Air Pollution and Acute Respiratory Infections among Children 0-4 Years: an 18-year Time-series Study

Web Appendix 1

Modeling Approach

In time-series analyses of the short-term associations between air pollution levels and exacerbation of respiratory disease potential confounders are factors which change over short periods of time (e.g., day of week, meteorology, holidays) and those that change seasonally and longer term, which are commonly unmeasured.

Seasonal and long-term trends. Seasonal and long term trends in the outcomes are driven by seasonal variation in circulating infections, as well as long term changes in the source population and ED usage patterns. Because these factors are unmeasured, we control for calendar time as a proxy of these factors by including cubic splines with 1 knot per month (12 knots/year) on the time variable (day of study). The choice of degrees of freedom in the splines is not straightforward. Basing this decision strictly on goodness of fit statistics (i.e., maximizing prediction) can yield biased effect estimates for pollution.[1] Figure E1 shows the predicted ED counts for the 2000-2008 time periods based solely on the smooth function of time used in our primary analyses (cubic splines with 12 knots/year). The splines account for the overall seasonal trend of these outcomes as well as some finer scale temporal variation, allowing pollution and the other model covariates to explain the remaining fine-scale temporal variation in ED counts. Our inclusion of a relatively high number of degrees of freedom in the splines may have reduced the possibility of a spurious association, but also potentially reduced our ability to detect true associations between pollutant concentrations and respiratory ED visits.

Because results were sensitive to the choice of degrees of freedom for some pollutant-outcome associations, in Figure E2 we present rate ratios per interquartile range increase in pollution for models with various degrees of freedom per year (2, 4, 6, 8, 10, 12, 14, and 16). Note that estimates for models with 2 degrees of freedom are typically outliers and reflect gross confounding by seasonal patterns. In general, for the positive associations observed in our primary analyses, inclusion of fewer degrees of freedom per year yielded slightly higher estimates and inclusion of additional degrees of freedom yielded slightly lower estimates.

Selection of meteorological control. Our models included cubic polynomials for the three-day moving average of mean dew point and maximum temperature (Celsius). Maximum temperature was chosen over mean and minimum temperature for the primary models because it was the stronger predictor of the pollutant concentrations (based on the AIC, adjusting for other time-varying covariates included in the primary models for respiratory ED visits). Mean, maximum, and minimum temperature were only weakly predictive of the respiratory outcomes. In sensitivity analyses we assessed (a) cubic splines with knots at the 25^{th} and 75^{th} percentile of lag 0-1-2 moving average maximum temperature (i.e., adding two degrees of freedom to the cubic polynomials) (b) inclusion of mean temperature instead of maximum temperature (c) separate cubic polynomials for lag 0 and lag 1-2 moving average maximum temperature and mean dew point temperature (d) interaction of the meteorological variables with the 4-level season variable (e) addition of indicator variables for the top and bottom 1% of temperature days, (f) inclusion of three-day moving average precipitation. None of these model variants resulted in meaningfully changes in the pollutant estimates. This was not surprising as these factors were not strong predictors of the respiratory outcomes after accounting for the smooth functions of time, which may have captured much of the temperature effect.

Weekday. Of the strongest confounders in in our analysis was day of week, which is related to both pollutant emissions as well as population ED usage patterns. The interaction terms between day of week and a four-level season variable (winter, spring, summer, fall) were highly significant and changed some of the pollutant estimates; finer-scale interaction by month of year (Jan, Feb, etc.) did not further meaningfully change the pollutant estimates so interaction with the 4-level season variable was deemed sufficient to control for confounding by weekday.

Holidays. Because holidays impact both air pollution levels and ED usage patterns we included indicator variables for federal holidays and lag holiday (i.e., whether a holiday occurred on lag 1 or lag 2). Examination of residuals led to identification of additional holiday periods that were not official federal holidays but had systematically under- or over-predicted ED counts: Thanksgiving weekend, December 26th, January 1-3rd and the week between Christmas and New Year's Day. Indicators for these holiday periods were added to the primary model.

Given the acute changes in behavior due to the December and January holidays which affect air pollution levels (fewer people commuting), respiratory infections (school vacations, travel, family gatherings) and utilization of the emergency department (e.g., doctor's offices are closed) we were concerned about the impact of the mid-December to mid-January period on the estimation of pollutant associations. However, analyses excluding December 15th through January 15th each year did not meaningfully change the pollutant estimates, thus we retained this time period in the main analyses.

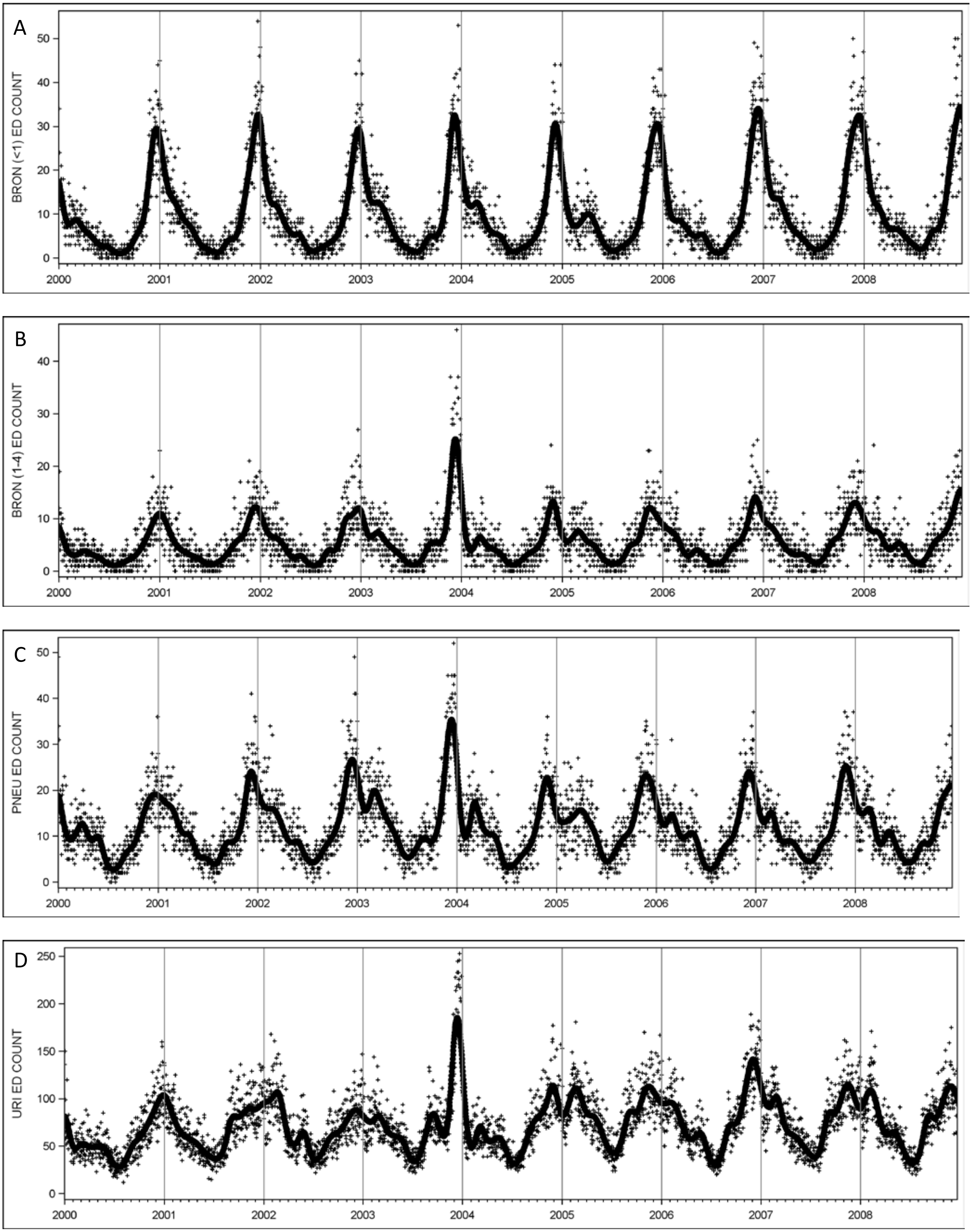
Autocorrelation. Due to autocorrelation in the residuals for the upper respiratory infection (URI) outcome only, we also used generalized estimating equations (GEE) with a stationary 4-dependent correlation structure to estimate pollutant effects for URI. Rate Ratios and confidence intervals for the pollutants were virtually identical to those obtained using generalized linear models.

Population-weighted daily air quality estimates. A detailed description of our approach to calculate daily population-weighted concentrations for each pollutant is provided in a previous publication (2). Briefly, we created statistical models to characterize the spatial variability of ambient air pollutant concentrations in the study area. At each monitor the measurements were log-transformed and then standardized using the mean and standard deviation at that monitor for a given study year. Daily surfaces for the study area were created by inverse distance-square weighting the standardized values. The daily pollutant concentrations at each Census tract centroid within 20-county metropolitan Atlanta were estimated by converting the standardized value back to a concentration using an isotropic model that relates the means and standard deviations of the concentrations to the distance between the centroid and the urban center. Model diagnostics are available (2). On each day we calculated the population-weighted average by weighting the estimated pollutant concentration at each Census tract by the number of people residing in that tract according to the 2000 U.S. Census.

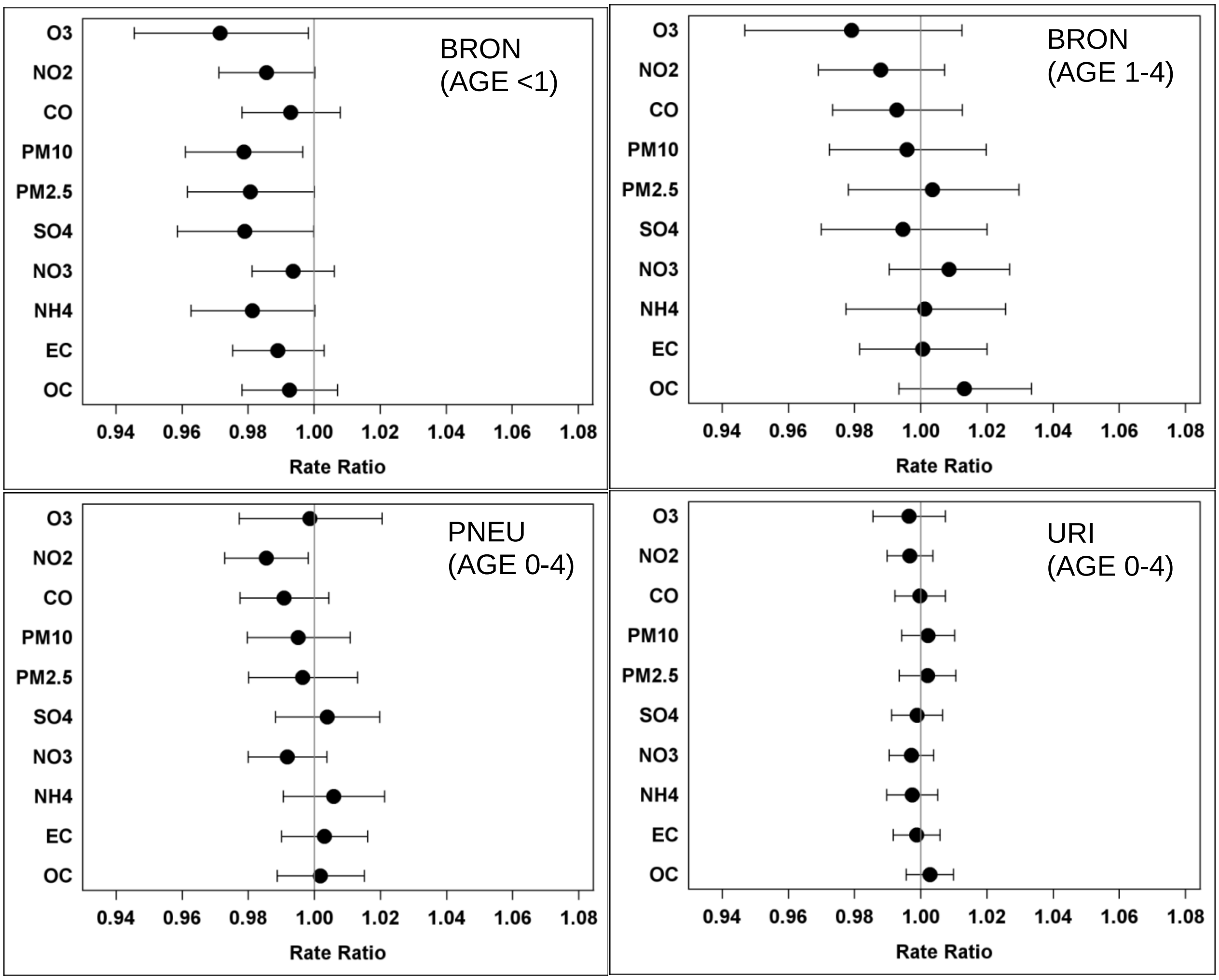
REFERENCES:

1. Peng RD, Dominici F, Louis TA. Model choice in time series studies of air pollution and mortality. *J Roy Stat Soc a Sta*. 2006;169:179-98.
2. Ivy D, Mulholland JA, Russell AG. Development of ambient air quality population-weighted metrics for use in time-series health studies. *J Air Waste Manag Assoc*. 2008;58:711–720.

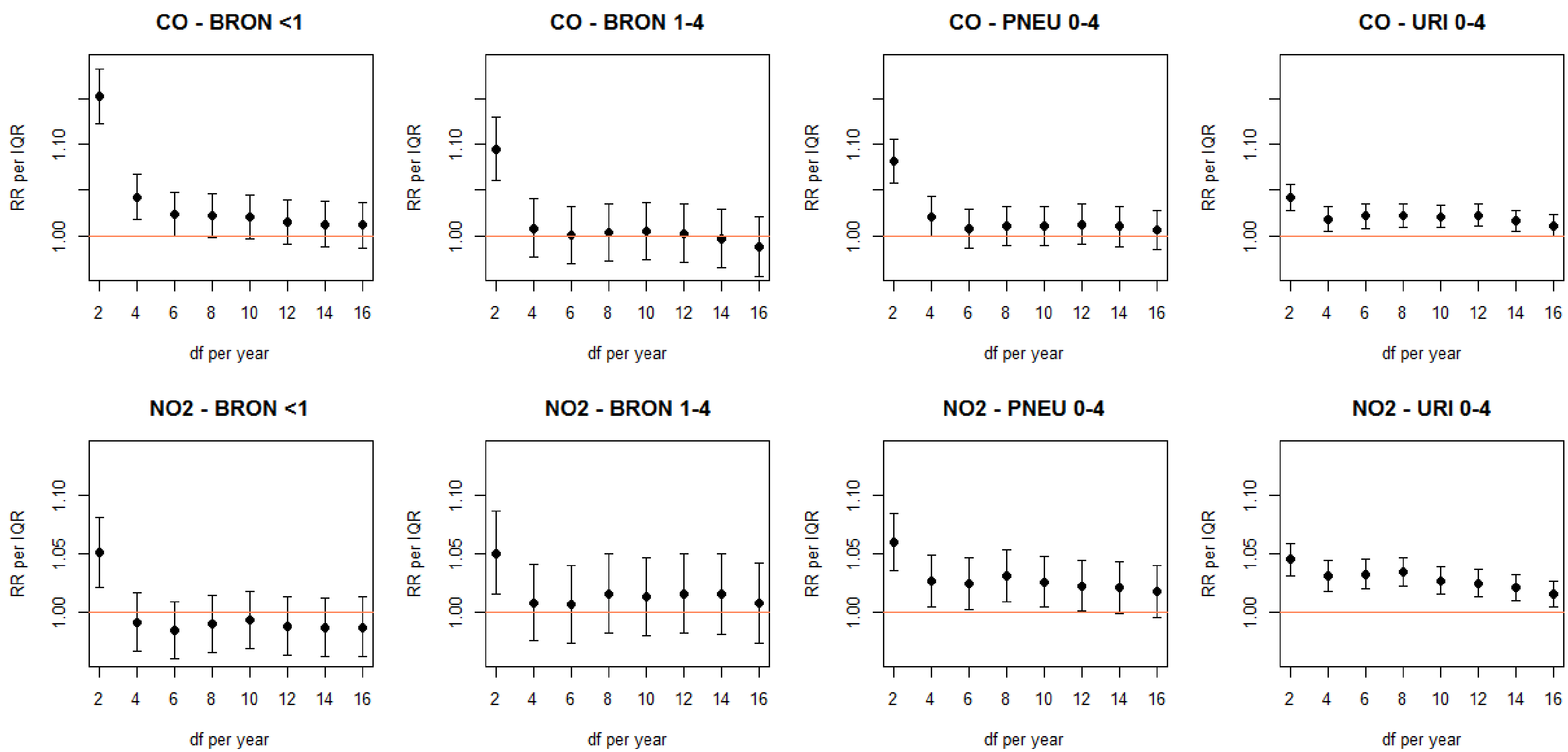
Web Figure 1. Daily observed ED visit counts (+) and predicted counts (line) based on the smooth control for time (cubic splines with 1 knot/month) for 2000 through 2008: bronchitis/bronchiolitis age<1 (A), bronchitis/bronchiolitis age 1-4 (B), pneumonia age 0-4 (C), upper respiratory infection age 0-4 (D). Vertical lines indicate January of each year. The impact of an influenza epidemic in December 2003 on ED visits coded as bronchitis, pneumonia and upper respiratory infection is evident in plots B, C and D.

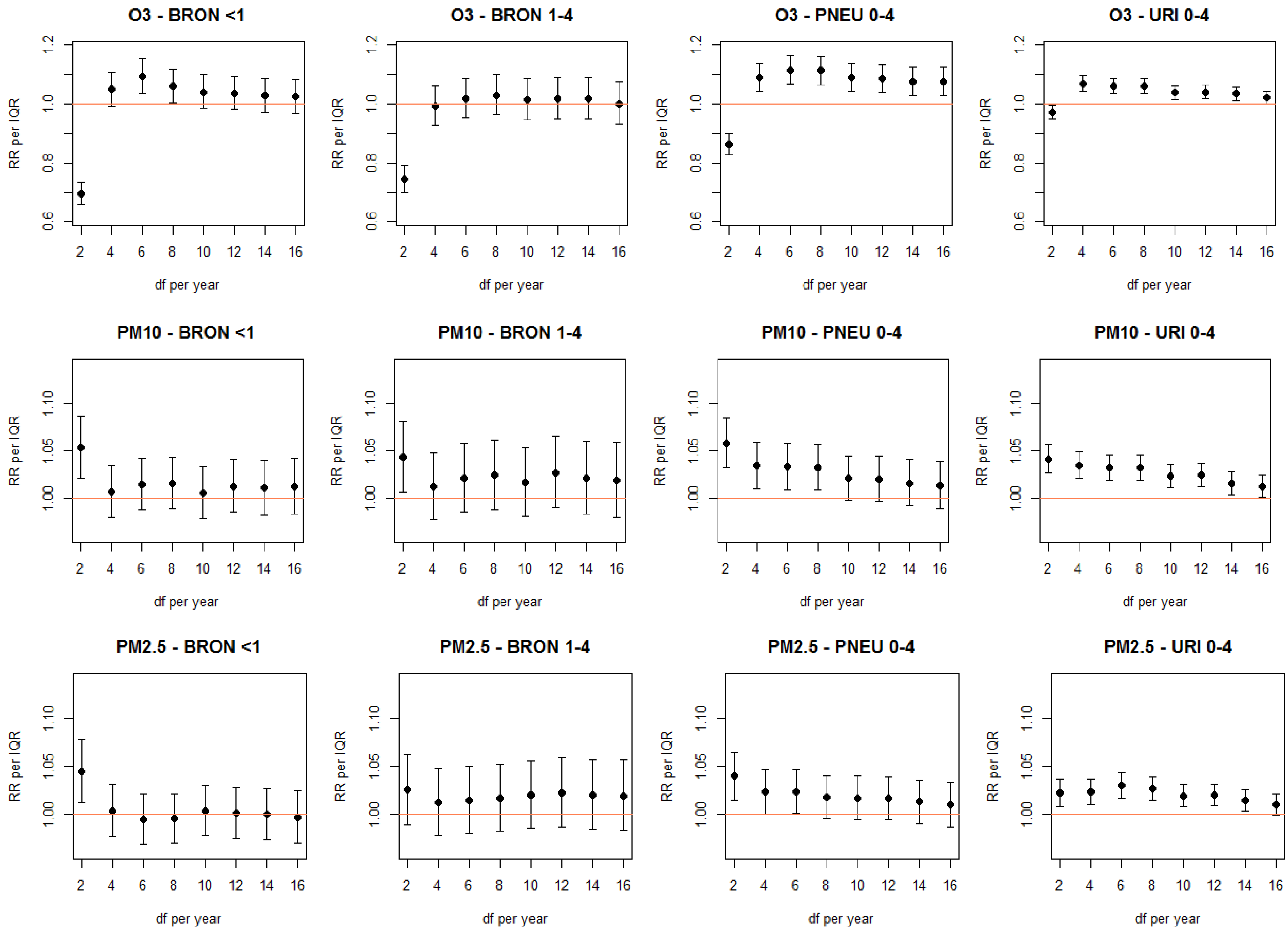


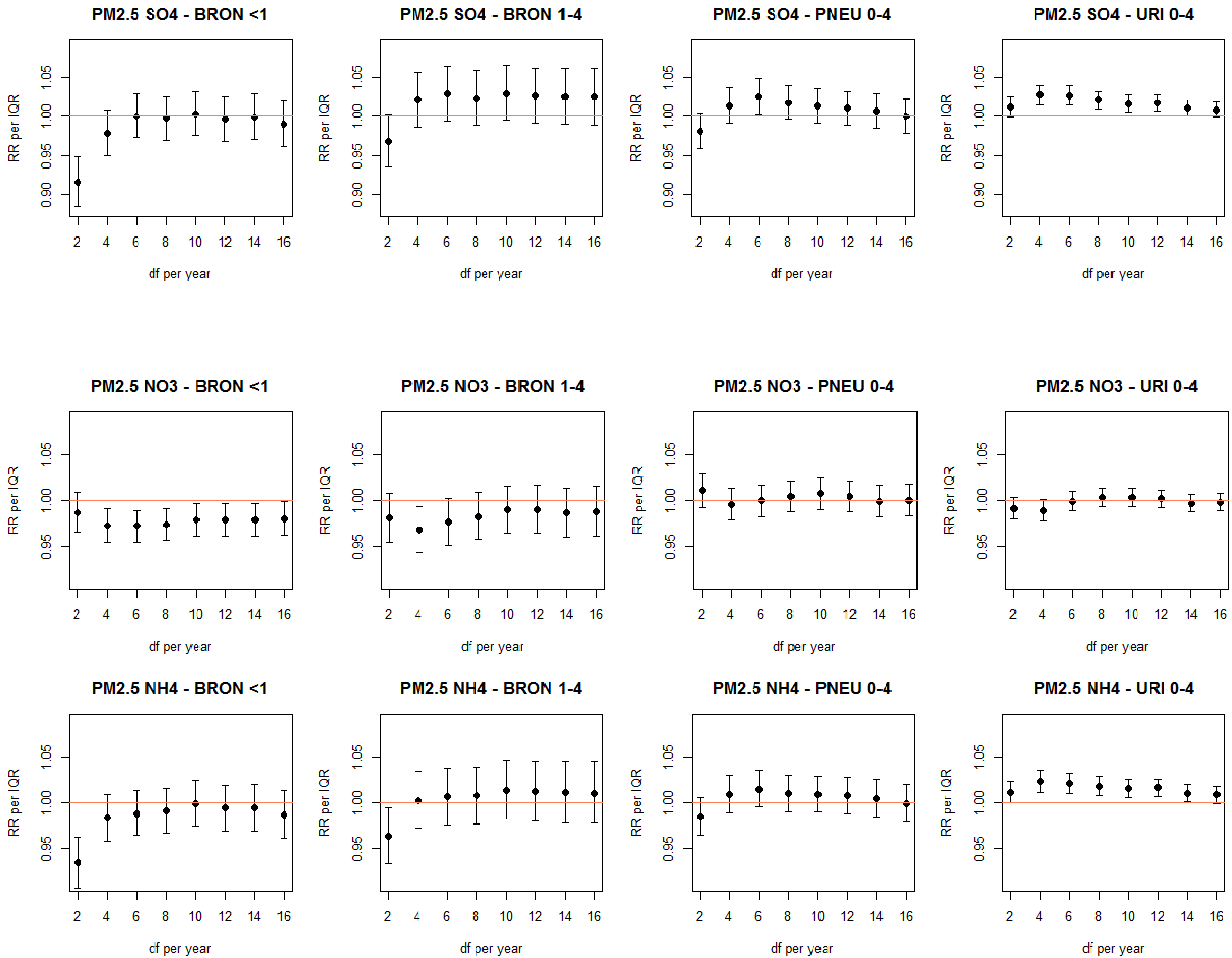
Web Figure 2. Future pollution model assessment: rate ratios and 95% confidence intervals per IQR increase in pollutant concentration for lag negative 1 exposure window (next day pollution). (Abbreviations: BRON <1= bronchiolitis and bronchitis in children <1 year, BRON 1-4= bronchiolitis and bronchitis in children 1-4 years, PNEU=pneumonia, URI= upper respiratory infection, O_3=ozone, NO_2=nitrogen dioxide, CO=carbon monoxide, PM_{10}=particulate matter <10μm in diameter, $PM_{2.5}$= particulate matter <2.5μm in diameter, SO_4= $PM_{2.5}$ sulfate, NO_3= $PM_{2.5}$ nitrate, NH_4= $PM_{2.5}$ ammonium, EC= $PM_{2.5}$ elemental carbon, OC= $PM_{2.5}$ organic carbon.)

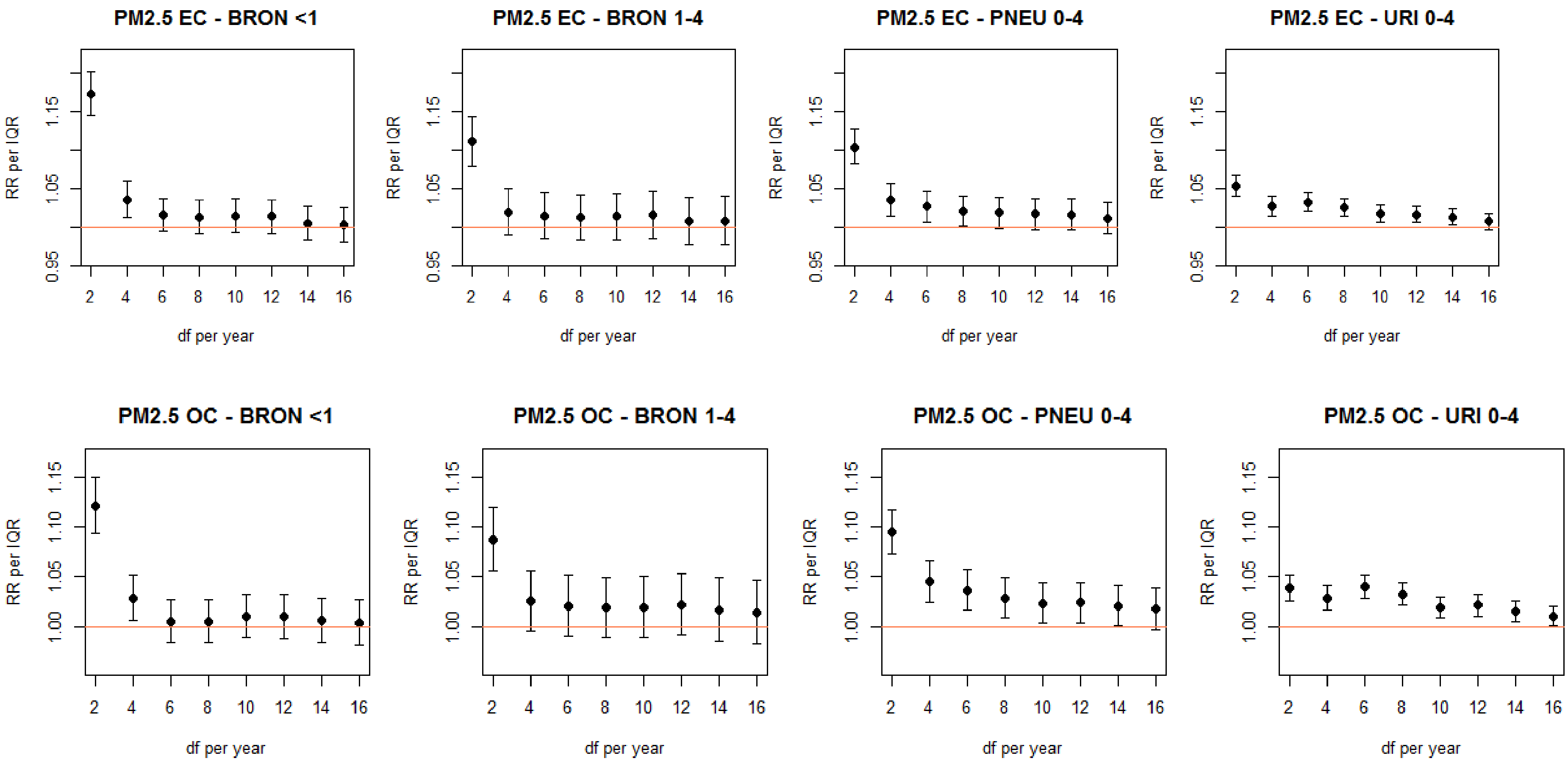


Web Figure 3. Estimated rate ratios per IQR increase in three-day moving average pollutant concentration (lag 0-1-2) for models with different degrees of freedom (knots) per year included in smooth functions of time (cubic splines). Primary analyses included 12 knots/ year. (Abbreviations: BRON <1= bronchiolitis and bronchitis in children <1 year, BRON 1-4= bronchiolitis and bronchitis in children 1-4 years, PNEU=pneumonia, URI= upper respiratory infection, O_3=ozone, NO_2=nitrogen dioxide, CO=carbon monoxide, PM_{10}=particulate matter <10μm in diameter, $PM_{2.5}$= particulate matter <2.5μm in diameter, SO_4= $PM_{2.5}$ sulfate, NO_3= $PM_{2.5}$ nitrate, NH_4= $PM_{2.5}$ ammonium, EC= $PM_{2.5}$ elemental carbon, OC= $PM_{2.5}$ organic carbon.)

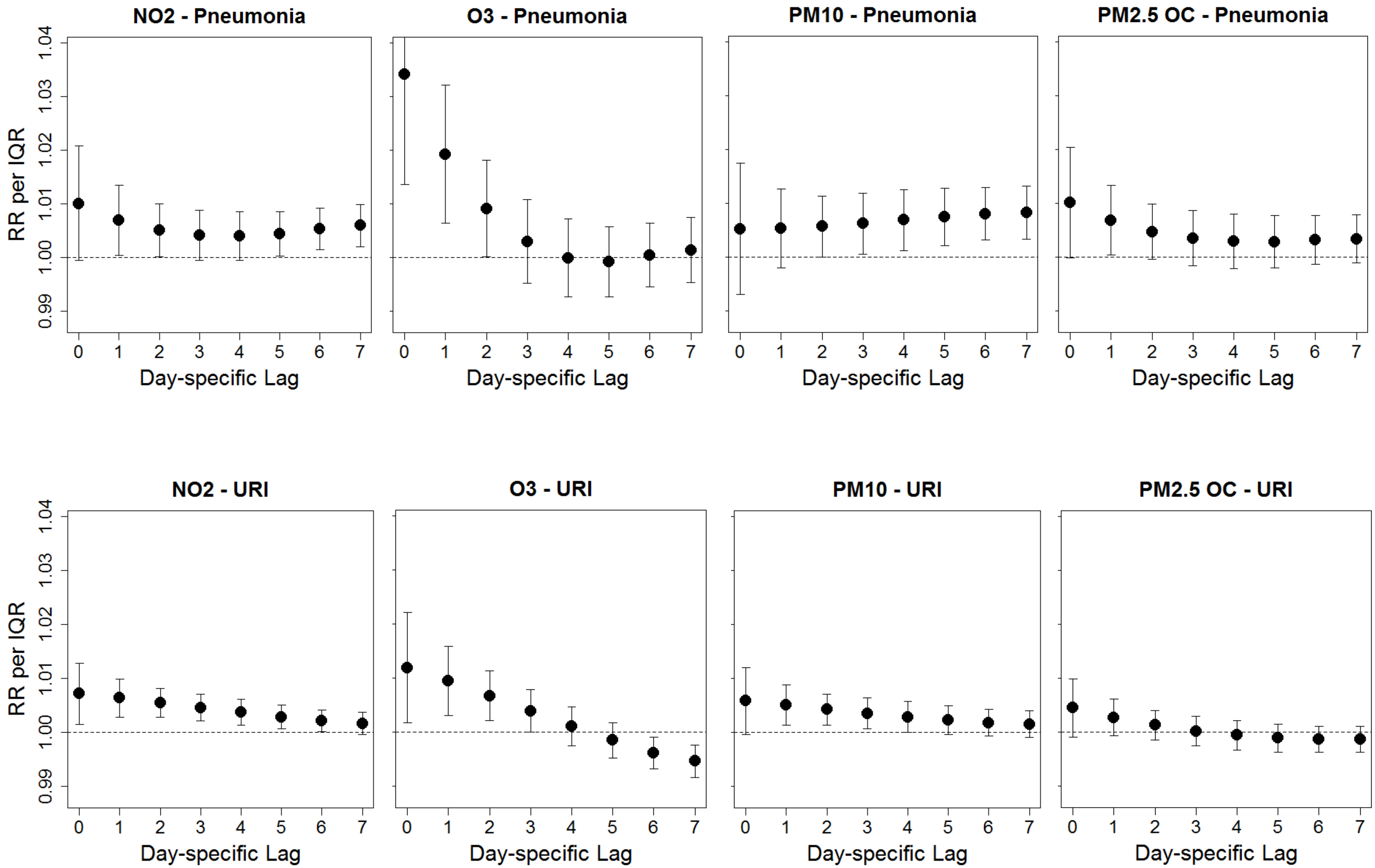



O3 - BRON <1
RR per IQR
0.6 0.8 1.0 1.2
2 4 6 8 10 12 14 16
df per year
O3 - BRON 1-4
RR per IQR
0.6 0.8 1.0 1.2
2 4 6 8 10 12 14 16
df per year
O3 - PNEU 0-4
RR per IQR
0.6 0.8 1.0 1.2
2 4 6 8 10 12 14 16
df per year
O3 - URI 0-4
RR per IQR
0.6 0.8 1.0 1.2
2 4 6 8 10 12 14 16
df per year
PM10 - BRON <1
RR per IQR
1.00 1.05 1.10
2 4 6 8 10 12 14 16
df per year
PM10 - BRON 1-4
RR per IQR
1.00 1.05 1.10
2 4 6 8 10 12 14 16
df per year
PM10 - PNEU 0-4
RR per IQR
1.00 1.05 1.10
2 4 6 8 10 12 14 16
df per year
PM10 - URI 0-4
RR per IQR
1.00 1.05 1.10
2 4 6 8 10 12 14 16
df per year
PM2.5 - BRON <1
RR per IQR
1.00 1.05 1.10
2 4 6 8 10 12 14 16
df per year
PM2.5 - BRON 1-4
RR per IQR
1.00 1.05 1.10
2 4 6 8 10 12 14 16
df per year
PM2.5 - PNEU 0-4
RR per IQR
1.00 1.05 1.10
2 4 6 8 10 12 14 16
df per year
PM2.5 - URI 0-4
RR per IQR
1.00 1.05 1.10
2 4 6 8 10 12 14 16
df per year


PM2.5 SO4 - BRON <1
PM2.5 SO4 - BRON 1-4
PM2.5 SO4 - PNEU 0-4
PM2.5 SO4 - URI 0-4
PM2.5 NO3 - BRON <1
PM2.5 NO3 - BRON 1-4
PM2.5 NO3 - PNEU 0-4
PM2.5 NO3 - URI 0-4
PM2.5 NH4 - BRON <1
PM2.5 NH4 - BRON 1-4
PM2.5 NH4 - PNEU 0-4
PM2.5 NH4 - URI 0-4
RR per IQR
df per year


PM2.5 EC - BRON <1
PM2.5 EC - BRON 1-4
PM2.5 EC - PNEU 0-4
PM2.5 EC - URI 0-4
PM2.5 OC - BRON <1
PM2.5 OC - BRON 1-4
PM2.5 OC - PNEU 0-4
PM2.5 OC - URI 0-4
RR per IQR
0.95
1.00
1.05
1.10
1.15
2
4
6
8
10
12
14
16
df per year

Web Figure 4. Constrained cubic polynomial distributed lag model results. Rate ratios are per IQR increase in pollutant concentration on individual lag days 0-7 controlling for pollutant concentrations for lag days 0-13. To enhance the stability of the lag estimates a cubic polynomial was fit to lags 0–13. (Abbreviations: PNEU=pneumonia, URI= upper respiratory infection, O_3=ozone, NO_2=nitrogen dioxide, PM_{10}=particulate matter <10μm in diameter, $PM_{2.5}$ OC= $PM_{2.5}$ organic carbon.)



Web Figure 5. Loess concentration-response estimates (solid line) and twice-standard error estimates (dashed lines) from generalized additive models (span=0.6) for associations between three-day moving average air pollutant concentrations and emergency department visits for pneumonia age 0-4 and upper respiratory infection age 0-4, Atlanta 1993-2010. The reference (denominator) for the rate ratio is the estimated rate at the 5th percentile of the pollutant concentration. Estimates are presented for the 5th percentile through the 95th percentile of pollutant concentrations due to instability in the concentration-response estimates at the distribution tails.