Analysis of Cardiac Birth Defects and Air Pollution in Texas

Laura Boehm

INTRODUCTION

Previous research has suggested a link between maternal exposure to various airborne pollutants and selected birth defects (Gilboa et al., Ritz et al.) Defining time periods of possible critical exposure over which to test a hypothesis is difficult. Should we consider the total pollution a mother is exposed to, or only a certain critical window? This study uses Bayesian Stochastic Search Variable Selection (SSVS) to determine critical windows and spatial random effects for each of the six study counties are considered to account for overdispersion while allowing nearby counties to borrow strength from their neighbors. The analysis is completed using WinBUGS.

THE DATA

In this study, we consider a dataset of births to mothers over age 18 in six Texas counties (Bexar, Dallas, Ellis, Harris, Tarrant and Travis) in the period 1997-2000. Average Carbon Monoxide (CO) concentrations were collected daily from the nearest active monitor to maternal residence. The daily average for the county was also measured. The study design is a matched case control, with the number of controls approximately twice the number with the most common defect, Ventricular Septal Defect (VSD), resulting in 4091 births observed. Of these, 3936 had with no missing covariates, including 2274 cases, and 1683 unique individuals with defects.

IMPORTANT COVARIATES

For this study, the covariates included are maternal age, plurality, the sex of the child, mother's education, indicator for maternal illness, and race/ethnicity. Maternal age ranged from 18-46 with median 27, and education ranged from 0-17 with median 12. Plurality is recorded as 1 (95% of births), 2 (4%) and 3 or more (.4%). Race categories are Hispanic White (41%), Non-Hispanic White (40%), Non-Hispanic Black (14%) and Other (4%), which consists primarily of Asian/Pacific Islanders, and a small number of Native Americans and Hispanic Blacks. Half the children were of each sex, and less than 8% of mothers were ill. These covariates were selected based on stepwise logistic regression; Additionally, SSVS conducted with all possible covariates (but not spatial effects) included these in the model with highest probability. Although only the "Other"race/ethnicity category was significant, the other categories are included for completeness.

DEFINING WINDOWS

In this analysis, 10 windows were defined by Gaussian-weighted averages; each window had a different center, mu, spaced approximately 14 days apart, with standard deviation of 7 days. In this way, there is considerable between each window. Weights on each window were adjusted so that they summed to one. Prior to the weighting, the daily pollution measurements were weighted and scaled. This results in ten new variables $\mathbf{z}_1,\ldots,\mathbf{z}_{10}$, where z_{ik} is the value of the k^{th} window for the i^{th} mother.

$$w^*_{tk} = e^{\frac{-1}{2\sigma^2}(t-\mu_k)^2}, \quad w_{tk} = \frac{w^*_{tk}}{\sum_{t=1}^{135} w^*_{tk}}, \quad z_{ik} = \sum_{t=1}^{135} w_{tk} Poll_{it}$$

The model was then defined as a logistic regression, with Y=1 being the presence of a specific birth defect, and x being the covariates listed above.

$$Y_i \sim Bin(1, \pi_i), \qquad \text{logit}\pi_i = x_i^T \gamma + \sum_{k=1}^{10} \alpha_k z_{ik}$$

Then, using SSVS methods, we determine which of those 10 z variables are most probably in the model. To do this, a binary variable g_k is included for each z, so that α, the coefficient on z, is drawn from a mixture of normal distributions, with τ^2, the precision when g_k=1, is small, and τ_2^2 is quite large.

$$g_k \sim Bernoulli(p_k), \qquad \alpha_k \sim Normal(\mu, \tau^2 + (1-g_k)\tau_2^2)$$

The priors chosen here are τ^2=11, τ_2^2=10000, p_k=.5.

SPATIAL RANDOM EFFECTS

Spatial random effects for the different counties were added to account for overdispersion in the model while allowing nearby counties to borrow strength from their neighbors. These effects were given a CAR prior, with the weights for neighbors equal to the inverse distance between the two county population centroids.

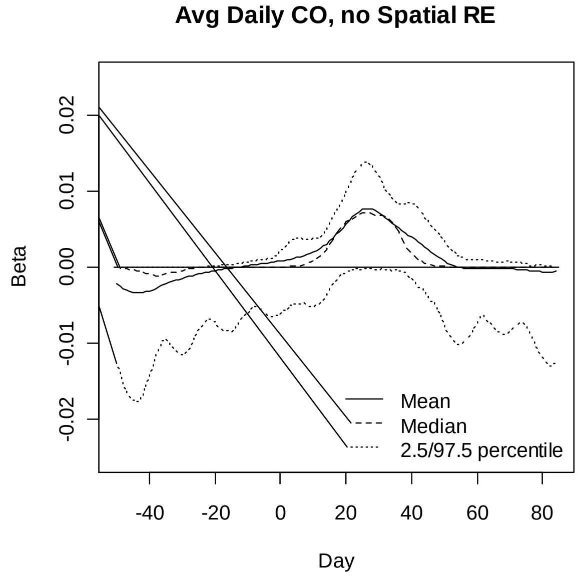
$$\phi_s \mid \phi_{r\neq s} \sim N\left(\frac{\sum_{r\neq s} c_{rs}\phi_r}{\sum_{r\neq s} c_{rs}} \;,\; \frac{1}{\tau \sum_{r\neq s} c_{rs}} \right), \qquad c_{rs} = \frac{1}{dist(county_r, county_s)}$$

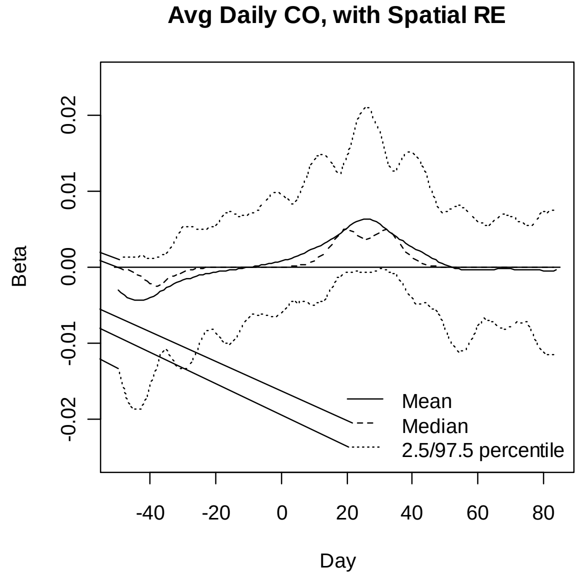
When these were added to the same model above, the effects of CO pollution were not changed very much, but there was more noise in the runs.

RESULTS

To create the quantiles and means on the plots above, the coefficients α were transformed into daily coefficients,β, by multiplying the weight matrix times α. The mean and quantiles were then found for each daily coefficient β. To summarize, we consider the change in odds of VSD if CO exposure increased one standard deviation for the entire study period by summing the daily coefficients, or similarly over a certain period by summing over those days. Here we consider the entire period and the period of weeks 2-6 after conception (most highly positive).

The other covariates included in the model had similar coefficients and credible intervals for the model with and without spatial random effects. The only random effect with consistent sign is the one corresponding to Dallas county (2) which is negative.





	Entire period	Weeks 1-6
With Random Effects	1.13 (.39, 1.52)	1.08 (.36, 3.2)
Without Random Effects	1.18 (.97, 1.33)	1.14 (.95, 1.58)

```
              mean    sd   2.5%    97.5%
Covariates:
plurality   0.50   0.15    0.21   0.77
age         0.02   0.01    0.01   0.04
sex        -0.18   0.07   -0.32  -0.04
education  -0.04   0.01   -0.06  -0.01
illness     0.43   0.13    0.17   0.68
NHWhite     0.10   0.10   -0.09   0.28
NHBlack    -0.07   0.14   -0.35   0.20
Other      -0.67   0.23   -1.14  -0.25
Spatial Parameters:
tau          153    107      24    427
phi[1]     -0.07   0.11   -0.28   0.14
phi[2]     -0.27   0.09   -0.43  -0.11
phi[3]      0.13   0.20   -0.24   0.53
phi[4]      0.09   0.08   -0.05   0.24
phi[5]      0.10   0.09   -0.08   0.27
phi[6]      0.01   0.17   -0.33   0.36
```

FURTHER RESEARCH

More exploration of the data can be done. Several things have been tried, including putting correlation between consecutive windows in the SVSS model, which results in a smoother graph. The Ozone data needs to be further analyzed also, and temperature information should be included; one way to do this may be to remove seasonal trend. A model with multivariate response (one for each defect) may also help in detecting associations by borrowing strength across defects.