Web Appendix

Bayesian Ensemble Averaging

The Bayesian ensemble averaging approach described below was developed by Balachandran et al., 2013 (1).

A drawback to typical ensemble averaging approaches is that the ensemble weights associated with each source apportionment (SA) method are assumed to be known and the same for each day. The standard ensemble weights are defined as the root mean-squared error (RMSE) between daily source impacts from each SA method to an overall average. The main objective of the Bayesian ensemble approach is to incorporate uncertainty in the ensemble weights. Moreover, to better represent reality, the ensemble weights are also allowed to vary across days. This is accomplished by assuming the true daily uncertainties arise from a distribution with mean equal to the overall RMSE. Specifically, S_{jlk}, the concentration from source j and method l on day k, can be viewed as an error-prone measure of the true source concentration and the ensemble of these SA methods, $\overline{S}_{jk}$, can be treated as the true source concentration. We also assume that these errors are normally distributed so that for any day k:

$$S_{jlk} - \overline{S}_{jk} \sim Normal\,(0, \tau^2_{jlk}) \qquad \text{(Web Equation 1)}$$

A Bayesian approach to estimate $\overline{S}_{jk}$ was adopted in order to obtain posterior distributions of τ^2_{jlk}, which could then be used to calculate an ensemble average. First an inverse-gamma (scaled-inverse-chi-squared) distribution was assigned to each variance component. The density function for the inverse-gamma (IG) distribution is specified with two known parameters α and β, and denoted as IG (α, β):

$$f(\tau_{jlk}^2 \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} (\tau_{jlk}^2)^{-\alpha-1} \exp(-\frac{\beta}{\tau_{jlk}^2}) \qquad \text{(Web Equation 2)}$$

The error of the data (S_{jl} (k=1) . . . S_{jl} (k=K)) with respect to the average $\overline{S}_{jk}$, has a likelihood given by the normal density:

$$f(\text{data} \mid \tau_{jlk}^2) = (2\pi\tau_{jlk}^2)^{-\frac{K}{2}} \exp(-\frac{1}{2\tau_{jlk}^2} \sum_{k=1}^{K} \left(S_{jlk} - \overline{S}_{jk}\right)^2) \qquad \text{(Web Equation 3)}$$

The posterior distribution of τ_{jlk}^2 given the data is expressed as:

$$f(\tau_{jlk}^2 \mid \text{data}) \propto f(\text{data} \mid \tau_{jlk}^2) \times f(\tau_{jlk}^2 \mid \alpha, \beta)$$

(Web Equation 4)

$$= (\tau_{jlk}^2)^{-(\alpha+\frac{K}{2})-1} \exp(-\frac{1}{\tau_{jlk}^2} \left[\beta + \frac{1}{2} \sum_{k=1}^{K} \left(S_{jlk} - \overline{S}_{jk}\right)^2 \right])$$

The previous expression is proportional to an inverse-gamma distribution:

$$IG\left(\alpha + \frac{K}{2}, \left[\beta + \frac{1}{2} \sum_{k=1}^{K} \left(S_{jlk} - \overline{S}_{jk}\right)^2 \right] \right) \qquad \text{(Web Equation 5)}$$

And this distribution has mean:

$$\frac{\left[\beta + \frac{1}{2} \sum_{k=1}^{K} \left(S_{jlk} - \overline{S}_{jk}\right)^2 \right]}{\alpha + \frac{K}{2}} (1) \qquad \text{(Web Equation 6)}$$

When α and β are small, the above mean is approximately the square of the RMSE. An advantage of a Bayesian approach is that prior information about τ_{jlk}^2 can be incorporated in α and β. To reflect our lack of knowledge about SA method uncertainties we chose non-informative priors by setting $\alpha = \beta = 0.0001$. Through this approach we were able to sample multiple realizations of weights that were used in ensemble-averaging. Stochastic variation in the sampled weights reflects uncertainties in the ensemble averages. Ensemble-averaging was conducted for 30 days in summer (July 2001) and 30 days in winter (January 2002). For each day in the ensemble, we used 30 samples from the posterior distributions, resulting in 30 ensemble-averaged source concentration estimates for each of 30 days in the short term period. To generate the long-term (8.5 year) time series used in the analysis, each day 10 realizations of the ensemble source impacts were used to obtain 10 source profiles based on CMB equation. These 10 source profiles were then used to obtain daily concentrations of each source.

Health Models

Poisson generalized linear regression was used to model the logarithm of the expected daily count of emergency department visits for pediatric asthma (Y) as a function of the covariates and $PM_{2.5}$ source (S_j). For each exposure window (lags 0-2 and lags 0-7) three different models were considered: a single-source model, single-source model with O3 control and all-source model.

Single-source model

This model considered source exposures individually and took the following form:

$$\log[E(Y)] = \alpha + \beta_1 S_{j0} + \beta_2 S_{j1} + \cdots + \beta_8 S_{j7} + g\big(\theta_1, \ldots \theta_n; day\ of\ study\big) + h(temp_{0-2}) + h(dewpt_{0-2}) + h(temp_{3-7}) + h(dewpt_{3-7}) + \sum_i \tau_i (day\ of\ week\ or\ FH)_i + \delta_1 (day\ after\ TG) + \delta_2 (day\ after\ Xmas) + \sum_j \varphi_j (hospital)_j + \sum_k \vartheta_k (season)_k + \sum_k \mu_k (season)_k \big(h(temp_{0-2})\big) + \sum_k \gamma_k (season)_k \big(h(temp_{3-7})\big) \quad \text{(Web Equation 7)}$$

The dependent variable (Y) was the count of emergency department visits for pediatric asthma. Our primary exposure (S_j) was the Bayesian-based ensemble source concentration estimate for source *j* (i.e., BURN, COAL, DUST, DV, GV or SOC). To model the 8-day exposure we included individual terms for lag 0 (S_{j0}) through lag 7 (S_{j7}) of the source concentration. Long-term temporal trends were controlled for with a cubic spline (g) with 8 knots per year. Meteorological trends were controlled for using two time intervals (lags 0-2 and lags 3-7) to better match our exposure windows. This control included: cubic polynomials [*h()*] for the 3-day moving average maximum temperature ($temp_{0-2}$), 3-day moving average dew point ($dewpt_{0-2}$), 5-day moving average maximum temperature ($temp_{3-7}$) and 5-day moving average dew point ($dewpt_{3-7}$). Indicator variables were included for day of week or federal holidays (*day of week or FH)*, with holidays receiving a separate indicator; the day after Thanksgiving (*day after TG)*; the day after Christmas (*day after Xmas)*; hospital, and season. Interaction terms between season and the cubic polynomials for maximum temperature lags 0-2 and lags 3-7 were also included.

To calculate the 3-day and 8-day exposures to a given source we exponentiated the sum of the beta coefficients for the relevant source exposures. For example, the rate ratio for a 3-day source exposure was calculated as RR = $\exp(\beta_1+ \beta_2+ \beta_3)$, while the rate ratio for an 8-day source exposure was $\exp(\beta_1+ \beta_2\ldots+ \beta_8)$. Note that the RR for the 3-day source exposure should be interpreted as the 3-day association *controlling* for exposure to lags 3-7, since those terms remained in the model. The standard errors were calculated using the estimated covariance matrix from the SAS "genmod" procedure.

Single-source model with O3 control

This model was identical to the model presented in Web Equation 7 with the addition of individual terms for the 8-day (lags 0-7) exposure to O_3 (8-hour daily maximum ozone) included in the model as: $\eta_1X_0 + \eta_2X_1 + \eta_3X_2 + \eta_4X_3 + \eta_5X_4 + \eta_6X_5 + \eta_7X_6 + \eta_8X_7$, where X_k represents O_3 concentration on lag day *k*. The 3-day and 8-day RR for the source exposures were calculated as described in the single-source model above. In addition, we calculated the cumulative effect of a 25ppb increase in O_3 for lags 0-7 using the following equation:

$$RR_{O3} = \exp(25\sum_{i=1}^{8}\eta_i) \qquad \text{(Web Equation 8)}$$

where η_i is the coefficient for a given O_3 lag.

All-sources model

To account for potential confounding by other sources we also considered a model that included all six $PM_{2.5}$ sources simultaneously. The main exposure source (for which the RR was estimated) was modeled using the same unconstrained distributed lag structure as in Web Equation 7. The remaining five sources were controlled for in the model with a single term for the 8-day moving average concentration as follows:

$$S_{j,0-7} = {\sum_{k=0}^{7} S_{jk}}\Big/{8} \qquad \text{(Web Equation 9)}$$

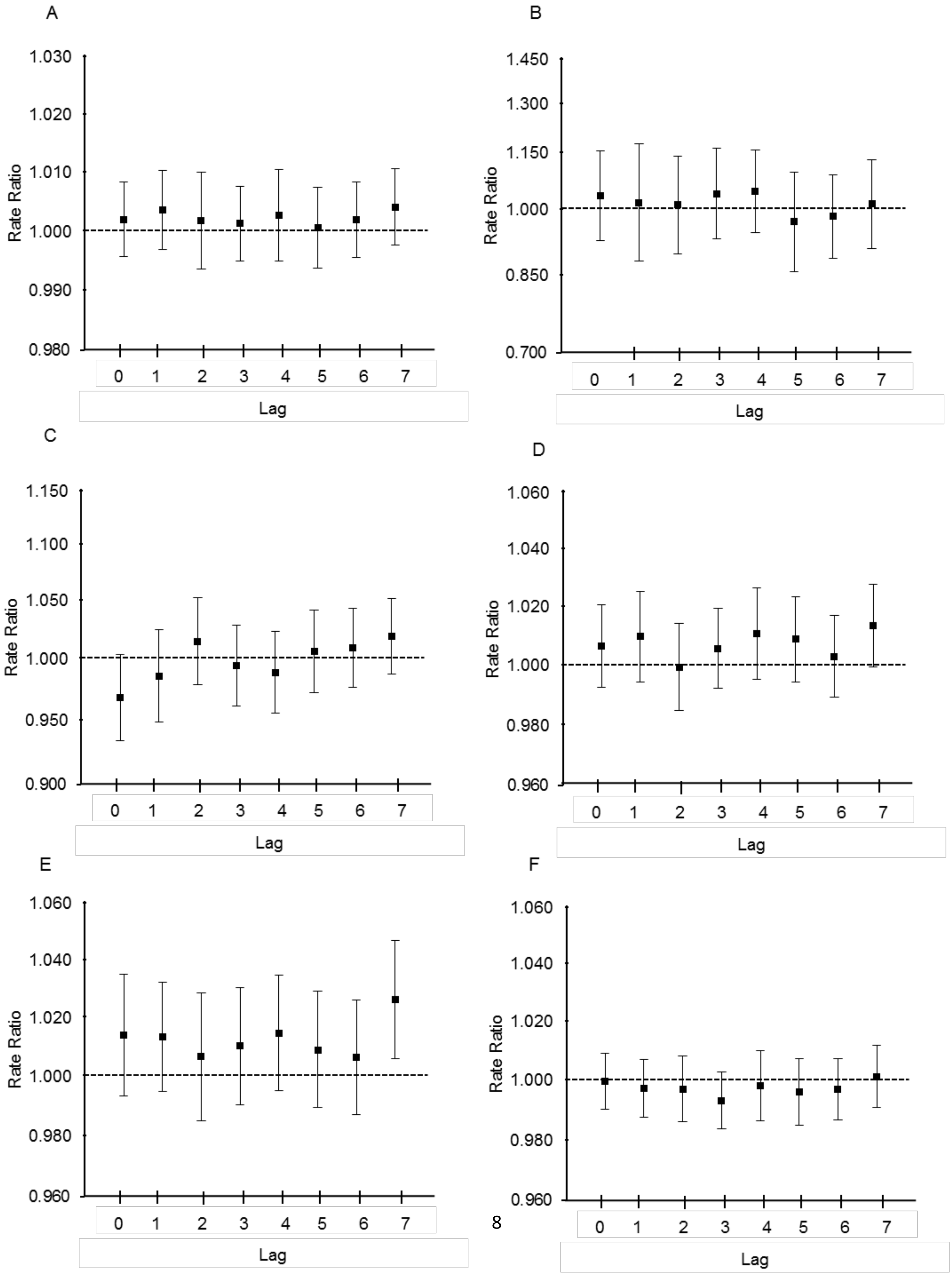
where S_{ji} is the concentration for source *j* on lag day *i*. We chose to control for the 8-day moving average source concentrations, rather than including multiple terms for the distributed lag for each of the 5 additional sources, in order to minimize issues of multicollinearity. Apart from the addition of the other five sources in the model, the covariate control was identical to that in Web Equation 7. The RR and standard errors for the main exposure source were calculated as described in the single-source model.

References

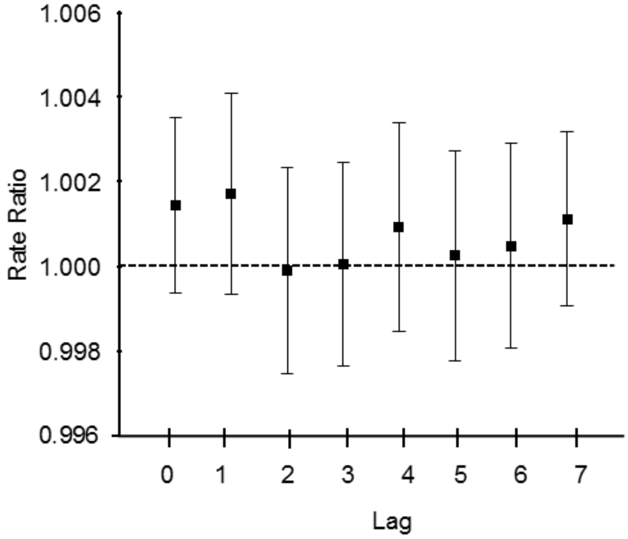
1. Balachandran S, Chang HH, Pachon JE, et al. Bayesian-based ensemble source apportionment of $PM_{2.5}$. *Environmental science & technology* 2013;47(23):13511-8.

Web Table 1. Spearman correlations between lag 0 (same day) and previous 7 days, for each source as well as total $PM_{2.5}$ mass and Ozone, from a single ensemble run in Atlanta, GA (2002-2010).

Lag	Biomass Burning	Primary Coal Combustion	Dust/ Resuspended Soil	Diesel Vehicles	Gasoline Vehicles	Secondary Organic Carbon	$PM_{2.5}$	O_3
Lag 0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Lag 1	0.59	0.31	0.63	0.38	0.32	0.54	0.62	0.77
Lag 2	0.49	0.18	0.39	0.18	0.13	0.44	0.31	0.64
Lag 3	0.45	0.12	0.31	0.14	0.09	0.43	0.22	0.59
Lag 4	0.46	0.13	0.30	0.17	0.11	0.44	0.23	0.58
Lag 5	0.42	0.13	0.30	0.14	0.13	0.42	0.22	0.57
Lag 6	0.43	0.12	0.32	0.17	0.13	0.39	0.21	0.56
Lag 7	0.41	0.13	0.32	0.14	0.13	0.38	0.19	0.57


A
1.030
1.020
1.010
1.000
0.990
0.980
Rate Ratio
0 1 2 3 4 5 6 7
Lag
B
1.450
1.300
1.150
1.000
0.850
0.700
Rate Ratio
0 1 2 3 4 5 6 7
Lag
C
1.150
1.100
1.050
1.000
0.950
0.900
Rate Ratio
0 1 2 3 4 5 6 7
Lag
D
1.060
1.040
1.020
1.000
0.980
0.960
Rate Ratio
0 1 2 3 4 5 6 7
Lag
E
1.060
1.040
1.020
1.000
0.980
0.960
Rate Ratio
0 1 2 3 4 5 6 7
Lag
F
1.060
1.040
1.020
1.000
0.980
0.960
Rate Ratio
0 1 2 3 4 5 6 7
Lag

Web Figure 1. Rate Ratios and 95% Confidence Intervals for the Single-Day Association of a 1 μg/m^3 Increase in Source Concentration on Pediatric Asthma Emergency Department Visits Presented for Lags 0 Through 7 in Atlanta, GA (2002-2010) for the Following Sources: A) Biomass burning, B) Primary Coal Combustion, C) Dust/Resuspended Soil, D) Diesel Vehicles, E) Gasoline Vehicles, and F) Secondary Organic Carbon. Results are generated from the single-source model with an unconstrained distributed lag structure. The rate ratios and 95% confidence intervals corresponding to this figure are in Web Table 3.



Web Figure 2. Rate Ratios and 95% Confidence Intervals for the Single-Day Effect of a 1 µg/m3 Increase in Total $PM_{2.5}$ Concentration on Pediatric Asthma Emergency Department Visits Presented for Lags 0 Through 7 in Atlanta, GA (2002-2010). Results are generated from the single-source model with an unconstrained distributed lag structure.

Web Table 2. Rate Ratios and 95% Confidence Intervals for the Association of $PM_{2.5}$ Sources, Total $PM_{2.5}$, and Ozone on Pediatric Asthma Emergency Department Visits in Atlanta, GA (2002-2010). These results correspond to Figures 1-3.

Source	Lag	Single-Source Model			Single-Source Model with O_3 Control			All-Sources Model		
		RR	**95% CI**	**Scale[a]**	**RR**	**95% CI**	**Scale[a]**	**RR**	**95% CI**	**Scale[a]**
Biomass burning[b]	0-2	1.007	0.997, 1.017	1.457	1.007	0.997, 1.018	1.430	1.001	0.989, 1.013	1.448
(BURN)	0-7	1.017	0.995, 1.04	1.457	1.018	0.998, 1.039	1.430	1.001	0.974, 1.029	1.448
Primary coal combustion[b]	0-2	1.059	0.822, 1.363	1.460	1.084	0.839, 1.402	1.433	1.025	0.803, 1.308	1.448
(COAL)	0-7	1.103	0.703, 1.731	1.460	1.152	0.717, 1.850	1.433	0.994	0.644, 1.534	1.448
Dust/resuspended soil[b]	0-2	0.967	0.922, 1.013	1.458	0.981	0.926, 1.038	1.431	0.949	0.903, 0.997	1.445
(DUST)	0-7	0.980	0.914, 1.050	1.458	1.025	0.940, 1.117	1.431	0.943	0.874, 1.019	1.445
Diesel Vehicles[b]	0-2	1.015	0.991, 1.039	1.456	0.999	0.973, 1.026	1.432	1.020	0.99, 1.051	1.448
(DV)	0-7	1.057	1.012, 1.104	1.456	1.020	0.970, 1.073	1.432	1.072	1.004, 1.144	1.448
Gasoline Vehicles[b]	0-2	1.033	1.001, 1.067	1.454	1.031	0.996, 1.067	1.430	1.020	0.974, 1.07	1.448
(GV)	0-7	1.103	1.039, 1.170	1.454	1.077	1.012, 1.147	1.430	1.068	0.962, 1.185	1.448
Secondary Organic Carbon[b]	0-2	0.993	0.977, 1.010	1.459	0.986	0.968, 1.004	1.430	0.984	0.967, 1.001	1.448
(SOC)	0-7	0.978	0.946, 1.011	1.459	0.963	0.930, 0.996	1.430	0.953	0.92, 0.987	1.448
PM_{25}[b]	0-2	1.003	1.000, 1.006	1.458	1.002	0.999, 1.005	1.431			
	0-7	1.006	1.002, 1.010	1.458	1.006	1.000, 1.011	1.431			
O_3 only[c]	0-7				1.115	1.034, 1.202	1.432			
O_3-BURN[c]	0-7				1.114	1.031, 1.203	1.430			
O_3-COAL[c]	0-7				1.117	1.034, 1.207	1.433			
O_3-DUST[c]	0-7				1.130	1.043, 1.224	1.431			
O_3-DV[c]	0-7				1.107	1.024, 1.197	1.432			
O_3-GV[c]	0-7				1.113	1.030, 1.202	1.430			
O_3-SOC[c]	0-7				1.145	1.057, 1.240	1.430			
O_3-PM_{25}[c]	0-7				1.073	0.988, 1.167	1.431			

[a]Defined as Pearson's chi-square, divided by the degrees of freedom; equivalent to $\sqrt{\phi}$, where ϕ is the overdispersion parameter
[b]Rate ratios are for a 1 μg/m^3 increase
[c]Rate ratios are for a 25ppb increase

Web Table 3. Rate Ratios and 95% Confidence Intervals for the Single-Day Association of a 1 µg/m3 Increase in Source Concentration on Pediatric Asthma Emergency Department Visits Presented for Lags 0 Through 7 in Atlanta, GA (2002-2010). These results correspond to the results presented in Web Figures 1 and 2.

	Biomass Burning		Primary Coal Combustion		Dust/Resuspended Soil		Diesel Vehicles		Gasoline Vehicles		Secondary Organic Carbon		Total $PM_{2.5}$	
LAG	**RR**	**95% CI**	**RR**	**95% CI**	**RR**	**95% CI**	**RR**	**95% CI**	**RR**	**95% CI**	**RR**	**95% CI**	**RR**	**95% CI**
0	1.002	0.996, 1.008	1.033	0.924, 1.154	0.968	0.934, 1.003	1.006	0.992, 1.02	1.014	0.993, 1.035	0.999	0.99, 1.009	1.001	0.999, 1.004
1	1.003	0.997, 1.010	1.015	0.878, 1.174	0.985	0.948, 1.024	1.009	0.994, 1.025	1.013	0.995, 1.032	0.997	0.987, 1.007	1.002	0.999, 1.004
2	1.002	0.994, 1.010	1.010	0.895, 1.139	1.014	0.978, 1.051	0.999	0.985, 1.014	1.006	0.985, 1.028	0.997	0.986, 1.008	1.000	0.997, 1.002
3	1.001	0.995, 1.007	1.038	0.929, 1.161	0.993	0.961, 1.027	1.006	0.992, 1.019	1.010	0.99, 1.030	0.993	0.984, 1.002	1.000	0.998, 1.002
4	1.003	0.995, 1.01	1.044	0.942, 1.157	0.988	0.955, 1.022	1.011	0.995, 1.026	1.014	0.995, 1.034	0.998	0.986, 1.010	1.001	0.998, 1.003
5	1.001	0.994, 1.007	0.968	0.856, 1.095	1.005	0.971, 1.040	1.009	0.994, 1.023	1.009	0.989, 1.029	0.996	0.985, 1.007	1.000	0.998, 1.003
6	1.002	0.995, 1.008	0.981	0.885, 1.087	1.009	0.976, 1.042	1.003	0.989, 1.017	1.006	0.987, 1.026	0.997	0.986, 1.007	1.000	0.998, 1.003
7	1.004	0.997, 1.011	1.012	0.907, 1.129	1.018	0.986, 1.051	1.013	0.999, 1.027	1.026	1.006, 1.047	1.001	0.991, 1.012	1.001	0.999, 1.003