Assessment of Critical Exposure and Outcome Windows in Time-to-event Analysis with Application to Air Pollution and Preterm Birth Study.

Supplementary Materials

Howard H Chang[1,*], Joshua L Warren[2], Lyndsey A Darrow[3], Brian J Reich[4], Lance A Waller[1]

[1] Department of Biostatistics and Bioinformatics, Emory University.

[2] Department of Biostatistics, Yale University.

[3] Department of Epidemiology, Emory University.

[4] Department of Statistics, North Carolina State University.

[*] Corresponding author: howard.chang@emory.edu

1. Estimation

Each element of the regression coefficient in $\boldsymbol{\delta}$ and the discrete-time baseline hazard $h(g)$ is assigned flat priors. Precision parameter σ^{-2} is assigned Gamma (0.001, 0.001); α is assigned Uniform (0,1); and ϕ is assigned Gamma (0.03, 0.005) with mean 6, chosen such that the correlation between weekly effects drops below 0.05 if they are greater than 16 weeks apart. We also considered two additional priors for σ^{-2}: Gamma (0.5, 0.0005) and Gamma (0.1, 0.0001). These two diffused priors minimize the strong peak near zero. We found the results to be nearly identical to the original prior.

Let c_i denote the indicator for whether pregnancy i is censored (full-term). For each pregnancy, we augment the data (c_i, g_i) to $[Y_i(27), ..., Y_i(g_i)]$, where $Y_i(g) = 0$ for $g < g_i$ and $Y_i(g_i) = 1 - c_i$. Therefore at each time point during the pregnancy, $Y_i(g)$ indicates whether a birth occurred during the at risk window weeks 27 to 36. The model for pregnancy i can then be written as

$$Y_i(g) \sim \text{Bernoulli}\left(\Phi\left[h(g) + \sum_{k=1}^{g} \theta_{g,k}\, x_{ik} + \boldsymbol{\delta}\, \mathbf{z}_i\right]\right),$$

and the full data likelihood is given by:

$$\left[\prod_{i=1}^{n}\prod_{g=27}^{g_i} \pi_{ig}^{Y_i(g)}(1-\pi_{ig})^{1-Y_i(g)}\right] \times \mathrm{N}(\boldsymbol{\theta}_{36}\,|\,\mathbf{0},\,(1-\alpha^2)^{-1}\Sigma_{36}) \times \prod_{g=27}^{35}\mathrm{N}(\boldsymbol{\theta}_g\,|\,\alpha\boldsymbol{\theta}_{g+1}^{[1:g]},\,\Sigma_g)$$
$$\times \prod_{g=27}^{36}[\,h(g)\,] \times [\,\boldsymbol{\delta}\,] \times [\,\alpha\,] \times [\,\phi\,] \times [\,\sigma^{-2}\,]$$

where π_{ij} is given by equation (3.1), the (k, k') element of Σ_g is given by $\sigma^2 \exp(-\frac{1}{\phi}|k - k'|)$, and $[\,\cdot\,]$ denotes the prior distribution described above.

The Bernoulli model for $Y_i(g)$ is equivalent to the model $Y_i(g) = I(Z_i(g) > 0)$, where $Z_i(g)$ is a latent variable with $Z_i(g) \sim \mathrm{N}(h(g) + \sum_{k=1}^{g} \theta_{g,k}\, x_{ik} + \boldsymbol{\delta}\, \mathbf{z}_i, 1)$. The latent variables $Z_i(g)$, $\boldsymbol{\theta}_g$, $h(g)$, $\boldsymbol{\delta}$ and σ^2 have conditional distributions in closed-form and Gibbs sampler was used to analyze the posterior distributions. Random-walk Metropolis-Hastings algorithm was used for α and ϕ. We used a log-normal proposal distribution for ϕ and a normal proposal for logit(α). R code for fitting the model and for generating an example dataset are provided in the Supplementary Materials. We generated 25,000 samples and discarded the first 10,000 samples as burn-in. Convergence was monitored using trace plots and autocorrelation plots for several representative parameters.

2. Confounders in the Preterm Birth Model

The PTB model includes the following confounders: a non-linear effect of maternal age modeled using natural cubic splines with 3 degrees of freedom, maternal education ($< 12^{\mathrm{th}}$ grade, high school or GED, some college or higher), race/ethnicity (non-Hispanic black, non-Hispanic white, Hispanic, Asian), reported tobacco use during pregnancy (yes or no), firstborn (yes or no), marital status (married or unmarried), infant sex (male or female), and percent population below poverty at the census block group level. To control for unmeasured time-varying confounders, we included the season of conception (winter: Dec-Feb, spring: Mar-May, summer: Jun-Aug, autumn: Sep-Nov), a conception date (natural cubic splines with 7 degrees of freedom) to capture long-term trends, and 1-week lag temperature (natural cubic splines with 3 degrees of freedom). We assumed the effects of all confounder covariates to be constant in time.

Higher risks of PTB were associated with male infants and those born to unmarried, less educated, non-Hispanic black mothers, as well as among those who reported tobacco use during pregnancy and were living in an area with higher proportion of households below poverty line (Supplementary Table S1). Supplementary Figure S2 shows the non-linear effects of maternal age and conception date. We found that higher rates of PTB were associated with younger and older mothers, and pregnancies conceived in recent years. The significant increasing trend of PTB rate may be attributed to increasing number of cesarean deliveries for medically-indicated PTB. The above associations between PTB and demographic variables are consistent with previous studies in different populations. We did not observe an association between PTB and high average 1-week lagged temperature.

3. Supplementary Tables

Table S1. Simulation study results for estimating cumulative $PM_{2.5}$ effects combined across outcome weeks: relative change in bias and length of 95% posterior intervals (PI), averaged across 20 simulated replicate datasets. The reference corresponds a time-to-event model where the exposure is averaged across known exposure and outcome weeks. The standard deviations across replicate datasets are given in parentheses.

	Exposure Weeks	Outcome Weeks	Prior Structure	Δ Bias ($\times$1000)	Relative Δ $PI_{95\%}$ length
Scenario 1	14-26	27-36	Known windows	Reference	
			Dynamic	-0.9 (0.47)	0.84 (0.06)
			Exchangeable	-4.2 (0.60)	0.93 (0.10)
			Independent	-5.8 (0.55)	1.00 (0.10)
Scenario 2	14-19	27-36	Known windows	Reference	
			Dynamic	-6.4 (1.9)	0.89 (0.05)
			Exchangeable	-10.9 (1.9)	0.80 (0.06)
			Independent	-12.7 (2.0)	0.81 (0.08)
Scenario 3	14-26	27-31	Known windows	Reference	
			Dynamic	-2.6 (1.3)	0.73 (0.06)
			Exchangeable	-8.6 (1.5)	0.87 (0.16)
			Independent	-11.0 (1.0)	0.93 (0.24)
Scenario 4	4-wk lag	27-36	Known windows	Reference	
			Dynamic	-4.8 (1.2)	0.81 (0.09)
			Exchangeable	-9.0 (1.1)	0.71 (0.07)
			Independent	-9.7 (1.1)	0.66 (0.05)
Scenario 5	4-wk lag	31-36	Known windows	Reference	
			Dynamic	-4.6 (1.0)	0.84 (0.07)
			Exchangeable	-6.8 (1.1)	0.81 (0.04)
			Independent	-4.6 (1.3)	0.80 (0.06)

Table S2. Posterior mean and 95% posterior interval (P.I.) for the relative increase in PTB risk associated with various factors.

	Estimate (95% P.I.)
Male	1.07 (1.04, 1.09)
Tobacco use during pregnancy	1.50 (1.42, 1.58)
Married	0.87 (0.85, 0.90)
Firstborn	1.01 (0.98, 1.03)
Tract-level % poverty (×10)	1.04 (1.02, 1.05)
Race/ethnicity	
Non-Hispanic white	Reference
Non-Hispanic black	1.48 (1.43, 1.53)
Hispanic	1.00 (0.96, 1.04)
Asian	1.00 (0.93, 1.06)
Mother's education (years)	
$< 12^{\text{th}}$ grade	1.10 (1.06, 1.14)
high school or GED	1.17 (1.13, 1.22)
some college or higher	Reference
Conception season	
Sep-Nov	Reference
Dec-Feb	1.00 (0.94, 1.03)
Mar-May	1.00 (0.95, 1.05)
Jun-Aug	0.99 (0.96, 1.04)

4. Supplementary Figures

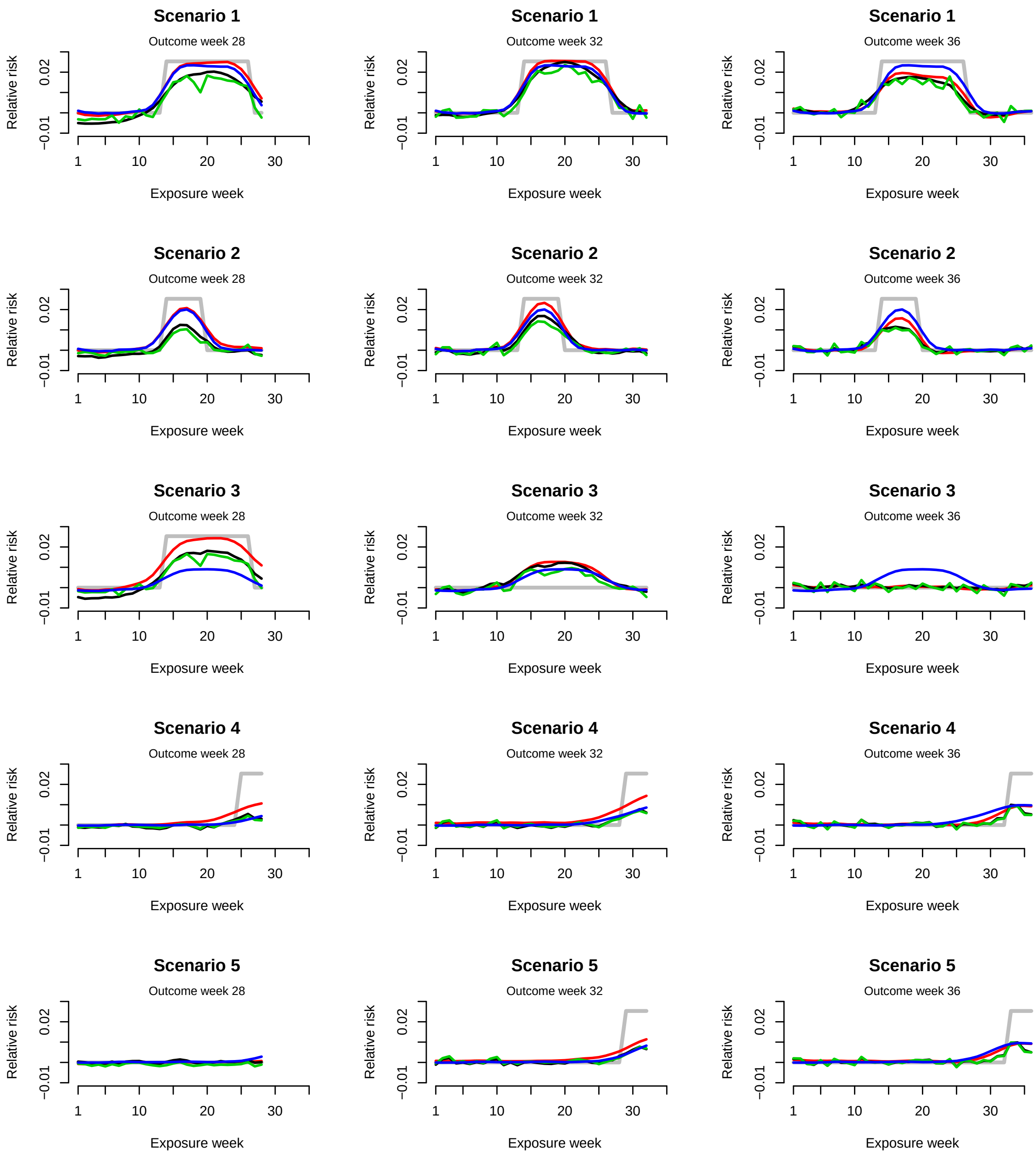

Fig. S1. Estimated weekly effects by outcome weeks (averaged across 20 simulation replicates). Four different prior structures were considered: dynamic (red), exchangeable (black), independent (green), and fixed (blue). The true effect is shown in grey.

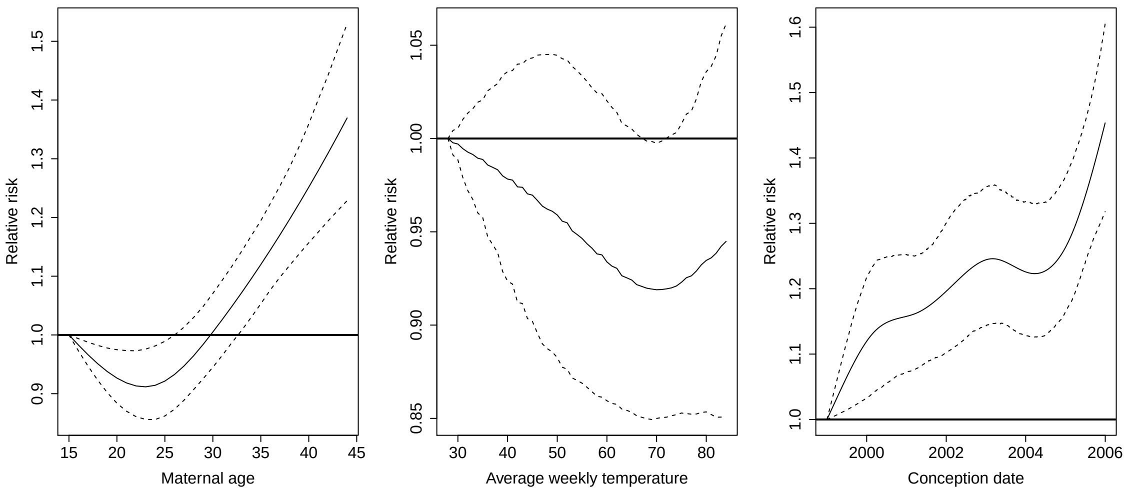

Fig. S2. Non-linear effects of maternal age, average 1-week lagged temperature, and conception date on the relative risk of preterm birth.

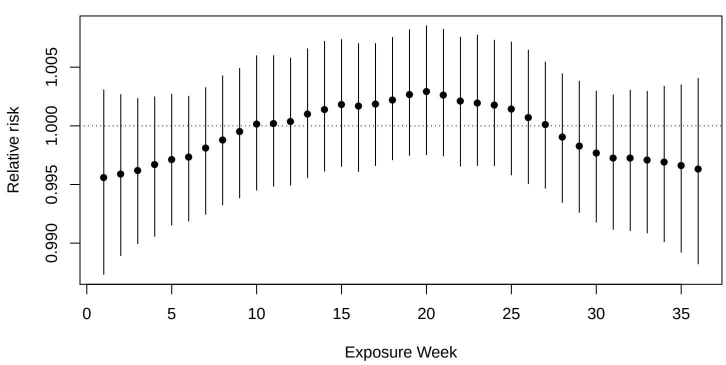

Fig. S3. Posterior means and 95% posterior intervals of the relative risks for preterm birth associated with an interquartile range (7.4 $\mu g/m^3$) increase in weekly $PM_{2.5}$ exposure, estimated from a modeling assuming weekly effects do not vary across outcome weeks.