Journal of Exposure Science and Environmental Epidemiology (2006) 16, 300–310
© 2006 Nature Publishing Group All rights reserved 1053-4245/06/$30.00
www.nature.com/jes

PM source apportionment and health effects: 2. An investigation of intermethod variability in associations between source-apportioned fine particle mass and daily mortality in Washington, DC

KAZUHIKO ITO,[a] WILLIAM F. CHRISTENSEN,[b] DELBERT J. EATOUGH,[c] RONALD C. HENRY,[d] EUGENE KIM,[e] FRANCINE LADEN,[f] RAMONA LALL,[a] TIMOTHY V. LARSON,[g] LUCAS NEAS,[h] PHILIP K. HOPKE[e] AND GEORGE D. THURSTON[a]

[a]*Institute of Environmental Medicine, New York University, Tuxedo Park, NY, USA*
[b]*Department of Statistics, Brigham Young University, Provo, UT, USA*
[c]*Department of Chemistry and Biochemistry, Brigham Young University, Provo, UT, USA*
[d]*Department of Civil and Environmental Engineering, University of Southern California, Los Angeles, CA, USA*
[e]*Center for Air Resources Engineering and Science, Clarkson University, Potsdam, NY, USA*
[f]*Department of Environmental Health, Harvard School of Public Health, Boston, MA, USA*
[g]*Department of Civil and Environmental Engineering, University of Washington, Seattle, WA, USA*
[h]*National Health & Environmental Effects Research Laboratory, U.S. EPA, Chapel Hill, NC, USA*

Source apportionment may be useful in epidemiological investigation of PM health effects, but variations and options in these methods leave uncertainties. An EPA-sponsored workshop investigated source apportionment and health effects analyses by examining the associations between daily mortality and the investigators' estimated source-apportioned $PM_{2.5}$ for Washington, DC for 1988–1997. A Poisson Generalized Linear Model (GLM) was used to estimate source-specific relative risks at lags 0–4 days for total non-accidental, cardiovascular, and cardiorespiratory mortality adjusting for weather, seasonal/temporal trends, and day-of-week. Source-related effect estimates and their lagged association patterns were similar across investigators/methods. The varying lag structure of associations across source types, combined with the Wednesday/Saturday sampling frequency made it difficult to compare the source-specific effect sizes in a simple manner. The largest (and most significant) percent excess deaths per 5–95th percentile increment of apportioned $PM_{2.5}$ for total mortality was for secondary sulfate (variance-weighted mean percent excess mortality = 6.7% (95% CI: 1.7, 11.7)), but with a peculiar lag structure (lag 3 day). Primary coal-related $PM_{2.5}$ (only three teams) was similarly significantly associated with total mortality with the same 3-day lag as sulfate. Risk estimates for traffic-related $PM_{2.5}$, while significant in some cases, were more variable. Soil-related PM showed smaller effect size estimates, but they were more consistently positive at multiple lags. The cardiovascular and cardiorespiratory mortality associations were generally similar to those for total mortality. Alternative weather models generally gave similar patterns, but sometimes affected the lag structure (e.g., for sulfate). Overall, the variations in relative risks across investigators/methods were found to be much smaller than those across estimated source types or across lag days for these data. This consistency suggests the robustness of the source apportionment in health effects analyses, but remaining issues, including accuracy of source apportionment and source-specific sensitivity to weather models, need to be investigated.
Journal of Exposure Science and Environmental Epidemiology (2006) **16,** 300–310. doi:10.1038/sj.jea.7500464; published online 23 November 2005

Keywords: *air pollution, source apportionment, health effects, mortality.*

Introduction

Numerous studies have reported short-term associations between ambient PM concentrations and mortality/morbidity (e.g., see U.S. EPA, 1996). PM was often implicated as the most significant predictor of the health outcomes among the air pollutants in these studies. However, PM is a chemically non-specific pollutant, and may originate or be derived from various emission source types. Thus, its toxicity may well vary depending on its source and chemical composition. If the PM toxicity could be determined based on source types, the regulation of PM may be implemented more effectively. One natural progression of the PM health effects research is therefore to conduct a source apportionment of PM using chemical speciation data, and to examine the associations between source-apportioned PM and health outcomes, rather than with PM mass overall. There have been only a few studies that conducted such analyses (Özkaynak and Thurston, 1987; Özkaynak et al., 1996; Laden et al., 2000; Tsai et al., 2000; (reanalyzed by Schwartz, 2003); Mar et al., 2000 (reanalyzed by Mar et al., 2003)). These studies have provided some suggestive evidence that PM from certain combustion sources (i.e., secondary

1. Address all correspondence to: Professor P.K. Hopke, Center for Air Resources Engineering and Science, Clarkson University, Potsdam, NY 13699-5708, USA. Tel.: +1 315 268 3861. Fax: +1 315 268 4410.
E-mail: hopkepk@clarkson.edu
Received 20 April 2005; accepted 10 October 2005; published online 23 November 2005

aerosols and traffic), but not other sources (e.g., soil), were associated with daily mortality. However, the results are far from conclusive, and more analyses of speciation data using cities with larger populations are needed.

While source apportionment may be a potentially powerful tool for source-oriented evaluation of PM health effects, there are several approaches to conduct source apportionment, and their associated model uncertainties have not been systematically examined. Even for a given source-apportionment approach, there are several options that an investigator may take: (1) which elements to include; (2) the number of factors to be specified; (3) the extent of rotation; and (4) criteria with which the "best" model is chosen. Furthermore, while the advantage of the multivariate factor analysis-based receptor modeling is its ability to "find" sources without prior source profiles, the resulting factors still need to be "named" (i.e., identified) based on the external (and often prior) knowledge regarding the sources. Thus, it is possible that, given the same factor solutions, different researchers may name the factors differently. Despite these options, which can lead to different results, the results from several recent studies that compared two source-apportionment techniques applied to the same sets of data (Huang et al., 1999; Poirot et al., 2001; Qin et al., 2002; Ramadan et al., 2003) suggest that the source-apportioned results between various two sets of methods were not remarkably different. However, this may be in part due to the fact that the same investigators were applying two different methods (except in the Poirot et al. study in which two investigators independently analyzed the same data), possibly minimizing the choice of options taken in each method. These issues motivated several members of the EPA PM Centers to organize a workshop on source apportionment in which the same data sets would be analyzed by at least several teams of investigators so that source apportionment results and their estimated health impacts would be compared (for details, see Thurston et al., 2005). This paper will present the results from the mortality analysis of Washington, DC data. In addition to the results presented at the workshop, we present results of additional sensitivity analyses that we conducted after the workshop. A more detailed comparison of source apportionment results are presented in Hopke et al. (2005).

Materials and methods

Data

The $PM_{2.5}$ speciation data for Washington, DC for the study period 08/31/1988–12/31/1997 were downloaded from http://vista.cira.colostate.edu/improve. Detailed descriptions of the IMPROVE data can be found on the web site, but briefly, the database contains $PM_{2.5}$ mass, 24 trace elements by energy dispersive X-ray fluorescence (EDXRF), anions (sulfate, nitrate) and cations (particulate ammonium) by ion chromatograph, and organic and elemental carbon (four and three levels of fractions from thermal optical analysis, respectively). The database also included uncertainty for each observation. The IMPROVE data were collected on Wednesdays and Saturdays for the study period. In addition to the IMPROVE data, we retrieved, processed, and distributed other air pollution data from EPA's Aerometric Information Retrieval System (AIRS, now called Air Quality System, or AQS) for the same period. Average values of multiple monitors' data (three or four monitors) for PM_{10} (available every 6th-day 24-h samples) and daily gaseous pollutants (24-h average of hourly values), O_3, SO_2, NO_2, and CO, were computed. Several weather variables from the Dulles airport were also distributed in this data set, as extracted from EarthInfo (Boulder, CO, USA) compact discs, including: daily mean temperature, dew point, relative humidity, precipitation, pressure, resultant wind speed and direction. The IMPROVE, AQS, and weather data were sent on a compact disc to 11 investigators in December 2002. To allow a consistent intercomparison of results across investigators, participants were requested to submit results in a standardized format and with a list of items to describe the details of source apportionment analysis (e.g., type and extent of rotation, treatment of outliers, criteria used to include species in the analysis, etc.).

The source apportionment methods used in these analyses include absolute principal components analysis (APCA; Thurston and Spengler, 1985), positive matrix factorization (PMF; Paatero and Tapper, 1993), Unmix (Henry and Norris, 2002), target-transformation (or specific-rotation) factor analysis (TTFA; Koutrakis and Spengler, 1987), and confirmatory factor analysis (Christensen et al., 2005). Description of these methods are found elsewhere (Hopke et al., 2005). We note that none of the investigators applied source profiles-based approach such as the chemical mass balance (CMB) method. Thus, the variation of the source apportionment results and corresponding variation in mortality risk estimates in this analysis is limited to this particular type (i.e., multivariate factor analysis based) of source apportionment methods.

These multivariate factoranalysis-based models yield a set of factors. The investigator then gives each of these factors a label that is indicative of a source or source type (e.g., "soil") based on the investigator's prior knowledge on "signature species". Thus, these factors could be labeled subjectively. Nevertheless, several commonly labeled factors were found in the results submitted by the investigators. At the workshop, we ascertained the assignment of each investigator's factors to commonly named factors. Thus, in the rest of the report, when we mention, for example, "soil factor", we mean the factor that was labeled as soil (or something equivalent) by the investigators. These factors were submitted in the form of estimated daily source contributions (e.g., via regression of

$PM_{2.5}$ on the factors). We call these estimated daily $PM_{2.5}$ source contribution as "source-apportioned $PM_{2.5}$" in the rest of this report.

We obtained nine sets of source-apportioned $PM_{2.5}$ data that could be analyzed for mortality associations. The Clarkson University team submitted two sets, and the New York University team submitted three sets of solutions using different methods. Thus, these nine sets are not from nine independent investigators, and therefore, we call this grouping as "investigators/methods".

Death records were extracted from the National Center for Health Statistics database for the period 8/31/88–12/31/97, and daily counts were aggregated for District of Columbia and the surrounding six counties: Montgomery Co., MD, USA; Prince George's Co., MD, USA; Fairfax Co., VA, USA; Alexandria city, VA, USA; Fairfax city, VA, USA; and, Falls Church city, VA, USA. Three categories of deaths were analyzed: (1) total non-accidental; (2) cardiovascular; and (3) cardiovascular plus respiratory. The total population of the area included was approximately 2.4 million. The IMPROVE-operated monitoring site was located in Washington, DC (longitude −77.0343; latitude 38.8761) near the Potomac River, approximately in the center of the geographic area covered (a 30-mile radius from the monitor contains essentially all of the study population).

Statistical Analysis

The nine source-apportioned $PM_{2.5}$ datasets were each used in the mortality analysis conducted at NYU. One source-apportioned $PM_{2.5}$ time series was included in the model at a time. We developed our "base" mortality model as a function of season and other temporal trends, day-of-week (note the data collections on Wednesdays and Saturdays only) and weather variables in Poisson Generalized Linear Models (McCullagh and Nelder, 1989). First, we fit a smooth function of time, using natural splines, to mortality to adjust for seasonal trend and unmeasured seasonal confounders including influenza epidemics. In addition to this epidemiological reasoning, the inclusion of a smooth function of time also has statistical benefit in that it removes or reduces residual autocorrelation and overdispersion in the mortality regression. Thus, the choice of the degrees of freedom for smoothing of time was based on the inspection of the fitted mortality series (to see if it captured broad peak influenza epidemics that vary from year to year) and based on the extent of autocorrelation of the residuals. We used natural splines with 38 degrees of freedom (approximately 4 degrees of freedom per year). To examine the sensitivity of estimated $PM_{2.5}$ mortality risks, we also ran the model using 2, 8, and 16 degrees of freedom per year. We then considered weather terms to be added to this model along with the day-of-week variable. The weather model specifications were in part based on the pattern observed in crosscorrelation function results, and in part based on the literature. We chose a relatively parsimonious weather model that included: (1) natural splines of the same-day temperature with 4 degrees of freedom to fit immediate temperature effects; (2) natural splines of the average of lag 1–3 days lagged temperature with 4 degrees of freedom to fit delayed temperature effects; and (3) an indicator for "hot" (daily mean temperature above 80 degrees) and "humid" (daily relative humidity above 70%) days to fit the interaction. Thus, our base model included these weather terms, a temporal trend term (described above), and a day-of-week indicator (1 degree of freedom).

To the base model, we added $PM_{2.5}$ or estimates of source-apportioned $PM_{2.5}$ from each of the investigators/methods. Lags between 0 and 4 days were examined. Note that, because the $PM_{2.5}$ speciation data at the Washington, DC air monitoring site were collected on Wednesdays and Saturdays, each lag corresponds to different sets of mortality days (e.g., 0-day lag effect: Wednesday and Saturday mortality only; 1-day lag effect: Thursday and Sunday mortality only, etc.). As these lags could induce complications in the lag structure of PM–mortality associations, for $PM_{2.5}$, we also ran separate regression models using Wednesdays mortality only and Saturdays mortality only.

Relative risks for the mortality series were computed for two types of mass increment: (1) per 5–95th percentile increment of source-apportioned $PM_{2.5}$ and (2) per 10 $\mu g/m^3$ increment of source-apportioned $PM_{2.5}$. The former would be useful for evaluating a relative risk increase for "low" vs. "high" pollution, while the latter may be useful to evaluate the relative toxicity of PM from different source types per equal mass basis. The former also "adjusts" for potential bias of a given monitor that may consistently measure higher or lower levels of pollution compared to the average levels across the metropolitan area due to the influence of local sources. Thus, we mainly focus on our results using the 5–95th percentile increment.

We further summarized the results across investigators/methods in two ways. First, we computed variance-weighted average risk estimates for each lag and estimated source type across investigators/methods. Second, we attempt to explain the variation in the estimated risks as a function of lag, estimated source type, and investigators/methods. This was done by regressing the percent excess deaths as the dependent variable and indicator variables for estimated source types (8 degrees of freedom), investigators/methods (8 degrees of freedom), and lags (4 degrees of freedom) in a general linear model, yielding an analysis of variance (ANOVA) table.

Owing to the concern about the rather peculiar lag structure of association between $PM_{2.5}$ and mortality (lag 3 day association) found in this dataset, we also conducted, after the workshop, a sensitivity analysis using five alternative weather models (models a–e) to examine if weather models affected the lag structure of associations. The first two models were similar to those used in recent multi-city time-series

studies: (a) four smoothing terms including natural splines of same-day temperature (df = 6), natural splines of the average of lag 1–3 day temperature (df = 6), natural splines of same-day dewpoint (df = 3), natural splines of the average of lag 1–3 day dewpoint (df = 3); and (b) two smoothing terms including one with natural splines of same-day temperature (df = 3) and another with natural splines of same-day dewpoint (df = 3). Model (a) is similar to that used in the mortality analyses of the National Morbidity and Mortality Air Pollution Study, or NMMAPS (Samet et al., 2000; Dominici et al., 2003). Model (b) is similar to that used in the Harvard Six Cities time-series analyses (Schwartz et al., 1996; Schwartz, 2003; Klemm et al., 2000; Klemm and Mason, 2003). In addition, to compare results with those from the other workshop data set, Phoenix, AZ, the model used by Mar et al. (2005) was used. This model, model (c), included a natural spline term of 1-day lag temperature with 5 degrees of freedom and a natural spline term of the same-day relative humidity with two degrees of freedom. We also considered model (d) with interaction terms of temperature and relative humidity at lag 0 and lag 2, in natural splines, both with 4 degrees of freedom. Finally, a model without any weather adjustment, model (e), was also examined.

Results

Descriptive Statistics

Detailed descriptions of the source apportionment results for this city can be found in a separate companion paper by Hopke et al. (2005). This paper will focus on the comparison of the mortality analysis results and will present the source apportionment results only to the extent that they will help to interpret the mortality analyses. Table 1 shows the means, standard deviations, and 5–95th percentiles for the source-apportioned $PM_{2.5}$ mass concentrations by source types identified and by investigators/methods. The mean, standard deviation, and 5–95th percentile for the observed total $PM_{2.5}$ were 17.8, 8.7, 28.7 $\mu g/m^3$, respectively. Four sources/pollution types, soil, traffic, secondary sulfate, and nitrate, were most commonly identified, and together explained more than 80% of the $PM_{2.5}$ on the average. The estimated secondary sulfate associated $PM_{2.5}$, identified in all investi-

Table 1. Estimated average $PM_{2.5}$ source contributions, (SD), and [5–95th increment] in $\mu g/m^3$.

Group/name analysis[a]	ID[b]	Number of factors[c]	Soil	Traffic (gasoline and/or diesel)[d]	Secondary sulfate	Nitrate	Residual oil	Wood smoke	Sea salt	Incinerator	Primary coal
BYU/DE UNMIX	A	6	0.8 (0.6) [1.3]	2.0 (1.6) [5.4]	8.4 (7.3) [21.4]	6.7 (5.6) [17.9]		0.2 (0.5) [0.6]		0.3 (0.4) [0.7]	
BYU/WC Confirmatory FA	B	8	0.5 (1.0) [2.7]	3.7 (2.5) [8.0]	5.1 (4.3) [10.4]	6.7 (3.5) [10.4]		1.9 (1.9) [5.7]	0.9 (0.9) [2.7]		
CU/EK, PH PMF	C	10	0.3 (0.5) [0.9]	4.1 (3.1) [8.0]	10.6 (7.1) [21.1]	1.6 (1.6) [5.0]	0.3 (0.4) [0.9]		0.4 (0.4) [1.1]	0.7 (0.5) [1.4]	
CU/XHS, PH PMF	D	8	0.5 (0.7) [1.3]	1.6 (1.2) [3.2]	8.4 (6.5) [19.0]	3.5 (3.0) [9.8]	0.6 (0.8) [2.0]		0.6 (0.5) [1.6]	1.0 (0.6) [1.6]	1.7 (1.1) [3.6]
HU/FL, LN TTFA:	E	3	3.7 (2.0) [4.8]	4.6 (2.0) [5.7]	7.6 (4.0) [12.2]						
NYU/GT, RL, KI APCA	F	10	1.0 (1.5) [2.8]	4.2 (2.8) [8.2]	9.6 (7.1) [22.0]		0.6 (1.2) [3.3]	0.2 (0.4) [0.6]	0.9 (1.4) [4.3]	0.6 (0.7) [1.4]	
NYU/GT, RL, KI PMF	G	10	0.5 (0.7) [1.4]	2.5 (1.4) [4.2]	10.7 (6.9) [21.5]	2.5 (2.5) [8.4]	0.3 (0.4) [1.0]	0.3 (1.1) [0.8]	0.2 (0.2) [0.7]		
NYU/GT, RL, KI Multiple regression[e]	H	7	1.1 (1.3) [2.0]	3.5 (1.3) [3.8]	9.8 (6.3) [18.5]	2.1 (1.9) [6.1]	0.4 (0.4) [1.2]	0.6 (0.7) [0.9]			2.1 (1.2) [3.8]
USC/RH Unmix	I	8	0.7 (1.1) [1.8]	4.7 (3.5) [9.7]	7.8 (6.8) [20.2]	2.5 (2.6) [8.8]	0.3 (0.3) [0.9]				1.2 (1.0) [3.2]

[a]Initials of the University or Institute, investigators (see the title page), and the abbreviations of methods (see text) are shown.
[b]These IDs are used in the figures.
[c]Only the factors analyzed in mortality regressions are shown in this table. For complete listing, see the companion paper by Hopke et al. (2005).
[d]Mass contributions for gasoline and diesel were combined when they were both reported.
[e]"Signature element or species" for each source was chosen *a priori*, and $PM_{2.5}$ was regressed on them to apportion mass concentrations.

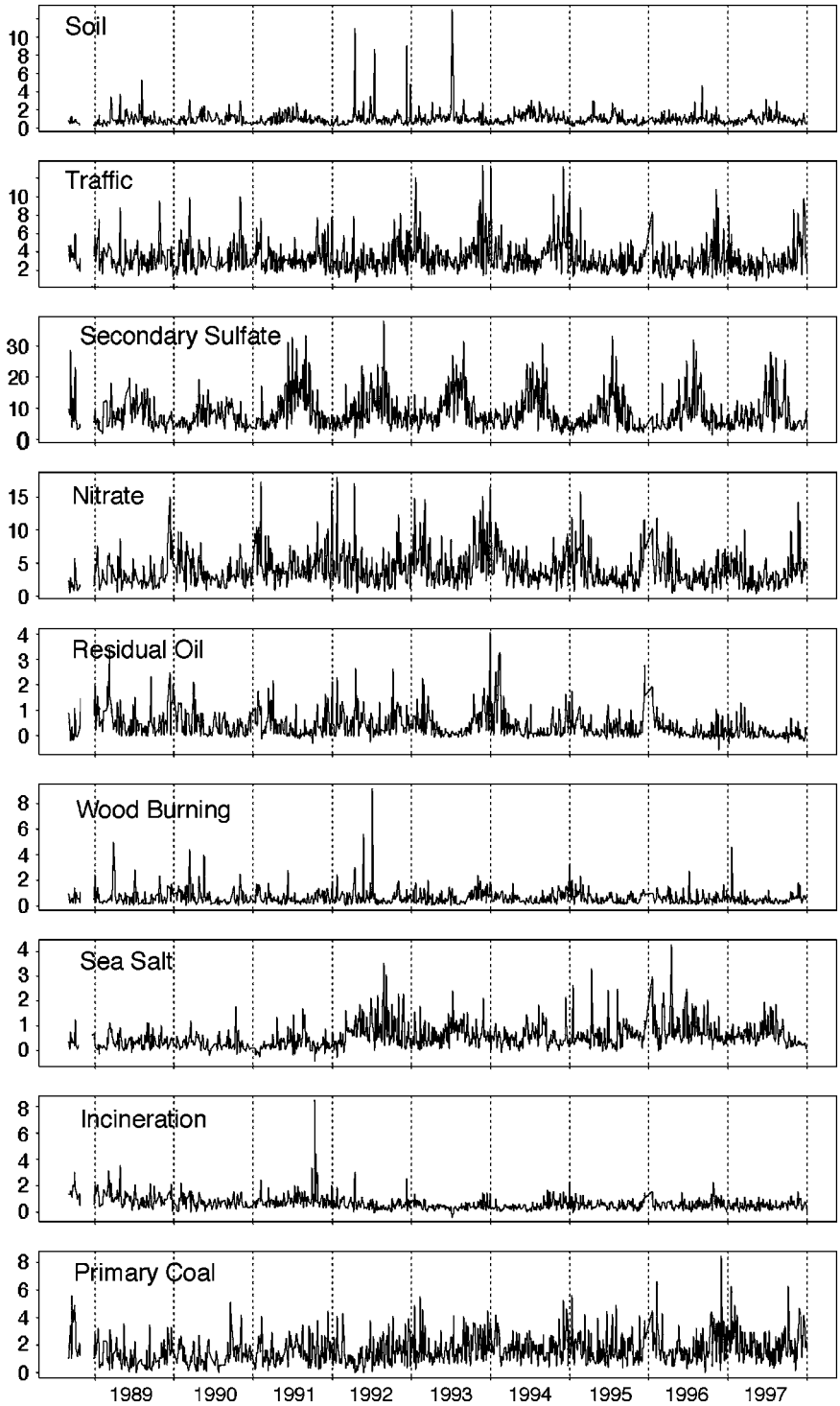

Figure 1. Time-series plot of source-apportioned $PM_{2.5}$ averaged across investigators/methods.

gators/methods' results, was generally the largest estimated constituent of $PM_{2.5}$, explaining up to 60% of the $PM_{2.5}$ mass on average. Soil associated $PM_{2.5}$ was estimated to be much smaller fraction of $PM_{2.5}$ (2–20%), but was identified by all the investigators/methods. Every investigator/method identified traffic-related $PM_{2.5}$, but two teams clearly separated diesel-associated $PM_{2.5}$ and gasoline-associated $PM_{2.5}$, while another team, using two methods, identified two traffic-associated $PM_{2.5}$ without clearly specifying diesel or gasoline. Therefore, to facilitate a comparison of risk

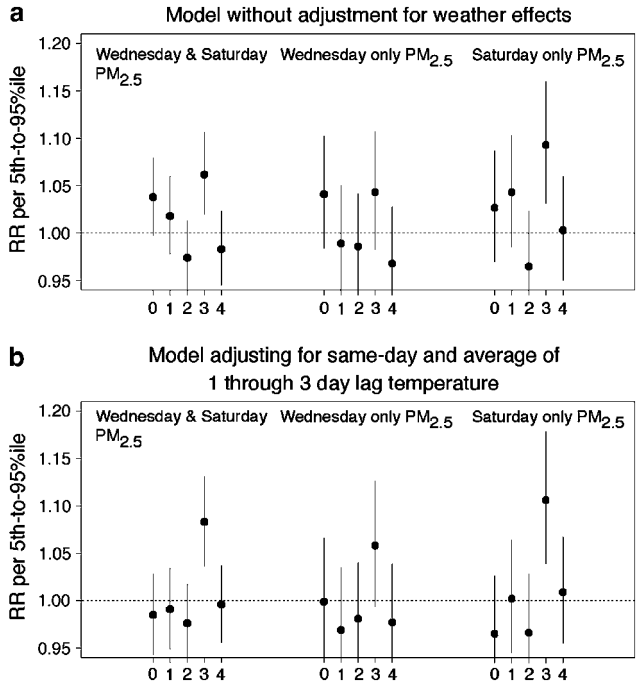

Figure 2. Lag structure of $PM_{2.5}$ associations with total mortality and the influence of weather model.

estimates for the estimated traffic $PM_{2.5}$, we conducted mortality analysis using the combined traffic $PM_{2.5}$ (i.e., diesel plus gasoline). Figure 1 shows the daily source-apportioned $PM_{2.5}$ mass concentrations averaged across investigators/methods. The secondary sulfate tends to be higher in summer, while traffic, nitrate, and residual oil tend to be higher in cold seasons. We also compared the mean values for the weekday (Wednesday) vs. weekend (Saturday) using *t*-test for these source-apportioned $PM_{2.5}$. We found that $PM_{2.5}$ apportioned to traffic, soil, and incinerator to have higher averages (21, 20, and 9%, respectively) on Wednesdays than on Saturdays.

Time-Series Daily Mortality Analyses

Figure 2 shows the estimated relative risks for $PM_{2.5}$ at lags 0–4 using the base model, as well as using the model without adjustment for weather effects. Note that a significant association is seen only at lag 3 day. This is rather a peculiar lag structure in that it is not "distributed" (i.e., positive coefficients at consecutive lags). However, examining associations at real consecutive days is not possible with this data set because of the Wednesday and Saturday sampling schedule. Thus, each lag represents the correlation of the

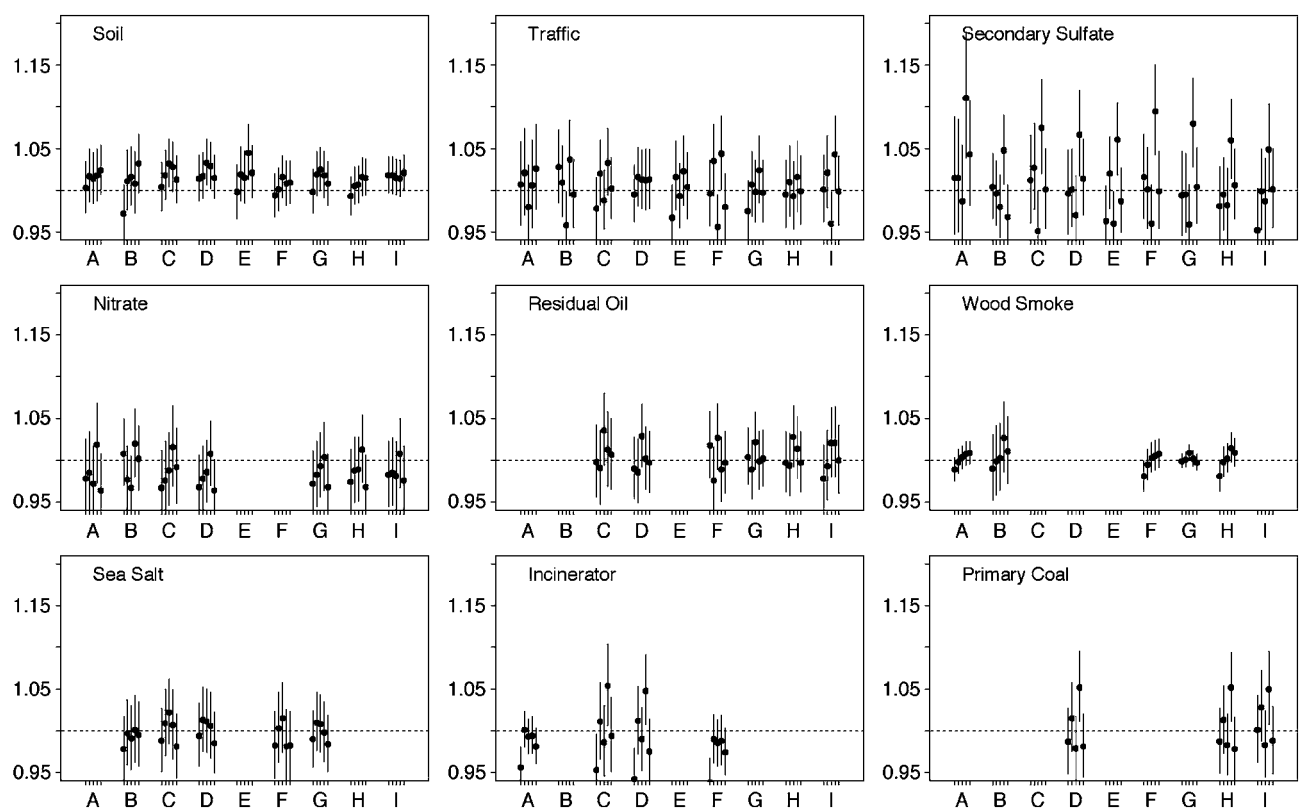

Figure 3. Estimated total non-accidental mortality relative risk per 5–95th percentile increment in source-apportioned $PM_{2.5}$ by source type and investigators/methods. See Table 1 for alphabetical keys. The four consecutive estimates are for lags 0–4 days.

pollution data with an entirely different set of mortality days, which may well add instability to the lag structure. However, as also shown in Figure 2, restricting the data to Wednesday only or Saturday only also resulted in the strongest association at lag 3 day. In the model without adjustment for weather, $PM_{2.5}$ risk estimates remained most significant for lag 3, but were also nearly significant at lag 0 day. Thus, the weather model does affect the lag structure of association and pollution risk estimates somewhat, but it is not inducing the association.

In the sensitivity analysis of $PM_{2.5}$ risk estimates to varying extents of temporal trend adjustment, the lag structure and magnitude of estimates were essentially unaffected. For example, the percent excess risk estimates for total non-accidental mortality at lag 3 were 7.9% (95% CI: 3.3, 12.6), 8.3% (95% CI: 3.7, 13.1), 8.3% (95% CI: 3.7, 13.2), and 8.1% (95% CI: 3.1, 13.2), for 2, 4 (the base model), 8, and 16 degrees of freedom per year, respectively. Therefore, the rest of the analyses using source-apportioned $PM_{2.5}$ were conducted using the base model with 4 degrees of freedom per year.

Figures 3 and 4 show the relative risk estimates per 5–95th percentile of source-apportioned $PM_{2.5}$ by source types across investigators/methods for total (non-accidental) and cardiovascular mortality series, respectively. The results for cardiorespiratory mortality were similar to those for cardiovascular mortality (results not shown). Risk estimates per 10 $\mu g/m^3$ of source-apportioned $PM_{2.5}$ were more variable across investigators/methods as well as across estimated source types, likely due to the differences in mean values across these groupings (results not shown). Interestingly, the lag structure of associations varies from source type to source type, and is generally consistent across the investigators/methods, as follows: the soil factor has mostly positive, although not significant, coefficients at multiple lags; the secondary sulfate factor shows the strongest association at lag 3 day in all the investigators/methods' results; the nitrate factor shows mostly negative coefficients except at lag 3 day; the residual oil factor shows its strongest association, although never significant, at lag 2 days; the wood-burning factor shows generally increasing association with increasing lag time (not significant); the incinerator factor consistently shows significantly negative associations at lag 0 day; and the primary coal factor, identified by three investigators, shows the same lag structure as the secondary sulfate, with a significant association at lag 3 day. In general, the variations in risk estimates across the lags are greater than the variations across investigators/methods.

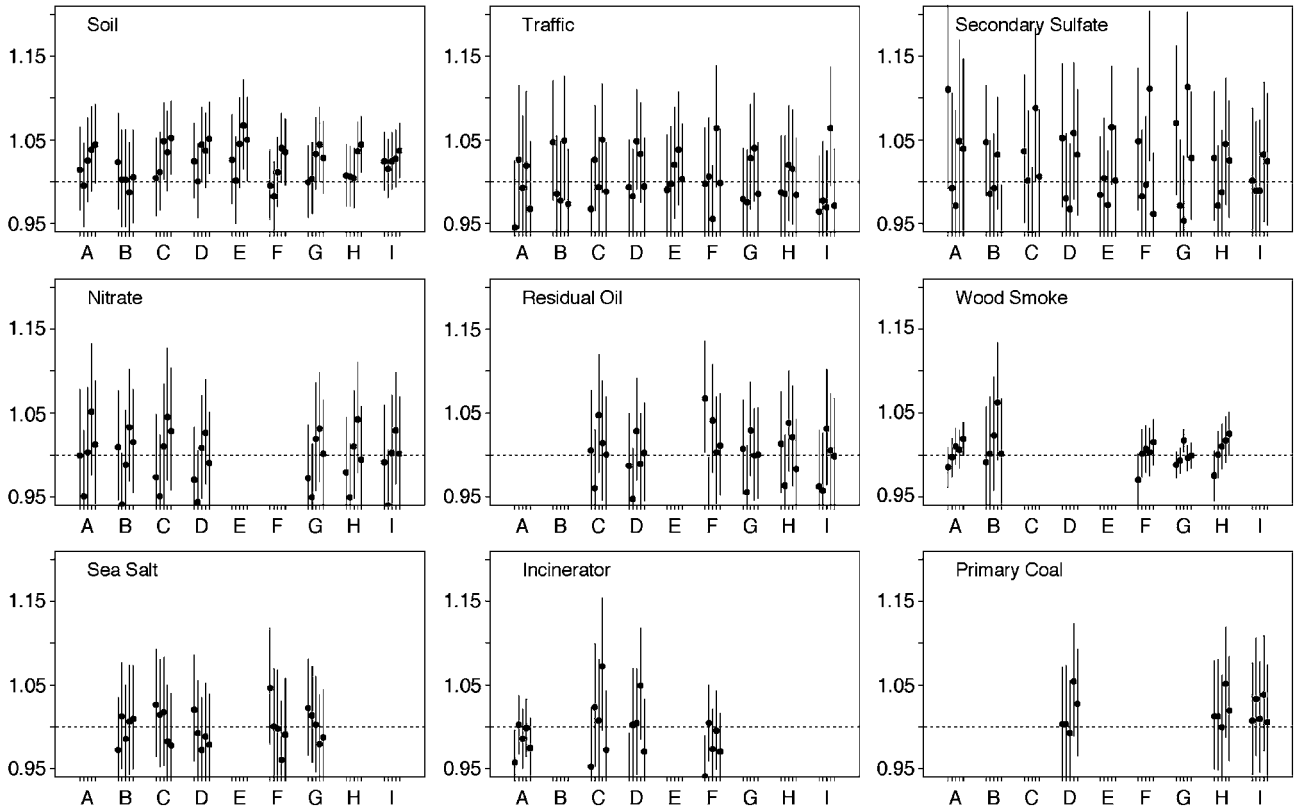

Figure 4. Estimated total cardiovascular mortality relative risk per 5–95th percentile increment in source-apportioned $PM_{2.5}$ by source type and investigators/methods. See Table 1 for alphabetical keys. The four consecutive estimates are for lags 0–4 days.

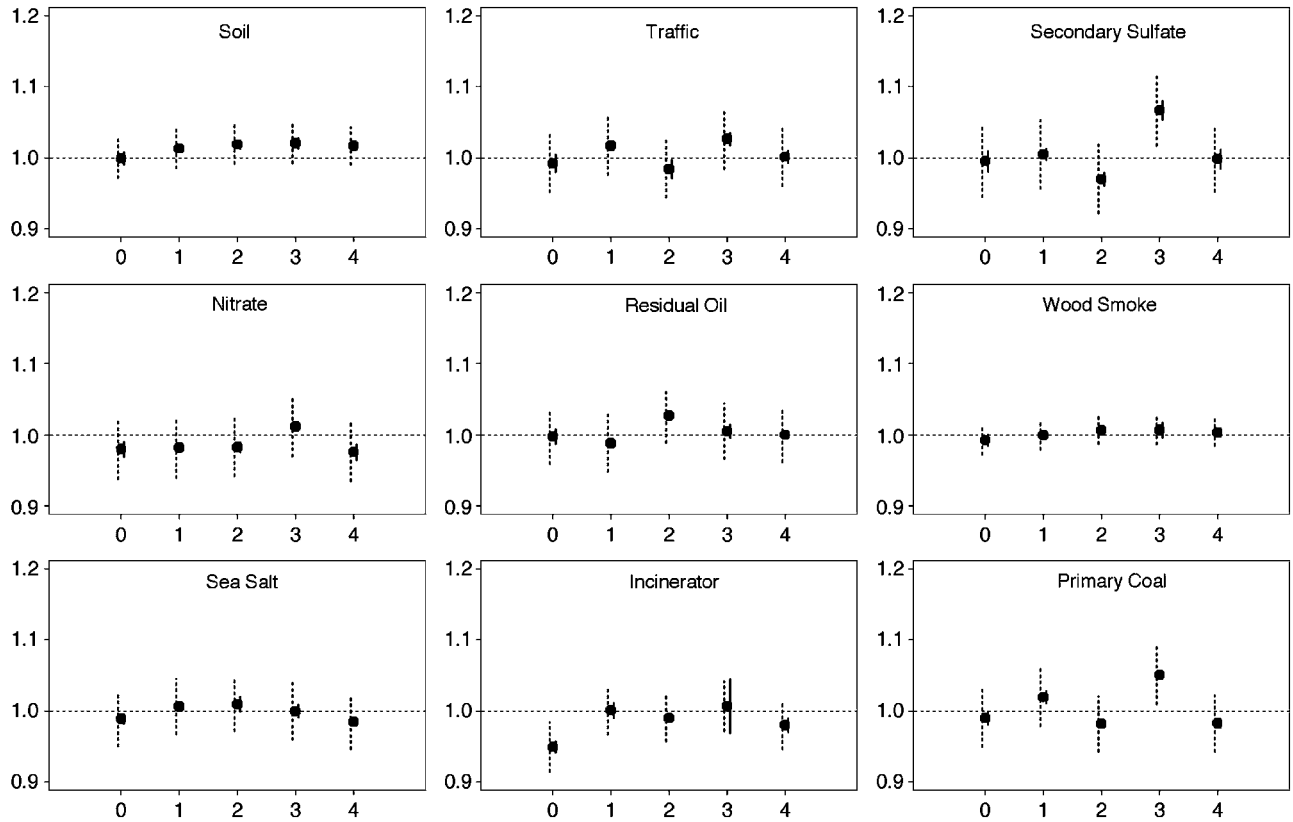

Figure 5. Relative risk and 95% CI of total mortality associated with estimated source-apportioned $PM_{2.5}$ (variance-weighted), averaged across investigators/methods. *Y*-axis: relative risk per 5–95th percentile increment of estimated source-apportioned $PM_{2.5}$. *X*-axis: lag 0–5 days. The dotted 95% CIs are the confidence bands based on the average regression standard errors. The solid 95% CIs are the confidence bands based on the variance of point estimates across investigators/methods.

To further summarize the variability of the risk estimates, we computed variance-weighted average risk estimates at each lag across investigators/methods, as shown in Figure 5 for total mortality. Also shown are the 95% CI from the variability across investigators/methods and the average 95% CIs from regressions (based on the average of standard errors of regression across investigators/methods). The variability of the estimated source-specific risk estimates due to investigators/methods is much smaller than that due to regression standard error with only one exception, the risk estimate for the incinerator factor at lag 3 day. Again, the differences in lag structure of associations among some of the source factors are seen. Figure 6 shows the variance-weighted mean risk estimates result for cardiovascular mortality. The pattern of risk estimates for cardiovascular mortality is similar to that for total mortality.

Table 2 shows the ANOVA result with the variance breakdown of total mortality percent risk estimates by estimated source types, investigators/methods, and lag days, from a general linear model. The pollution source types and lag days significantly explained the variation in risk estimates, whereas the investigators/methods did not. Table 3 shows the ANOVA result for cardiovascular mortality risk estimates, which was similar to that for total mortality. These results suggest that the variation in risk estimates due to investigators/methods was not as important as those due to source types and lag days.

The varying lag structure of associations across source types made it difficult to compare source-specific effect sizes in a simple manner. One approach, while this may bias the estimates upwards, was to compare the risk estimates across source types at the most consistently significant lag for each source type. For total mortality, the largest (and most significant) estimated relative risks (RR) per 5–95th percentile increment of source apportioned PM was found for secondary sulfate (variance-weighted mean percent excess mortality = 6.7% (95% CI: 1.7, 11.7)) at lag 3 days. The second largest estimated risks were found for primary coal-related $PM_{2.5}$, identified by only three teams, with the same lag structure as sulfate (mean percent excess mortality = 5.0% (95% CI: 1.0, 9.1). Residual oil factor showed the most consistent positive estimate at lag 2 days (mean percent excess mortality = 2.7% (95% CI: −1.1, 6.5)). Risk estimates for traffic-related $PM_{2.5}$, while significant in some

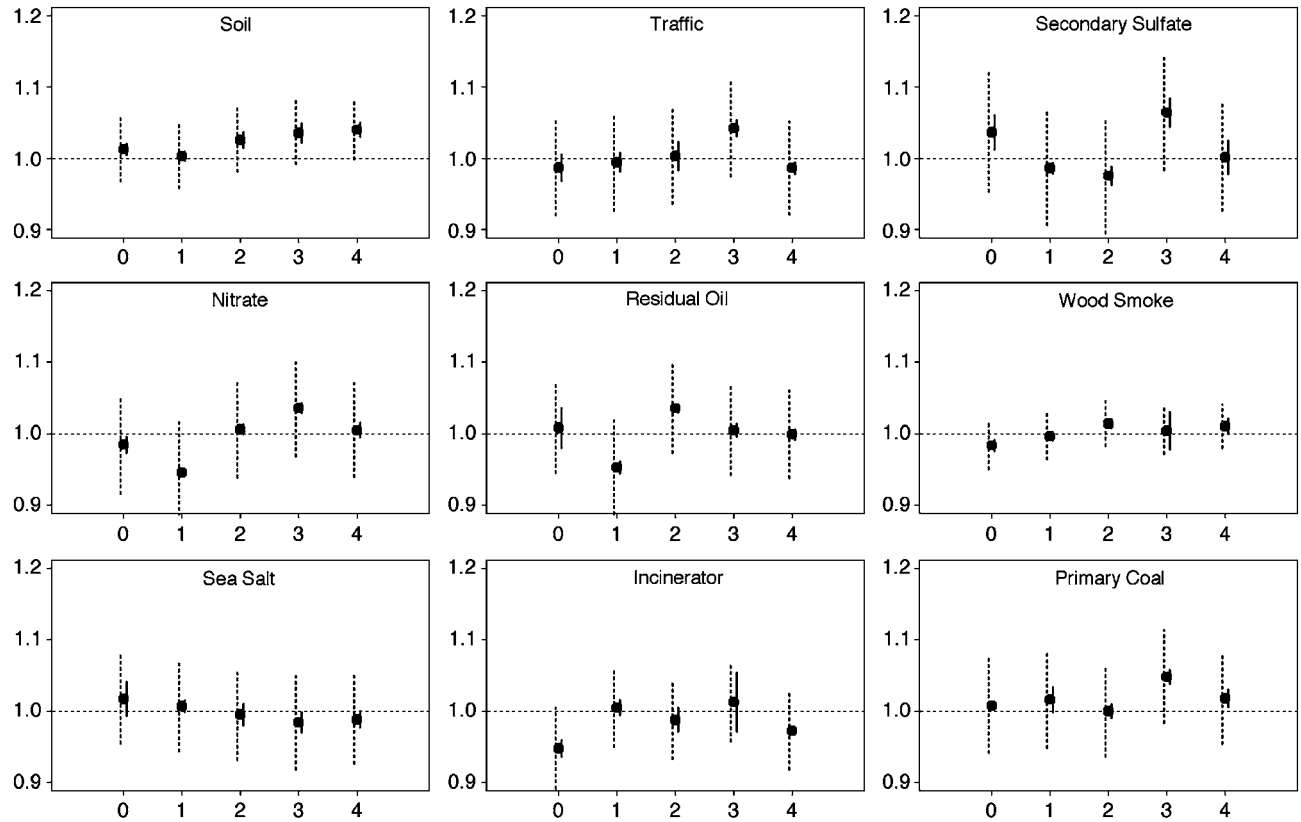

Figure 6. Relative risk and 95% CI of cardiovascular mortality associated with estimated source-apportioned $PM_{2.5}$ (variance-weighted), averaged across investigators/methods. *Y*-axis: relative risk per 5–95th percentile increment of estimated source-apportioned $PM_{2.5}$. *X*-axis: lag 0–5 days. The dotted 95% CIs are the confidence bands based on the average regression standard errors. The solid 95% CIs are the confidence bands based on the variance of point estimates across investigators/methods.

Table 2. Analysis of variance results: variation in total mortality percent excess risk estimates as a function of source type, investigators/methods, and lag days in a general linear model.

Source of variation	Degrees of freedom	Sequential sum of squares	Adjusted sum of squares	Adjusted mean squares	*F*-value	*P*-value
Pollution source type	8	219.3	230.5	28.8	8.19	<0.001
Investigators/methods	8	23.9	23.9	3.0	0.85	0.560
Lag days	4	458.2	458.2	114.6	32.56	<0.001
Error	264	928.9	928.9	3.5		
Total	284	1630.4				

Table 3. Analysis of variance results: variation in cardiovascular mortality percent excess risk estimates as a function of source type, investigators/methods, and lag days in a general linear model.

Source of variation	Degrees of freedom	Sequential sum of squares	Adjusted sum of squares	Adjusted mean squares	*F*-value	*P*-value
Pollution source type	8	312.6	309.0	38.6	5.29	<0.001
Investigators/methods	8	28.6	28.6	3.6	0.49	0.863
Lag days	4	592.1	592.1	148.0	20.28	<0.001
Error	264	1926.9	1926.9	7.3		
Total	284	2860.3				

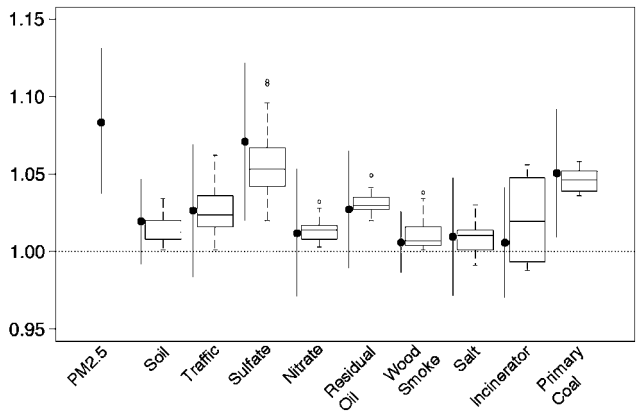

Figure 7. Total non-accidental mortality relative risks (RR) per 5–95th percentile of source-apportioned $PM_{2.5}$. The point estimates (except the total $PM_{2.5}$) are the variance-weighted average across investigators/methods, and the 95% CI bars are based on the average of standard error across investigators/methods at the most consistently significant lag for each source type, both are from the base model. The Box plots show the distribution of point estimates across investigators/methods using all of the six alternative weather models.

cases, were more variable across investigators (mean percent excess mortality $=2.6\%$ (95% CI: -1.6, 6.9)). Soil-related $PM_{2.5}$ showed smaller effect size estimates (mean percent excess mortality $=2.1\%$ (95%CI: -0.8, 4.9), but they were more consistently positive at multiple lags. Figure 7 shows these mean percent excess deaths for all the estimated source categories as well as the distribution of point estimates from the sensitivity analysis that examined the effects of alternative weather models. It is not surprising that the estimated effect size per 5–95th percentile for the sulfate-related $PM_{2.5}$ is similar to that for the total $PM_{2.5}$, as the sulfate-related $PM_{2.5}$ explain more than half of the total $PM_{2.5}$ in this city. The point estimates from the base model were mostly close to the median of the estimates from all the six weather models, except for sulfate, in which other models tended to yield somewhat smaller estimates than the base model.

Discussion

This analysis of source-apportioned $PM_{2.5}$ provided useful insights into the effects of PM source apportionment method variations on source apportionment health effects modeling, but also raised several issues. The ultimate goal of the application of source-apportionment to the PM health effects analysis is to identify component(s) of PM that may be especially harmful. However, as observed in this analysis, comparing the risk estimates across various source-apportioned $PM_{2.5}$ is not always straightforward.

First, the results of this study suggest that the lag structure of associations appears to vary across source types. For example, in this data set, the strongest associations for sulfate-related $PM_{2.5}$ were found at lag 3 day. In contrast, the risk estimates for soil-related $PM_{2.5}$ were more consistently positive across the four (0–4) lag days examined, although their effect size estimates were smaller than those for sulfate-related $PM_{2.5}$. The Wednesday/Saturday sampling schedule did not allow us to examine the distributed lag model, and therefore we could not compare the risk estimates across source types using the sums of the effects over several days. Thus, the use of $PM_{2.5}$ mass on one time lag may obscure the individual source components' lag effects, likely underpredicting the sum of the individual source's $PM_{2.5}$ effects.

Another complication is that the difference in lag structure of associations may be caused in part by the difference in the correlation between the source-apportioned PM and weather/temporal trend adjustment terms. Varying degrees of correlation with regression covariates across source types make it difficult to perform a "fair" comparison of source-specific PM effects, especially if these estimates are sensitive to alternative weather models. This issue needs to be examined using multiple cities where source types vary. An interesting implication of the possible difference in lag structure of health effects associations across source types is that, depending on the dominant source type(s) of the city, the lag structure of PM–health outcome associations may also vary across cities. Again, analyses of $PM_{2.5}$ speciation data in multiple cities should shed light on this factor.

In this data set, sulfate-related $PM_{2.5}$, the largest estimated fraction of $PM_{2.5}$, was most significantly associated with mortality (and with the largest effect size per the same distributional increment increase). Sulfate, being a secondary transported PM, tends to be uniformly distributed within the metropolitan Washington, DC area, and thus likely has relatively small exposure characterization error. In contrast, we expect the source types such as traffic, incinerator, and residual oil to be more locally influenced and therefore have larger exposure characterization error, which could have attenuated mortality associations. To evaluate the relative importance of source-specific PM health effects, we need to take into account these (possible) differential exposure errors across source types. We did not have quantitative information on these errors in this data set. A recent analysis of speciation data from three monitors in New York City suggests that $PM_{2.5}$ components vary spatially across source types (Ito et al., 2004). Interpretations of source-apportioned PM may need to incorporate such information.

While the risk estimates were reasonably consistent across investigators using different multivariate receptor models (note that no CMB type models were used), this does not guarantee the accuracy of the apportioned $PM_{2.5}$, or risk estimates that were apportioned to each of the estimated source types. Further research is needed in validating the accuracy of the apportioned mass concentrations.

Given the limitations and issues discussed above, our analysis still provided a great deal of useful information.

Despite the variety of source-apportionment techniques employed and the number of investigators involved, the variation of estimated risks across source types were larger than the variation of risks across investigators/methods. The lag structure and effect size of mortality associations appears to vary across source types, and they were generally consistent across investigators/methods. Sulfate-related $PM_{2.5}$, a major fraction (over 50%) of total $PM_{2.5}$, showed the largest excess risk estimates per 5–95th percentile increment among the source types identified. Analyses of $PM_{2.5}$ speciation data from multiple cities should resolve some of the issues raised in this study.

Acknowledgements

The Workshop was organized under the auspices of the participating U.S. EPA PM Health Effects Research Centers (Grant R827351 at NYU, R827351 at the University of Washington, R827353 at Harvard University, and R927354 at the University of Rochester). Kaz Ito's effort on this project has been supported by U.S. EPA STAR grant (R827997010) and NYU-NIEHS Center Grant (ES00260). We thank the individual researchers who undertook participation in this workshop, often on their own time and resources. Support for the organization and administration of the Workshop was also provided by the New York State Energy Research and Development Authority (NYSERDA Grant 375-34215).

References

Christensen W.F., Schauer J.J., and Lingwall J.W. Iterated confirmatory factor analysis for pollution source apportionment. *Environmetrics.* 2005 (in press).

Dominici F., McDermott A., Daniels M, Zeger S.L., and Samet J.M. Mortality among residents of 90 cities. In: *Revised Analyses of Time-Series Studies of Air Pollution and Health*. Special report. Health Effects Institute, Boston, MA, 2003, pp. 9–24. Available: http://www.healtheffects.org/pubs-special.htm [Accessed May 16, 2003].

Henry R.C., and Norris G.A. *UNMIX 2.3 User Guide*. U.S. EPA, Research Triangle Park, NC, 2002.

Huang S., Rahn K.A., and Arimoto R. Testing and optimization two factor-analysis techniques on aerosol at Narragansett, Rhode Island. *Atmos Environ* 1999: 33: 2169–2185.

Hopke P.K., Ito K., Mar T., Christensen W.F., Eatough D.J., Henry R.C., Kim E., Laden F., Lall R., Larson T.V., Liu H., Neas L., Pinto J., Stölzel M., Suh H., Paatero P., and Thurston G.D. PM source apportionment and health effects: 1. Intercomparison of source apportionment results. *J Exp Anal Environ Epidemiol* 2005, submitted.

Ito K., Xue N., and Thurston G.D. Spatial variation of $PM_{2.5}$ chemical species and source-apportioned mass concentrations in New York City. *Atmos Environ* 2004: 38: 5269–5282.

Klemm R.J., and Mason R. Replication of reanalysis of Harvard Six-City mortality study. In: *Revised Analyses of Time-Series Studies of Air Pollution and Health*. Special Report. Health Effects Institute, Boston, MA, 2003, pp. 165–172 Available: http://www.healtheffects.org/pubs-special.htm [Accessed May 16, 2003].

Klemm R.J., Mason Jr R.M., Heilig C.M., Neas L.M., and Dockery D.W. Is daily mortality associated specifically with fine particles? Data reconstruction and replication of analyses. *J Air Waste Manage Assoc* 2000: 50: 1215–1222.

Koutrakis P., and Spengler J.D. Source apportionment of ambient particles in Steubenville, OH using specific rotation factor analysis. *Atmos Environ* 1987: 21: 1511–1519.

Laden F., Neas L.M., Dockery D.W., and Schwartz J. Association of fine particulate matter from different sources with daily mortality in six US cities. *Environ Health Perspect* 2000: 108: 941–947.

McCullagh P., and Nelder J.A. *Generalized Linear Models*, 2nd edn. Chapman & Hall, New York, 1989.

Mar T.F., Ito K., Koenig J.Q., Larson T.V., Christensen W., Eatough D.J., Henry R.C., Kim E., Laden F., Lall R., Neas L., Hopke P.K., and Thurston G. An investigation of inter-method variability in associations between source apportioned fine particle mass and daily mortality in Phoenix, AZ. *J Expo Anal Environ Epidemiol* 2005, submitted.

Mar T.F., Norris G.A., Koenig J.Q., and Larson T.V. Associations between air pollution and mortality in Phoenix, 1995 to 1997. *Environ Health Perspect* 2000: 108: 347–353.

Mar T.F., Norris G.A., Larson T.V., Wilson W., and Koenig J.Q. Air pollution and cardiovascular mortality in Phoenix, 1995–1997. In: *Revised Analyses of Time-Series Studies of Air Pollution and Health*. Special report. Health Effects Institute, Boston, MA, 2003 pp. 177–182. Available: http://www.healtheffects.org/pubs-special.htm [Accessed May 16, 2003].

Özkaynak H., and Thurston G.D. Associations between 1980 US mortality rates and alternative measures of airborne particle concentration. *Risk Anal* 1987: 7: 449–460.

Özkaynak H., Xue J., Zhou H., and Raizenne M. Associations between daily mortality and motor vehicle pollution in Toronto, Canada. Boston, MA: Harvard University School of Public Health, Department of Environmental Health; March 25, 1996.

Paatero P., and Tapper U. Chemom. analysis of different modes of factor analysis as least squares fit problems. *Intell Lab Syst* 1993: 18: 183–194.

Poirot R.L., Wishinski P.R., Hopke P.K., and Polissar A.V. Comparative application of multiple receptor methods to identify aerosol sources in northern Vermont. *Environ Sci Technol* 2001: 35(23): 4622–4636.

Qin Y., Oduyemi K., and Chan L.Y. Comparative testing of PMF and CFA models. *Chemometr Intell Lab Systems* 2002: 61: 75–87.

Ramadan Z., Eickhout B., Song X.H., Buydens L., and Hopke P.K. Comparison of positive matrix factorization (PMF) and multilinear engine (ME-2) for the source apportionment of particulate pollutants. *Chemometr Intell Lab Systems* 2003: 66(1): 15–28.

Samet J., Zeger S., Dominici F., et al. *The National Morbidity, Mortality, and Air Pollution Study Part II: Morbidity, Mortality, and Air Pollution in the United States*. Health Effects Institute, Cambridge, MA, 2000.

Schwartz J. Daily deaths associated with air pollution in six US cities and short-term mortality displacement in Boston. In: *Revised Analyses of Time-Series Studies of Air Pollution and Health*. Special report. Health Effects Institute, Boston, MA, 2003 pp. 219–226. Available: http://www.healtheffects.org/pubs-special.htm [Accessed May 16, 2003].

Schwartz J., Dockery D.W., and Neas L.M. Is daily mortality associated specifically with fine particles? *J Air 6 Waste Manage Assoc* 1996: 46: 927–939.

Thurston G.D., Ito K., Mar T., Christensen W.F., Eatough D.J., Henry R.C., Kim E., Laden F., Lall R., Larson T.V., Liu H., Neas L., Pinto J., Stölzel M., Suh H., Paatero P., and Hopke P.K. The workshop on the source apportionment of PM health effects: inter-comparison of results and implications. *Environ Health Persp* 2005 (in press).

Thurston G.D., and Spengler J.D. A quantitative assessment of source contributions to inhalable particulate matter pollution in metropolitan Boston. *Atmos Environ* 1985: 19: 9–25.

Tsai F.C., Apte M.G., and Daisey J.M. An exploratory analysis of the relationship between mortality and the chemical composition of airborne particulate matter. *Inhal Toxicol* 2000: 12(Suppl.): 121–135.

U.S. EPA. *Air quality criteria for particulate matter* 1996. EPA/600/P-95/001.