0.746	NA	NA
BKMR-DLM	5.266	0.994	0.785	0.901	NA
BKMR-DLM-quad	2.924	0.995	0.762	0.888	NA
		$n = 500$, noise: sd$(\epsilon) = 3.0$			
BKMR	1.246	0.441	0.746	NA	0.480
BKMR-DLM	0.560	0.956	0.348	0.916	0.489
BKMR-DLM-quad	0.538	0.903	0.317	0.928	0.977
DLM	0.803	0.688	0.651	0.856	NA
DLNM	0.773	0.837	0.868	0.909	NA
		$n = 500$, noise: sd$(\epsilon) = 7.5$			
BKMR	1.406	0.592	0.746	NA	0.473
BKMR-DLM	1.333	0.991	0.618	0.904	0.498
BKMR-DLM-quad	1.080	0.934	0.541	0.916	0.622
DLM	1.273	0.859	0.895	0.860	NA
DLNM	1.352	0.885	0.961	0.887	NA
		$n = 500$, noise: sd$(\epsilon) = 15.0$			
BKMR	1.638	0.763	0.746	NA	0.477
BKMR-DLM	2.560	0.999	0.741	0.897	0.520
BKMR-DLM-quad	1.642	0.976	0.693	0.897	0.541
DLM	2.216	0.906	1.100	0.818	NA
DLNM	2.378	0.914	1.117	0.859	NA

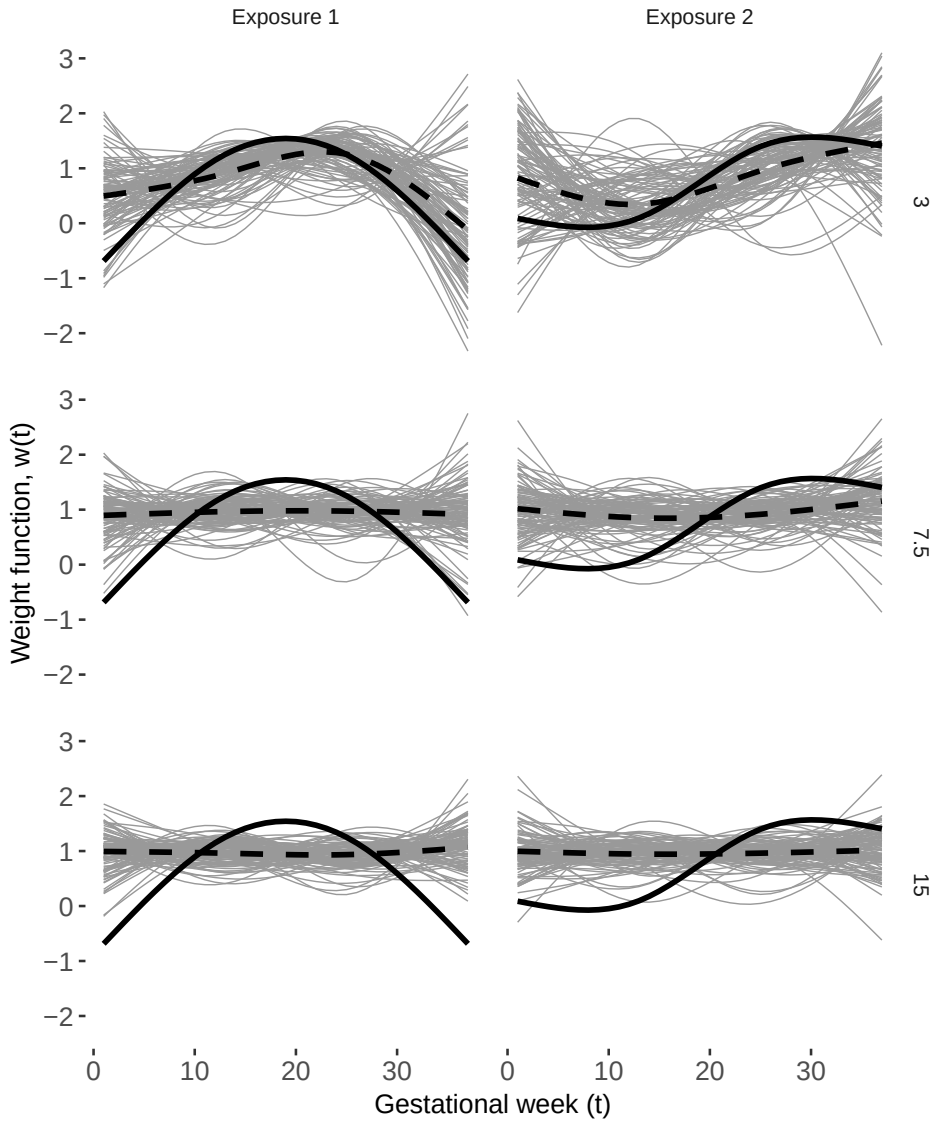

Figure S6: Simulated and estimated weight functions for scenario C. The thin black line shows the true weight function. The thin grey lines show the estimated weight functions for the first 100 simulated data sets. The dashed black line shows the mean weight function across all simulated data sets. Results from scenario C with $n = 100$ with BKMR-DLM with a Gaussian kernel. The rows show the residual SD of 3, 7.5, and 15, representing the high, medium, and low signal-to-noise settings.

has lower RMSE in some settings but had very low coverage. BKMR-DLM-quad was also best at making inference on the weight functions. The method had the lower RMSE and interval coverage close the the nominal level. A primary difference is that BKMR-DLM-quad had lower power to identify critical windows than in scenario B. Additive DLM and DLNM had higher power than BKMR-DLM but again had a very high rate of identifying incorrect critical windows on the two exposures not associated with the outcome. Similar to in scenario B, BKMR-DLM and BKMR-DLM-quad had a very low rate of identifying

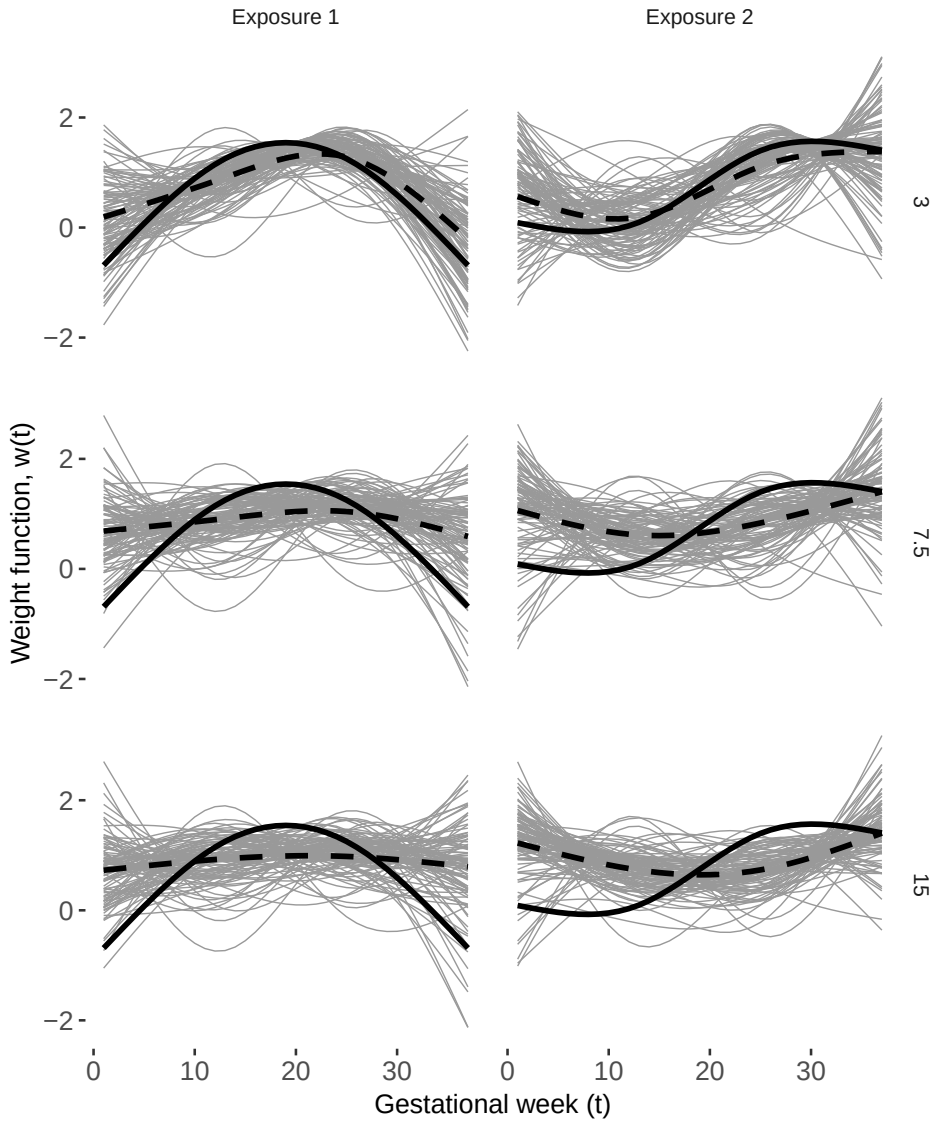

Figure S7: Simulated and estimated weight functions for scenario C. The thin black line shows the true weight function. The thin grey lines show the estimated weight functions for the first 100 simulated data sets. The dashed black line shows the mean weight function across all simulated data sets. Results from scenario C with $n = 100$ with BKMR-DLM with a quadratic kernel. The rows show the residual SD of 3, 7.5, and 15, representing the high, medium, and low signal-to-noise settings.

incorrect windows and had high precision.

Table S2: Simulation results for scenario D, three active exposure and two null exposures. The table shows (from left to right) RMSE for the exposure response function and 0.95 interval coverage for the exposure-response function, the average pointwise RMSE and pointwise 0.95 coverage for the weight functions for the three active exposures, the frequency of detecting a window in the three active exposures, the frequency of detecting a window in the two exposures not associated with the outcome, and the precision for window detection which is number of correct windows identified divided by total number of windows identified.

Model	RMSE h	Coverage h	RMSE $w(t)$	Coverage $w(t)$	Pr(window|active)	Pr(window|null)	Precision
			$n = 100$, noise: sd(ϵ) = 3.0				
BKMR	1.644	0.630	NA	NA	NA	NA	NA
BKMR-DLM	1.886	0.941	0.662	0.937	0.000	0.000	NaN
BKMR-DLM-quad	1.292	0.948	0.621	0.916	0.078	0.005	0.959
DLM	1.551	0.918	1.418	0.720	0.451	0.417	0.618
DLNM	1.466	0.872	1.404	0.652	0.497	0.435	0.632
			$n = 100$, noise: sd(ϵ) = 7.5				
BKMR	1.994	0.780	NA	NA	NA	NA	NA
BKMR-DLM	3.334	0.981	0.683	0.937	0.000	0.000	NaN
BKMR-DLM-quad	2.407	0.983	0.713	0.924	0.002	0.000	1.000
DLM	3.530	0.930	1.399	0.783	0.312	0.317	0.596
DLNM	3.171	0.906	1.384	0.715	0.313	0.302	0.609
			$n = 100$, noise: sd(ϵ) = 15.0				
BKMR	2.525	0.902	NA	NA	NA	NA	NA
BKMR-DLM	5.936	0.992	0.684	0.937	0.000	0.000	NaN
BKMR-DLM-quad	4.299	0.992	0.737	0.926	0.002	0.000	1.000
DLM	6.970	0.932	1.379	0.799	0.283	0.283	0.600
DLNM	6.119	0.917	1.371	0.743	0.283	0.265	0.615
			$n = 500$, noise: sd(ϵ) = 3.0				
BKMR	1.185	0.587	NA	NA	NA	NA	NA
BKMR-DLM	0.774	0.966	0.441	0.918	0.362	0.005	0.991
BKMR-DLM-quad	0.734	0.930	0.446	0.904	0.466	0.043	0.942
DLM	0.920	0.844	1.410	0.586	0.787	0.675	0.636
DLNM	0.915	0.886	1.325	0.700	0.511	0.419	0.647
			$n = 500$, noise: sd(ϵ) = 7.5				
BKMR	1.715	0.592	NA	NA	NA	NA	NA
BKMR-DLM	2.148	0.993	0.653	0.936	0.000	0.000	NaN
BKMR-DLM-quad	1.363	0.963	0.619	0.916	0.090	0.000	1.000
DLM	1.744	0.908	1.424	0.713	0.477	0.400	0.641
DLNM	1.802	0.906	1.379	0.720	0.385	0.308	0.653
			$n = 500$, noise: sd(ϵ) = 15.0				
BKMR	2.012	0.656	NA	NA	NA	NA	NA
BKMR-DLM	3.079	1.000	0.685	0.937	0.000	0.000	NaN
BKMR-DLM-quad	2.169	0.990	0.688	0.928	0.003	0.000	1.000
DLM	3.250	0.918	1.403	0.769	0.348	0.318	0.622
DLNM	3.340	0.919	1.360	0.767	0.273	0.288	0.588

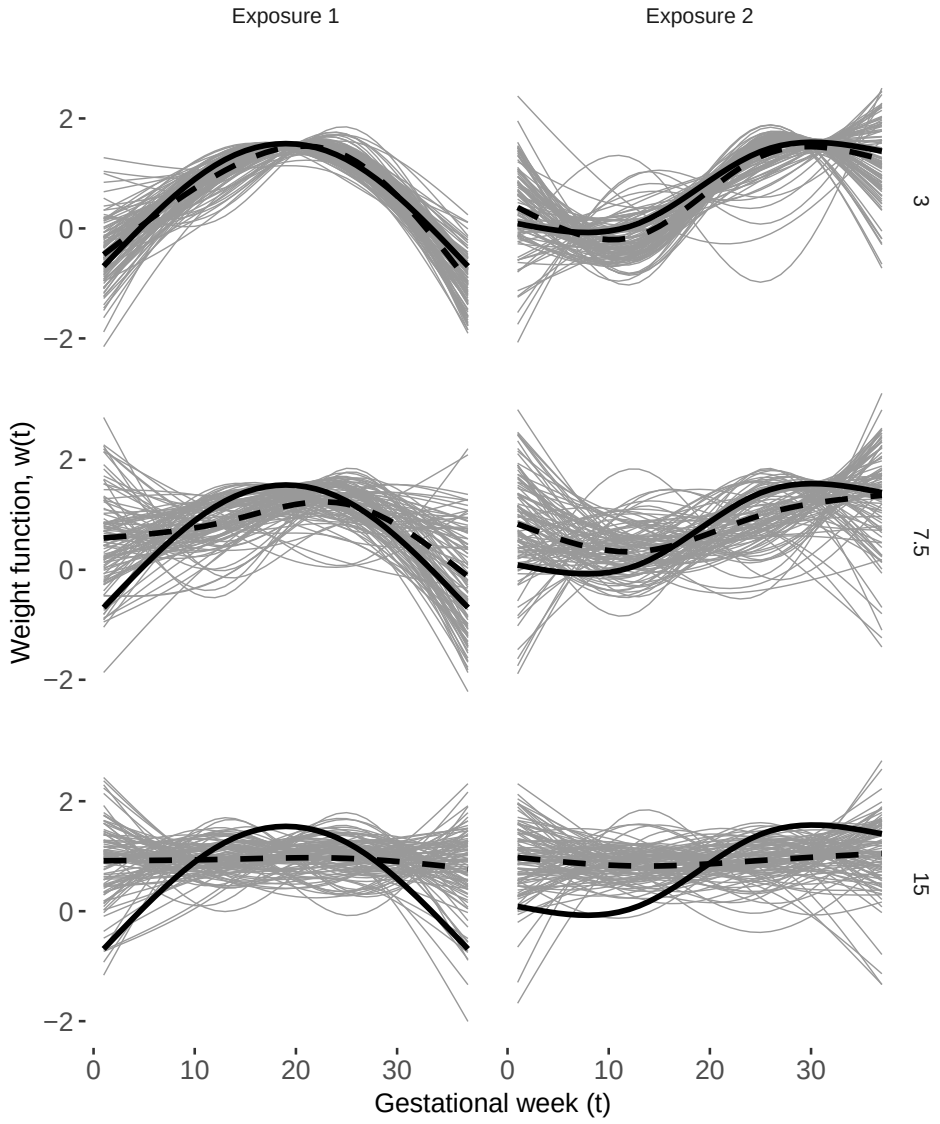

Figure S8: Simulated and estimated weight functions for scenario C. The thin black line shows the true weight function. The thin grey lines show the estimated weight functions for the first 100 simulated data sets. The dashed black line shows the mean weight function across all simulated data sets. Results from scenario C with $n = 500$ with BKMR-DLM with a Gaussian kernel. The rows show the residual SD of 3, 7.5, and 15, representing the high, medium, and low signal-to-noise settings.

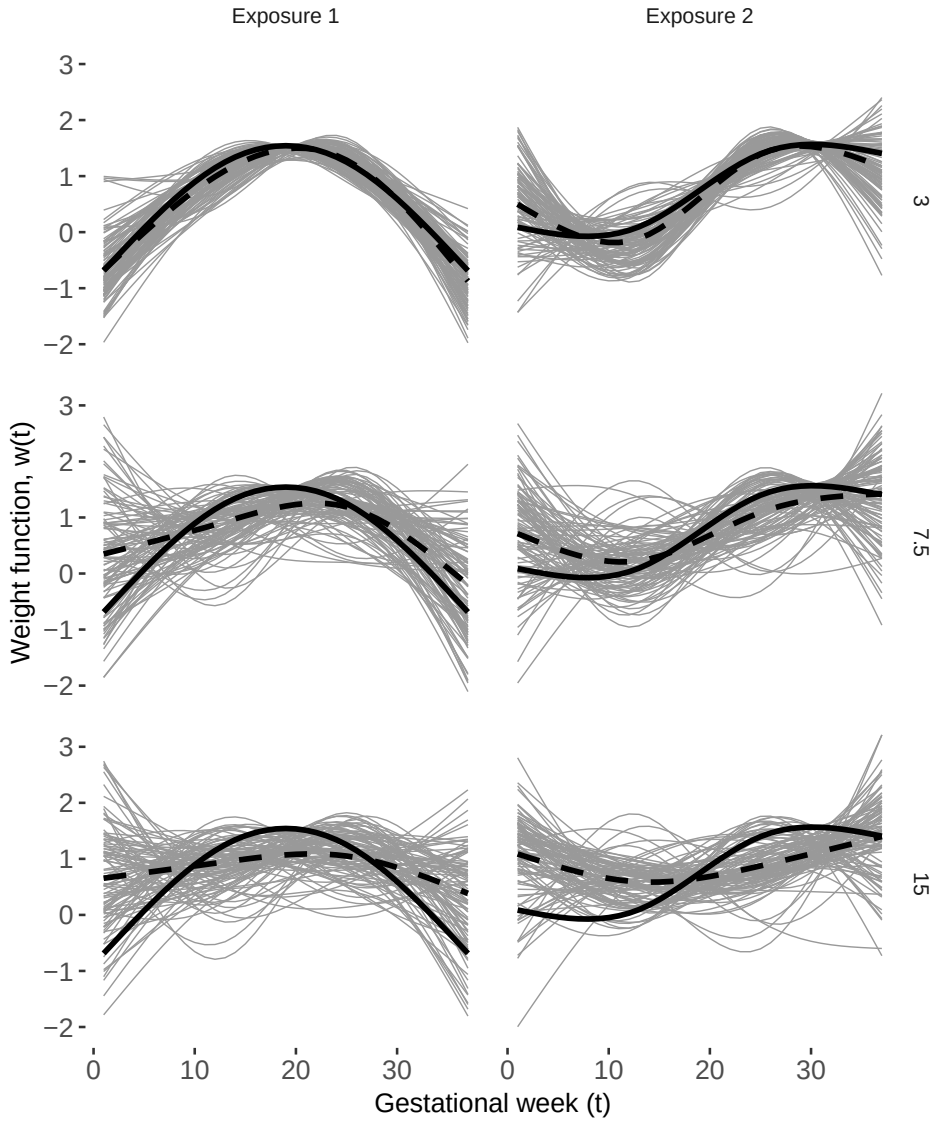

Figure S9: Simulated and estimated weight functions for scenario C. The thin black line shows the true weight function. The thin grey lines show the estimated weight functions for the first 100 simulated data sets. The dashed black line shows the mean weight function across all simulated data sets. Results from scenario C with $n = 500$ with BKMR-DLM with a quadratic kernel. The rows show the residual SD of 3, 7.5, and 15, representing the high, medium, and low signal-to-noise settings.

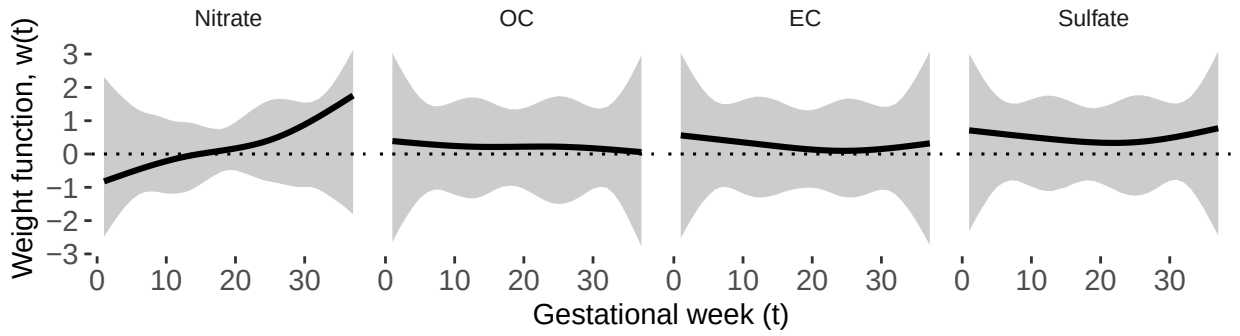

Figure S10: Estimated weight functions from the analysis of BWGAz in ACCESS with BKMR-DLM using a quadratic kernel. This uses the full ACCESS cohort. The weight function is constrained and does not reflect the magnitude of the association or the direction of the association. It only reflects the timing of the association.

.

F BKMR-DLM analysis of the full cohort

The data analysis in Section 5 of the main test included only observations for male babies with obese mothers. For completeness, we repeated the analysis using BKMR-DLM with a quadratic kernel using all full term births ($n = 661$).

Figure S10 shows the estimated weight functions. No critical windows are identified. Figure S11 shows the estimated exposure-response function. There is evidences of a main effect of nitrate and some evidence of interactions. Figure S12 shows results of the stratified DLM analysis for the full cohort.

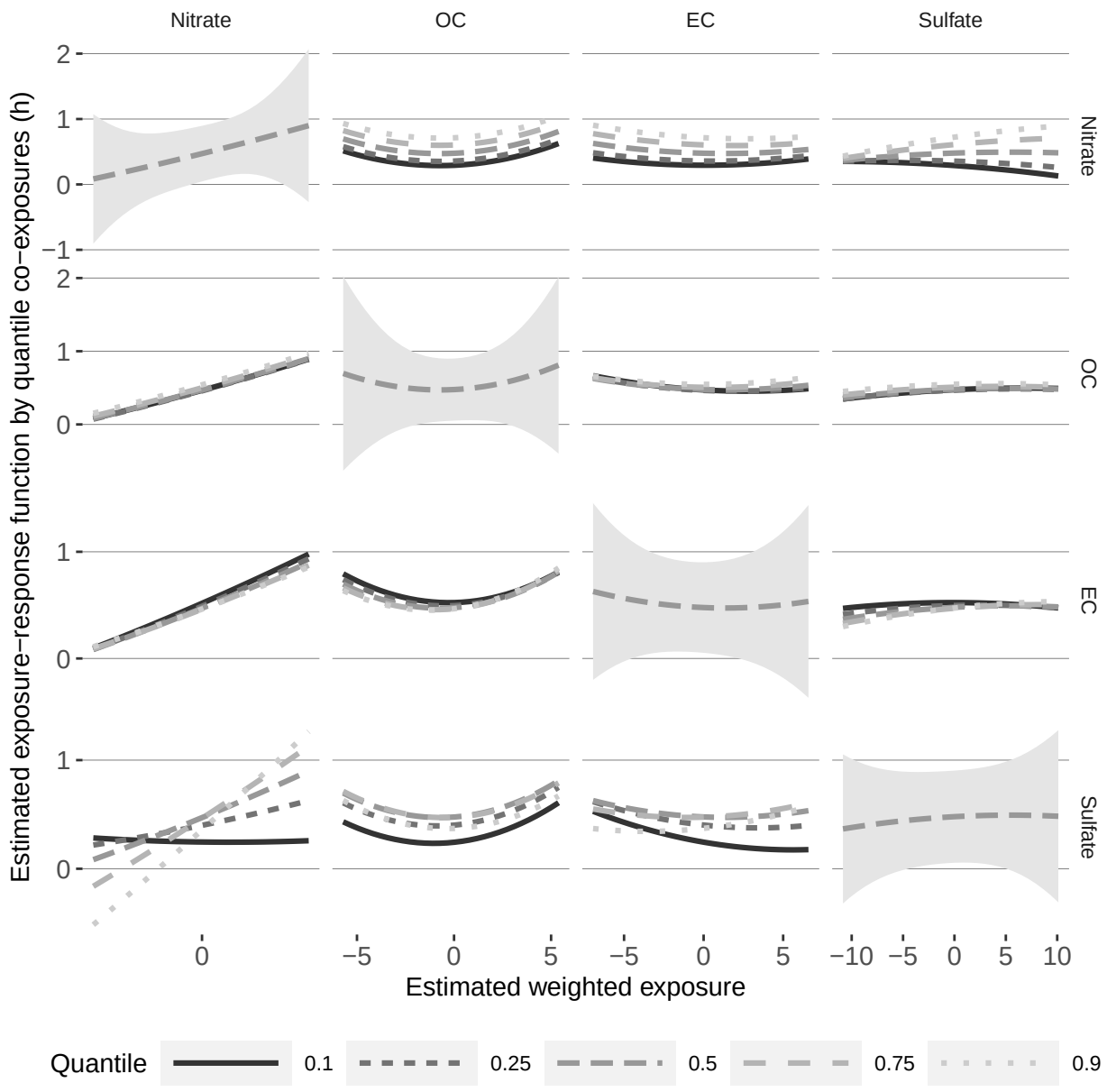

Figure S11: Cross sections of the estimated exposure-response function ($\hat{h}$) from BKMR-DLM with a quadratic kernel. This uses the full ACCESS cohort. The panels on the diagonal show the main effect, which is the association between a weighted exposure (x-axis) and the outcome at the median level of all other weighted exposures. The dashed line represents the posterior mean and the shaded ribbon represents the 0.95 credible interval. The off-diagonals show the exposure-response function at different quantiles of a single co-exposure. For example, the top right panel shows the sulfate exposure-response relationship at different quantiles of nitrate and median levels of OC and EC. A fanning or deviation from parallel lines in the exposure-response relationship represents evidence of an interaction.

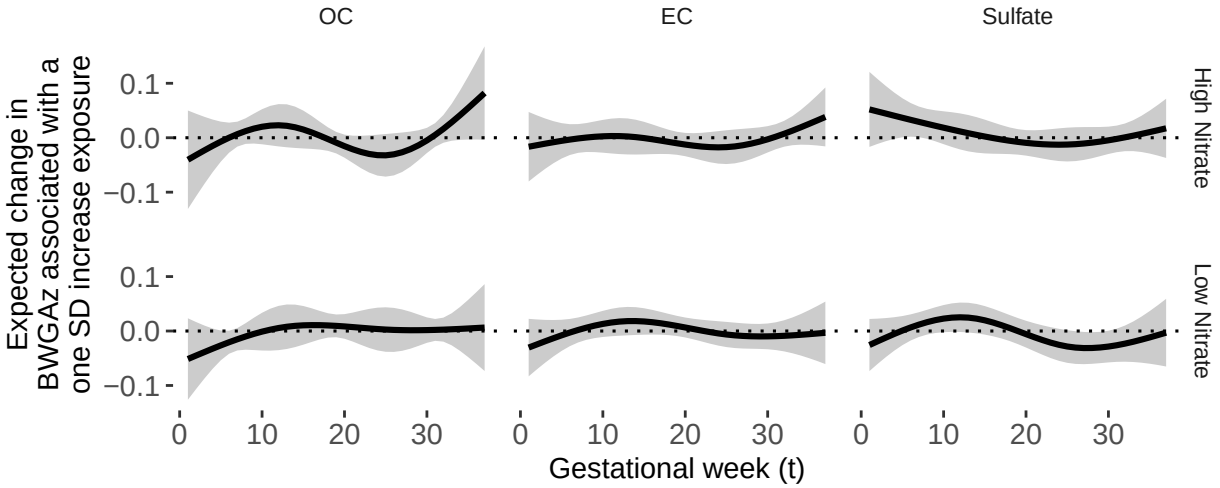

Figure S12: Estimated distributed lag function between exposures and birth weight for gestational age z-score in ACCESS using the stratified DLMs. This uses the full ACCESS cohort. The DLM for each exposure is stratified by mean nitrate level over pregnancy (below and above median nitrate value). The function represents the estimated expected difference in BWGAz per one standard deviation difference in exposure (y-axis) as a function of gestational week (x-axis).

References

Bobb, J. F., Valeri, L., Claus Henn, B., Christiani, D. C., Wright, R. O., Mazumdar, M., Godleski, J. J., and Coull, B. A. (2015). Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. *Biostatistics*, 16(3):493–508.

Murray, I., Adams, R. P., and MacKay, D. J. C. (2009). Elliptical slice sampling. *Journal of Machine Learning Research: W&CP*, 9:541–548.