Examining the effects of air pollution composition on within region differences in $PM_{2.5}$ mortality risk estimates Lisa K. Baxter, Sc.D^{1*}, Rachelle M. Duvall, Ph.D¹, and Jason Sacks, MPH²

¹ National Exposure Research Laboratory, U.S. Environmental Protection Agency, RTP, NC

² National Center for Environmental Assessment, U.S. Environmental Protection Agency, RTP, NC

*Corresponding author Lisa K. Baxter 109 T.W. Alexander Drive MD E205-2 RTP, NC 27711 Phone: 919-541-0671 Fax: 919-541-4787 E-mail: baxter.lisa@epa.gov

Running title: PM_{2.5} Composition and mortality estimates

Keywords: air pollution composition, PM_{2.5} mortality risk estimates, source apportionment

Abstract

Multi-city population-based epidemiologic studies have observed significant heterogeneity in both the magnitude and direction of city-specific risk estimates, but tended to focus on regional differences in PM_{2.5}- mortality risk estimates. Interpreting differences in risk estimates is complicated by city-to-city heterogeneity observed within regions due to city-to-city variations in the PM_{2.5} composition and the concentration of gaseous pollutants.

We evaluate whether variations in $PM_{2.5}$ composition and gaseous pollutant concentrations play a role in explaining the heterogeneity in $PM_{2.5}$ - mortality risk estimates observed in 27 U.S. cities from 1997 to 2002. Within each region, we select the two cities with the largest and smallest mortality risk estimate. We compare for each region the within- and between-city concentrations and correlations of $PM_{2.5}$ constituents and gaseous pollutants. We also attempt to identify source factors through Principal Component Analysis (PCA) for each city.

The results of this analysis indicate that identifying a PM constituent(s) that explains the differences in the $PM_{2.5}$ mortality risk estimates is not straight forward. The difference in risk estimates between cities in the same region may be attributed to a group of pollutants, possibly those related to local sources such as traffic.

Introduction

Multi-city epidemiologic studies conducted in the U.S. have observed clear regional patterns in PM-mortality risk estimates with the magnitude consistently being larger in the eastern U.S. (Franklin, et al., 2007; Peng, et al., 2005). The observed regional differences in PM-mortality risk estimates have often been attributed to a variety of factors including geographic

variability in particle composition, spatial heterogeneity of constituents, and differences between cities in the distribution of the population potentially at greatest risk of an air pollutant-related health effect (Levy, et al., 2012). Differences in topography and county size between eastern and western counties, which may lead to regional differences in exposure error, have also been hypothesized to contribute to the regional heterogeneity in PM mortality risk estimates observed (Davis, et al., 2011).

Although it is important to identify regional differences in $PM_{2.5}$ - mortality risk estimates, interpreting these differences is complicated by the city-to-city heterogeneity in risk estimates observed within regions. These differences in $PM_{2.5}$ - mortality risk estimates may be attributed to city-to-city variations in the composition of $PM_{2.5}$ and the concentration of gaseous pollutants such as carbon monoxide (CO), nitrogen dioxide (NO₂), ozone (O₃) and sulfur dioxide (SO₂) (Bell, et al., 2007; Dominici, et al., 2003; Sarnat, et al., 2001).

We evaluated whether variations in $PM_{2.5}$ composition and gaseous pollutant concentrations play a role in explaining the heterogeneity in the association between ambient $PM_{2.5}$ concentrations and mortality observed in 27 U.S. cities from 1997 to 2002 as reported by Franklin et al. (2007). The 27 cities were divided into 4 regions: Northeast, South, Midwest, and West. Within each region, we selected the two cities with the largest and smallest mortality risk estimate. We then conducted a semi-quantitative analysis by examining the within- and betweencity concentrations and correlations of the $PM_{2.5}$ components and gaseous pollutants. Finally, we identified source factors through Principal Component Analysis (PCA) for each city within each region.

Methods

Franklin et al. (2007) used a time-stratified case-crossover design to examine the association between $PM_{2.5}$ and all-cause (non-accidental), cardiovascular, and respiratory mortality using 1.3 million deaths in 27 U.S. cities between 1997 and 2002. For the purposes of this analysis we focused on the all-cause mortality risk estimates. We divided the 27 cities into 4 regions (Northeast, South, Midwest, and West) as defined by the United States Census Bureau. Within each region we then chose the two cites with the largest and smallest mortality risk estimate. Table 1 identifies the cities chosen along with their corresponding mortality risk estimates and 95% confidence intervals.

This analysis used data generated from the U.S. Environmental Protection Agency's (EPA) Aerometric Information Retrieval Service (AIRS) Database (United States Environmental Protection Agency, 2009) consisting of 1-in-3 or 1-in-6 day ambient concentration measurements of PM_{2.5} and PM_{2.5} constituents from the EPA's Chemical Speciation Network (CSN) and daily concentrations of gaseous pollutants (CO, NO₂, O₃, and SO₂) from 2001-2005. These years do not span the same years as the analysis conducted by Franklin et al. (2007) because PM speciation data was not collected by EPA until 2001. To include a sufficient number of measurements for this analysis, the study period was expanded to 2005. Where available, gaseous pollutant concentrations were obtained from collocated monitors, otherwise we used the next nearest non-CSN monitor within 20 miles. For comparability with the 24-hour filter-based measurements of PM constituents, hourly measurements of gaseous pollutants were combined into 24-hour averages.

We compared the city pairs using a number of analyses including 1) calculating the daily contribution (%) of various daily $PM_{2.5}$ constituents to daily $PM_{2.5}$ total concentrations, and daily concentrations of gaseous pollutants; 2) examining the correlations between cities for each

pollutant; 3) calculating the coefficient of divergence for each city and pollutant; 4) calculating the correlations between the $PM_{2.5}$ constituents and gaseous pollutants with $PM_{2.5}$ for each city; and 5) identifying the source factors for each city through principal components analysis (PCA).

In the first analysis, we examined potential differences in the PM_{2.5} mixture in the city pairs by calculating the percent contribution of each constituent to the total PM_{2.5} mass. We used this approach instead of focusing on the absolute concentration of each constituent because the concentration of a constituent could fluctuate daily (i.e., a constituent concentration on a given day could be due to a high aggregate PM_{2.5} concentration) and not accurately reflect the PM_{2.5} composition. For each day with PM_{2.5} constituent data, we divided the PM_{2.5} constituent concentrations by the aggregate PM_{2.5} concentration and multiplied by 100. We also considered the potential that differences in gaseous pollutant concentrations in the city pairs could be influencing the observed differences and, therefore, also examined the distribution of gaseous pollutant concentrations in the cities with the highest PM_{2.5} mortality risk estimates may have larger concentrations in the cities with the highest PM_{2.5} mortality risk estimate. We used a Wilcoxon signed rank sum test to determine significant differences between the city pairs.

In the second and third analyses, we calculated the Spearman correlation (ρ) and coefficient of divergence (COD) for each city pair and pollutant to examine both temporal and spatial variability, respectively. Correlations less than 0.5 were considered weak. If a pollutant is well correlated between cities then we would not expect that pollutant to explain the difference in mortality risk estimates.

While correlations can help determine if concentrations covary over time, the COD provides information as to whether overall concentrations are similar across communities. The COD for each city pair and pollutant was calculated as:

$$COD_{j}^{i,k} = \sqrt{\frac{1}{p_{j}^{i,k}} \sum_{t(j \in i,k)}^{p_{j}^{i,k}} \left(\frac{x_{t,j}^{i} - x_{t,j}^{k}}{x_{t,j}^{i} + x_{t,j}^{k}}\right)^{2}}$$
(1)

where $x_{t,j}^i$ and $x_{t,j}^k$ are the concentration of constituent *j* on day *t* at monitor site (i.e. city) *i* and *k*, respectively, and $p_j^{i,k}$ the number of days with data for constituent *j* for both monitor sites *i* and *k*. If the coefficient approaches 0, the measurements have high similarity, whereas if the coefficient approaches 1, the measurements have low similarity. If measurements were identical, the COD would equal 0. A COD > 0.2 indicates significant spatial heterogeneity between cities for an individual pollutant (Blanchard, et al., 1999; Wilson, et al., 2005), and identifies it as a potential candidate for explaining the difference between mortality risk estimates.

Since we are interested in identifying why there are differences in PM_{2.5} mortality risk estimates between cities in the same region, we calculated the Spearman correlations between each constituent and gaseous pollutant and PM_{2.5} for each city in the fourth analysis. In order for a pollutant to explain the difference in PM_{2.5}-mortality risk estimates it would have to be correlated with PM_{2.5}. If the pollutant and PM_{2.5} are very highly correlated than it would not explain the difference in mortality risk estimates. In contrast, if a pollutant and PM_{2.5} are not correlated than it would have no effect on the PM_{2.5}- mortality risk estimate. We used the Fisher r-to-z calculation to determine if there was a significant difference between the two correlation coefficients (Rosner, 2000). We would expect a pollutant that contributes to the difference in mortality risk estimates to be more highly correlated with PM_{2.5} in the city with the higher mortality risk estimate.

Finally, we evaluated the composition of the air pollution mixes by determining potential sources of air pollution for each city pair using PCA. PCA is a variable reduction procedure that is useful when you have obtained data on a number of variables that may be highly correlated with one another. PCA is used to reduce the observed variables into a smaller number of uncorrelated principal components (i.e., artificial variables) that will account for most of the variance in the observed variables (Beebe, et al., 1998). We hypothesize that it may be a particular source that is responsible for the higher PM_{2.5}-mortality risk estimates in one city compared to another. We applied PCA in each city using an oblique rotation method to selected PM_{2.5} constituents including aluminum (Al), ammonium (NH₄), bromine (Br), calcium (Ca), chlorine (Cl), copper (Cu), elemental carbon (EC), iron (Fe), manganese (Mn), nickel (Ni), nitrate (NO₃), organic carbon (OC), potassium (K), selenium (Se), silicon (Si), sodium (Na), sulfate (SO₄), sulfur (S), titanium (Ti), vanadium (V), zinc (Zn) as well as CO, NO₂, O₃, and SO₂. These species were selected because they are reliably measured by x-ray fluorescence (XRF) and are often used as key molecular markers for specific PM sources. We also wanted to focus on pollutants that have previously been shown to have adverse health effects (United States Environmental Protection Agency, 2009). All analyses were conducted with SAS 9.1 (SAS Institute Inc., 2003).

Results

Daily percent contributions and concentrations

We performed a Wilcoxon signed rank sum test to determine significant differences between the city pairs for both $PM_{2.5}$ constituents and gaseous co-pollutants. In Table 2 we indicate whether there are significant differences between the city pair and for which city the

percent daily contribution or concentration is higher. For ease of presentation the contribution of $PM_{2.5}$ constituents are categorized, using the median percent, into high (over 5%), medium (between 0.1% and 5%), and low (less than 0.1%) percent contributions.

Focusing on species that are significantly different between the city pairs, in the northeastern region percent daily contributions were higher in Boston for EC, OC, Ca, Cl, Na, Ni, Ti, V, and NO₂ while percent daily contributions were higher in Pittsburgh for NH₄, SO₄, S, K, , Zn, Br, Cu, Mn, and Se. When examining constituent percent contributions in the southern region, Birmingham and Memphis, the percent contribution of constituents was higher in Memphis for only NH₄, NO₃, SO₄, S, Cl, and K. The percent contribution for most of the constituents in the southern region were greater in Birmingham: Al, Ca, Fe, Si, Na, Zn, Cu, Mn, Ni, Se, and V. In the Midwest, percent contributions were higher in Milwaukee for NO₃, OC, Al, Cl, K, Na, Br, Mn, Ni, and Se and higher in Detroit for EC, SO₄, S, , Ca, Fe, Zn, and Cu. Similar to the southern region the majority of daily contributions in the West were similar or higher in one of the communities, in this case San Diego. However, the percent contribution was higher in Riverside for NH₄, NO₃, Ca, K, Si, and Zn.

We observed significantly higher concentrations of CO, O_3 , and SO_2 in Pittsburgh compared to Boston, and higher concentrations of NO_2 in Boston. For the southern region, CO and SO_2 concentrations were higher in Birmingham while Memphis exhibited higher concentrations of O_3 . Of note, a comparison of NO_2 concentrations in the South region was not possible due to the lack of an NO_2 monitoring site within 20 miles of the Birmingham CSN monitor. In the Midwest region, concentrations of NO_2 and SO_2 were higher in Detroit than in Milwaukee. Finally, in the western region, concentrations of NO_2 and SO_2 were higher in San Diego whereas O_3 concentrations were higher in Riverside. Figures S-1 through S-4 present the distributions of the daily percent contribution of $PM_{2.5}$ constituents to $PM_{2.5}$ mass for each city pair. Summary statistics for the $PM_{2.5}$ and $PM_{2.5}$ constituents for each city are presented in Tables S-1 through S-4 and for the concentrations of gaseous pollutants are shown on Table S-5.

Correlations between cities

Spearman correlation coefficients between city pairs for each pollutant are presented in Figure 1. Horizontal reference lines indicate a correlation of 0.50 and 0.70. We generally considered strong correlations as those over 0.70 and a moderate correlation as one between 0.50 and 0.70. We observed moderate to strong correlations between OC ($\rho = 0.51$), NO₂ ($\rho = 0.53$), and O₃ ($\rho = 0.65$) for Boston and Pittsburgh. For the southern communities (Birmingham and Memphis), moderate correlations were observed for NH₄, NO₃,OC, Si, and O₃, and strong correlations for SO₄ and S.. Moderate correlations between Detroit and Milwaukee were observed for NH₄, OC, K, CO, NO₂, and O₃, and strong correlations for NO₃, SO₄, and S. Compared to the other regions, a larger number of pollutants were more strongly correlated in the West region (San Diego and Riverside) including NH₄, EC, Fe, NO₃, SO₄, S, NO₂ and O₃. Moderate correlations were observed for remaining pollutants with the exceptions of Al, Cl, Cu, Mn, Ni, Se, and SO₂.

Coefficient of divergence

We examined the spatial heterogeneity of percent daily $PM_{2.5}$ contributions between cities in each region using the coefficient of divergence (COD) with values of 0.20 and above indicating significant heterogeneity. The CODs were greater than 0.20 (Table S-6) for the majority of city pairs and pollutants. The exceptions were OC and S for Birmingham-Memphis, and SO₄ and S for Detroit-Milwaukee which had COD values below 0.20.

Within-City correlations with PM_{2.5}

Table 3 presents the Spearman correlations coefficients between each pollutant and $PM_{2.5}$ mass for each city. Numbers shaded in grey indicate where the correlation coefficients are significantly different between city pairs based on a Fisher r-to-z calculation. Here we focus results on constituents where the higher correlation is observed in the city with the larger mortality risk estimate. We also examined the magnitude of the correlations.

In the northeast, NO₃ and V were more correlated to $PM_{2.5}$ in Boston. For NO₃, the correlation was much greater in Boston ($\rho = 0.62$) compared to Pittsburgh ($\rho = 0.13$). In contrast, correlations between $PM_{2.5}$ and V were weak ($\rho < 0.5$) in both cities. For the southern cities only NH₄ was more highly correlated with $PM_{2.5}$ in Memphis - the city with the higher mortality risk estimate. However, correlations between NH₄ and PM_{2.5} were strong ($\rho > 0.7$) in both Memphis and Birmingham. In the Midwest, NH₄, EC, NO₃, Fe, K, Zn, Cu, and Mn had higher correlations in Milwaukee. In both Milwaukee and Detroit we observed strong or moderate correlations for NH₄, NO₃, and K and weak correlations for Cu and Mn. Larger differences between cities were seen for EC, Fe, and Zn. Finally, in the west region, a higher correlation with PM_{2.5} was observed in San Diego than in Riverside for EC, Fe, Si, Zn, NO₂, and SO₂. EC, Fe, Zn, and NO₂ were at least moderately correlated with PM_{2.5} in San Diego and weakly correlated with PM_{2.5} in Riverside. Correlations were weak in both cities for Si and SO₂.

Principal Components Analysis

The results from PCA are summarized in Tables 4 –7. We initially tested a different number of factors ranging from 4-7 for each city. For consistency across cities, we used 5 factors for each city in the PCA analysis five source factors were identified for each city. PCA identified two consistent source factors, coal combustion and crustal in all of the cities. The coal combustion/secondary sulfate source factor contained high loadings on NH₄, SO₄, S, Se (Birmingham, Milwaukee, and Detroit only), and SO₂ (Birmingham and Detroit only). The crustal material source factor consisted of high loadings on Al, Si, Ca (Boston, Milwaukee, Detroit, San Diego, and Riverside only), and Ti (Boston, Birmingham, Detroit, and San Diego only) which are often associated with crustal material. The remaining source factors included traffic-related (OC, EC, Zn, CO, NO₂), sea salt (Cl, Na), brake wear (Cu), industrial (Fe, Mn, Zn), and residual oil (Ni, V) which differed among the cities or grouped with each other. Factors determined for Boston were clearly identified. However, an unknown factor was found for the remaining cities including high loadings on Cu/K/V (Pittsburgh), Cu/Ni (Birmingham), Cu/K/ Ti/V (Memphis), Cu/K (Detroit), Cu/K/Ti (Milwaukee and Riverside), and Ni (San Diego). In some cases, the constituents normally characteristic of a particular source did not always load on the same factor. For example, gaseous pollutants often associated with traffic, CO and NO₂, did not always load on the traffic-related factor.

Discussion

The semi-quantitative analysis conducted within this study to examine potential differences in the air pollution mixture between cities represents a novel approach to identify those pollutants that contribute to the city-to-city heterogeneity in PM-mortality risk estimates observed within regions of the U.S. In addition the use of PCA addresses the fact that air

pollution exposures are the result of a complex mixture and those differences in the mixture could potentially explain the observed heterogeneity. We were unable to identify systematic differences in air pollution components associated with higher mortality rates in cities across the country. However, the methodological approach used in this study to examine potential compositional differences between cities could be used to further characterize the air pollution mixture between cities.

Previous epidemiologic studies that have attempted to identify the most toxic constituents of PM_{2.5} have often relied on second stage analyses to identify if a specific constituent acts as an effect modifier of the PM-mortality relationship (Bell, et al., 2009; Franklin, et al., 2008), or have focused on examining associations with mortality using the constituents that are both highly correlated with and that dominate PM_{2.5} mass (Levy, et al., 2012). Of these studies, only Levy et al. (2012) attempted to examine whether there are differences in mortality risk estimates on a regional scale for each constituent, which could explain the heterogeneity in PM mortality risk estimates. Although this type of analysis is informative, it neglects the issue of the city-to-city heterogeneity in both the magnitude and direction of PM_{2.5}-mortality risk estimates often observed within a region, which complicates the interpretation of regional results.

Some approaches to identifying potential differences in the mixture between cities were more informative than others. For example, although significant differences were observed between cities when examining the percent contribution of individual constituents to the total PM_{2.5} mass and the concentration of gaseous pollutants, overall, the relative magnitude of this difference was not that large (Figures S-1 through S-4 and Table S-5). Similarly, the COD analysis found that there was significant spatial heterogeneity between cities for most pollutants.

Correlations between cities for PM_{2.5} constituents and gaseous pollutants in conjunction with correlations within cities for PM_{2.5} constituents and gaseous pollutants with PM_{2.5} mass were found to be more informative. We identified pollutants with correlations less 0.7 between cities and significantly different correlations with PM_{2.5} within a city. Using this approach and focusing on cities with the higher PM-mortality risk estimate in each region the following pollutants were identified that fit the above criteria: NO₃ and V in Boston, and NH₄ EC, Fe, K, and Zn in Milwaukee, and Zn in San Diego. No pollutants that met these criteria were identified for the southern cities. Interestingly across two of the cities, Zn was identified to be less correlated between cities and moderately correlated with PM_{2.5} within the two cities with the largest risk estimate (Milwaukee and San Diego).

While Zn correlations between cities and with $PM_{2.5}$ could potentially inform the difference in risk estimates, overall we were unable to clearly identify one constituent that is responsible for the difference in risk estimates between cities. It is rather more likely that a mixture of pollutants are contributing to the observed health effects lending support to conducting a source-based analysis, such as PCA, to identify potential sources of air pollution in each city.

Across the cities, PCA identified source factors often associated with a traffic-related, coal combustion, crustal, industrial, salt, nitrate, or brake wear source as either a single source factor or a mixed source factor. The only source factors consistent between cities were representative of coal combustion and crustal sources. More differences between the cities were observed for local sources (e.g. motor vehicles, industry). While it is clear that each city is impacted by different air pollution source mixtures, it is unclear which sources contribute to the differences in risk estimates between the cities.

Although a semi-quantitative analysis of PM composition between cities, as described in this study, may provide some information regarding the difference in the air pollution mixture it still remains difficult to identify the group of pollutants that contribute to the observed heterogeneity in PM-mortality risk estimates between cities. This is especially true when examining two cities in the same region of the U.S. that have previously been shown to have strong associations with PM_{2.5} mortality such as the Northeast (e.g., Pittsburgh and Boston) and industrial midwest (e.g., Detroit and Milwaukee) (Franklin, et al., 2007; Peng, et al., 2005), where the difference in risk estimates between the two cities is only in magnitude and not direction. The observed heterogeneity may actually not be attributed to compositional differences, but instead exposure differences.

These exposure differences could be due to either exposure measurement error or cityspecific exposure factors that influence personal-ambient exposure relationships. Bell et al. 2011 demonstrated strong spatial heterogeneity among PM_{2.5} constituents at the city level, leading to different levels of exposure misclassification by constituent. Therefore, the use of county-wide average PM_{2.5} constituent concentrations may not accurately reflect the true difference in exposures between two cities in the same region. Additionally, Baxter et al. 2011 demonstrated that city-specific exposure factors may explain some of the city-to-city variability in PM_{2.5} mortality risk estimates. For example, in the cities examined in this analysis, the majority of cityspecific exposure factors are similar between cities in the same region (i.e., median year of homes built, median normalized leakage, and annual and seasonal temperatures). However, within each region, the city with the higher mortality risk estimate has a much larger percent of people living in apartments and in the case of the East and West regions, the city with the smaller mortality risk estimate has a much larger prevalence of homes with central air conditioning.

Overall, these examples indicate that more than likely the difference in $PM_{2.5}$ mortality risk estimates is not as simplistic as identifying differences in $PM_{2.5}$ composition or potential air pollution sources between cities, but also includes a variety of potential exposure differences.

There are several limitations for this analysis. First, the years of PM constituent and gaseous pollutant data are slightly different than the years of data used in Franklin et al. 2007 (2001-2005 compared to 1997-2002). It is assumed that the $PM_{2.5}$ constituent and gaseous pollutant concentrations were relatively similar between these two time periods. This analysis is also limited by the 1-in-3 or 1-in-6 day sampling schedule employed by CSN monitors. We may therefore miss day-to-day variations or episodes of high concentrations such as due to wildfires. Additionally, the reported concentrations of PM constituent data from CSN sites for some of the pollutants are below the known monitor detection limit. However, for source apportionment sometimes species that tend to have >50% of data below MDL are included in an analysis because they are key tracers for sources (e.g. Se for coal combustion, V for oil combustion) and/or have been shown to cause adverse health effects. Our city selection criteria are also based on climatic region and ignore distance. Higher correlations would be expected between cities that are closer in distance (e.g. San Diego and Riverside) compared to cities further from one another (e.g. Boston and Pittsburgh). Finally, the difference in PM_{2.5}-mortality risk estimates between cities could be due to differences in individual- or population-level characteristics between cities such as the age distribution of the population or the distribution of the population encompassing a certain socioeconomic status (Franklin, et al., 2007; Ostro, et al., 2006; Ostro, et al., 2008; Zeka, et al., 2006).

Conclusions

This analysis attempts to identify the key constituents or groups of pollutants (i.e., constituents and gaseous pollutants) that contribute to the observed difference in $PM_{2.5}$ mortality risk estimates between cities in the same region. The semi-quantitative approach used within this study further supports the hypothesis that the difference between cities is due to differences in the air pollution mixture and not only a few pollutants. The systematic approach used in this study to examine potential compositional differences between cities could be used to further characterize the air pollution mixture between cities (e.g. comparing cities within the same region with increased mortality risk only or decreased risk only). We also used a source apportionment tool to help identify contributions of air pollution sources. Overall, the results of this analysis indicate that identifying a PM constituent(s) that explains the difference in the associations between ambient concentrations and mortality within a region is not straight forward. The difference in $PM_{2.5}$ mortality risk estimates between cities in the same region can likely be attributed to differences in the total air pollution mixture as well as differences in exposure (i.e., exposure measurement error or city-specific exposure factors).

Acknowledgments

The authors would like to thank Lucas Neas of the U.S. EPA's National Health and Environmental Effects Research Laboratory. The authors would like to thank Tom Luben of the U.S EPA's National Center for Exposure Assessment and Kathie Dioniosio of the U.S. EPA's National Exposure Research Laboratory for their review of this paper.

Disclaimer

The United States Environmental Protection Agency through its Office of Research and Development funded and managed the research described here. It has been subjected to Agency's administrative review and approved for publication.

References

Baxter L., Franklin M., Özkaynak H., Schultz B., and Neas L. The use of improved exposure factors in the interpretation of fine particulate matter epidemiological results. *Air Qual Atmos Health* 2011: 1-10.

Beebe K.R., Pell R.J., and Seasholtz M.B. *Chemometrics: A Practical Guide*. John Wiley &Sons, Inc.: New York, NY, 1998.

Bell M.L., Dominici F., Ebisu K., Zeger S.L., and Samet J.M. Spatial and temporal variation in PM_{2.5} chemical composition in the United States for health effects studies. *Environ Health Perspect* 2007: 115(7): 989-995.

Bell M.L., Ebisu K., Peng R.D., Samet J.M., and Dominici F. Hospital Admissions and Chemical Composition of Fine Particle Air Pollution. *Am J Respir Crit Care Med* 2009: 179(12): 1115-1120.

Bell M.L., Ebisu K., and Peng R.D. Community-level spatial heterogeneity of chemical constituent levels of fine particulates and implications for epidemiological research. *J Expo Sci Environ Epidemiol* 2011: 21(4): 372-384.

Blanchard C.L., Carr E.L., Collins J.F., Smith T.B., Lehrman D.E., and Michaels H.M. Spatial representativeness and scales of transport during the 1995 integrated monitoring study in California's San Joaquin Valley. *Atmos Environ* 1999: 33(29): 4775-4786.

Davis J.A., Meng Q., Sacks J.D., Dutton S.J., Wilson W.E., and Pinto J.P. Regional variations in particulate matter composition and the ability of monitoring data to represent population exposures. *Science of The Total Environment* 2011: 409(23): 5129-5135.

Dominici F., McDermott A., Zeger S.L., and Samet J.M. National maps of the effects of particulate matter mortality: exploring geographic variation. *Environ Health Perspect* 2003: 111(1): 39-43.

Franklin M., Zeka A., and Schwartz J. Association between PM_{2.5} and all-cause and specificcause mortality in 27 US communities. *J Expo Sci Environ Epidemiol* 2007: 17(3): 279-287.

Franklin M., Koutrakis P., and Schwartz J. The role of particle composition and the association between PM_{2.5} and mortality. *Epidemiology* 2008: 19(5): 680-689.

Levy J.I., Diez D., Dou Y., Barr C.D., and Dominici F. A Meta-Analysis and Multisite Time-Series Analysis of the Differential Toxicity of Major Fine Particulate Matter Constituents. *Am J Epidemiol* 2012.

Ostro B., Broadwin R., Green S., Feng W.Y., and Lipsett M. Fine particulate air pollution and mortality in nine California counties: results from CALFINE. *Environ Health Perspect* 2006: 114(1): 29-33.

Ostro B.D., Feng W.-Y., Broadwin R., Malig B.J., Green R.S., and Lipsett M.J. The impact of components of fine particulate matter on cardiovascular mortality in susceptible subpopulations. *Occup Environ Med* 2008: 65(11): 750-756.

Peng R.D., Dominici F., Pastor-Barriuso R., Zeger S.L., and Samet J.M. Seasonal Analyses of Air Pollution and Mortality in 100 US Cities. *Am J Epidemiol* 2005: 161(6): 585-594.

Rosner B. Fundamental of biostatistics, 5th edn. Brooks/Cole: Pacific Grove, Ca, 2000.

Sarnat J.A., Schwartz J., Catalano P.J., and Suh H.H. Gaseous pollutants in particulate matter epidemiology: confounders or surrogates? *Environ Health Perspect* 2001: 109(10): 1053-1061.

SAS Institute Inc. Version 9.1 of the SAS System for Windows. *Copyright 2002-2003*: Cary, NC USA, 2003.

United States Environmental Protection Agency. Aerometric Information Retrieval Service (AIRS) Database. <u>http://www.epa.gov.air/data/info.html</u>, 2009. Accessed.

United States Environmental Protection Agency. Integrated Science Assessment for Particulate Matter. Office of Research and Development, National Center for Environmental Assessment-RTP Division: Research Triangle Park, NC,2009.

Wilson J.G., Kingham S., Pearce J., and Sturman A.P. A review of intraurban variations in particulate air pollution: Implication for epidemiological research. *Atmos Environ* 2005: 39: 6444-6462.

Zeka A., Zanobetti A., and Schwartz J. Individual-Level Modifiers of the Effects of Particulate Matter on Daily Mortality. *Am J Epidemiol* 2006: 163(9): 849-859.

Region	City	% (95% CI)
Northeast	Boston, MA Pittsburgh, PA	2.1 (0.01, 4.2) 0.48 (-0.88, 1.8)
South	Memphis, TN Birmingham, AL	4.0 (1.4, 6.5) -2.4 (-4.5, -0.33)
Midwest	Milwaukee, WI Detroit, MI	10.5 (8.5, 12.5) 0.43 (-0.59, 1.5)
West	San Diego, CA Riverside, CA	1.4 (0.31, 2.5) -0.83 (-1.7, 0.07)

Table 1. List of selected communities and corresponding percent increases in all-cause mortality for a 10 μ g/m³ increase in the previous days PM_{2.5} concentrations (from Franklin et al. 2007)

		No	ortheast	South		Midwest		West	
		Boston	Pittsburgh	Memphis	Birmingham	Milwaukee	Detroit	San Diego	Riverside
ıt	$\rm NH_4$	а	X ^b	Х					Х
cei	EC	Х					Х		
Perutio	NO_3			Х		Х			Х
r trib	OC	Х				Х		Х	
ligh	SO_4		Х	Х			Х	Х	
ΗO	S		Х	Х			Х	Х	
nt	Al				Х	Х			
ICEI	Ca	Х			Х		Х		Х
Pe	Cl	Х		Х		Х			
uo	Fe				Х		Х		
ian tributi	K		Х	Х		Х			Х
	Si				Х				Х
Aec Con	Na	Х			Х	Х		Х	
20	Zn		Х		Х		Х		Х
nt	Br		Х			Х		Х	
rce	Cu		Х		Х		Х	Х	
Pe	Mn		Х		Х	Х		Х	
utic	Ni	Х			Х	Х		Х	
rib	Se		Х		Х	Х		Х	
ow ont	Ti	Х							
ΟΓ	V	Х			Х			Х	
S	СО		Х		Х				
ant	NO_2	Х		N/A ^c	N/A		Х	Х	
sec llut	O_3		Х		Х				Х
Po	SO_2		Х		Х		Х	Х	

Table 2. Comparison of percent daily contributions of PM_{2.5} species to daily PM_{2.5} mass and gaseous pollutant concentrations between city pairs

^a blank indicates that species was not significantly different between city pairs based on a Wilcoxon rank sum test ^b "X" denotes city with significantly (p < 0.05) higher percent daily contribution based on a Wilcoxon rank sum test ^c NO₂ concentrations not available in Birmingham so no comparison was made



Figure 1. Spearman correlations between cities for select PM_{2.5} constituents and gaseous pollutants in the 4 regions a) Northeast: Boston, MA and Pittsburgh, PA; b) South: Birmingham, AL and Memphis, TN; c) Midwest: Detroit, MI and Milwaukee, WI; and d) West: San Diego, CA and Riverside, CA

	Northeast		S	South			est	We	West		
	Boston	Pittsburgh	Memphis	Birmingham		Milwaukee	Detroit	San Diego	Riverside		
NH ₄	0.78	0.90	0.81	0.71		0.92	0.88	0.69	0.91		
EC	0.47	0.66	0.47	0.66		0.57	0.43	0.61	0.39		
NO ₃	0.62	0.13	0.14	0.12		0.72	0.62	0.79	0.94		
OC	0.69	0.79	0.67	0.81		0.62	0.67	0.63	0.72		
SO_4	0.78	0.88	0.78	0.76		0.83	0.83	0.39	0.70		
S	0.80	0.88	0.80	0.77		0.84	0.83	0.38	0.70		
Al	0.10	0.09	0.04	0.25		0.12	0.09	0.16	0.002		
Ca	0.41	0.58	0.35	0.62		0.48	0.39	0.32	0.10		
Cl	-0.002	0.09	-0.07	0.12		0.19	0.20	0.03	0.45		
Fe	0.46	0.60	0.44	0.61		0.54	0.35	0.53	0.30		
Κ	0.61	0.59	0.27	0.64		0.69	0.57	0.49	0.40		
Si	0.34	0.53	0.21	0.63		0.39	0.36	0.38	0.13		
Na	0.10	0.12	-0.06	0.04		-0.02	0.46	-0.12	0.08		
Zn	0.38	0.47	0.50	0.45		0.65	0.43	0.65	0.33		
Br	0.59	0.52	0.51	0.53		0.56	0.64	0.59	0.79		
Cu	0.30	0.47	0.26	0.42		0.40	0.24	0.24	0.28		
Mn	0.30	0.44	0.33	0.51		0.42	0.30	0.33	0.25		
Ni	0.35	0.31	0.001	0.02		0.20	0.26	0.38	0.52		
Se	0.34	0.49	0.35	0.38		0.20	0.53	-0.09	0.16		
Ti	0.27	0.50	0.35	0.53		0.30	0.32	0.39	0.28		
V	0.44	0.20	0.17	0.18		0.11	0.25	0.47	0.62		
CO	0.26	0.36	0.22	0.12		0.37	0.35	0.37	0.40		
NO ₂	0.41	0.46	0.27	N/A ^a		0.55	0.59	0.57	0.37		
O ₃	-0.04	0.14	0.29	0.30		0.20	0.47	0.27	0.06		
SO_2	0.34	0.42	-0.01	0.19		0.52	0.47	0.23	-0.03		

Table 3. Spearman correlations for selected $PM_{2.5}$ constituents and gaseous pollutants with $PM_{2.5}$ mass for each city

^a nearest NO₂ monitor was more than 20 miles away from STN monitor so measurements were excluded

Shaded numbers indicate that correlation coefficients are significantly different between city pairs

			Boston		Pittsburgh					
	Traffic Related/					Industrial/		Salt/		
	Residual Oil	Coal	Crustal	Salt	Brake Wear	Traffic	Coal	Nitrate	Unknown	Crustal
Al	-0.18	-0.12	0.82	0.08	0.06	-0.11	-0.13	-0.09	0.31	0.86
NH_4	0.15	0.92	-0.07	-0.01	0.08	-0.07	0.92	0.20	0.05	-0.04
Br	0.20	0.51	0.12	0.39	0.09	0.26	0.24	0.58	0.09	0.01
Ca	0.08	0.07	0.78	0.18	-0.08	0.46	0.17	-0.16	0.38	0.28
Cl	-0.03	-0.07	-0.06	0.92	0.03	0.09	-0.08	0.57	0.03	0.01
Cu	-0.12	-0.15	-0.06	0.01	0.84	0.29	0.13	0.19	0.54	-0.01
EC	0.72	0.13	0.21	-0.10	0.06	0.62	0.25	0.30	0.07	0.05
Fe	0.35	0.10	0.73	-0.09	0.11	0.73	0.07	0.02	0.28	0.19
Mn	0.07	0.02	0.46	-0.14	0.39	0.87	-0.09	-0.03	0.05	0.01
Ni	0.67	0.05	-0.13	0.12	0.19	0.57	0.08	-0.07	0.00	-0.11
NO_3	0.64	0.32	-0.06	0.10	0.13	-0.16	-0.06	0.82	0.10	-0.05
OC	0.27	0.54	0.23	-0.20	0.16	0.38	0.50	0.16	0.13	0.10
Κ	-0.15	0.19	0.29	0.15	0.38	0.02	0.03	-0.06	0.63	0.31
Se	0.12	0.36	-0.03	-0.02	0.45	0.17	0.19	0.28	-0.16	0.17
Si	-0.03	0.10	0.89	0.03	-0.13	-0.21	-0.10	0.09	-0.23	0.86
Na	-0.15	-0.04	-0.03	0.90	0.04	0.23	0.17	-0.12	-0.36	0.15
SO_4	-0.11	0.98	-0.04	-0.04	0.02	-0.02	0.98	-0.12	-0.02	-0.02
S	-0.11	0.97	-0.01	-0.05	0.03	-0.01	0.97	-0.13	0.01	-0.02
Ti	-0.02	0.00	0.82	-0.07	0.09	0.25	0.17	-0.19	0.49	0.40
V	0.53	0.16	-0.09	0.08	0.28	-0.04	0.02	0.06	0.67	-0.05
Zn	0.33	-0.02	0.21	0.41	-0.14	0.69	-0.07	0.34	0.06	0.03
CO	0.73	0.07	0.01	0.06	-0.02	0.65	0.01	0.24	-0.13	0.11
O_3	-0.70	0.50	0.21	0.00	-0.23	-0.13	0.53	-0.66	0.04	0.08
NO_2	0.79	0.07	0.18	-0.02	-0.14	0.37	0.17	0.59	0.01	0.07
SO_2	0.82	0.02	0.04	0.10	-0.04	-0.22	0.39	0.51	0.11	0.03

Table 4. Source factor loadings from principal components analysis for Boston, MA and Pittsburgh, PA (bold indicates factor loadings over 0.50)

	Memphis						Birmingham					
						Industrial/						
	Coal	Unknown	Traffic Related	Crustal	Salt	Salt	Coal	Crustal	Traffic Related	Unknown		
Al	-0.23	0.55	0.05	0.54	0.19	-0.20	-0.02	0.96	-0.20	0.02		
NH4	0.92	0.04	0.01	-0.18	-0.02	-0.15	0.82	-0.09	0.30	0.20		
Br	0.34	0.00	0.42	-0.10	0.50	0.06	-0.04	0.01	0.63	0.33		
Ca	0.08	-0.03	0.77	0.28	0.19	0.47	0.19	0.32	0.31	0.06		
Cl	-0.07	-0.09	0.01	-0.12	0.70	0.55	-0.28	0.00	0.32	0.09		
Cu	-0.01	0.72	0.07	-0.23	-0.11	0.37	-0.05	0.12	0.18	0.53		
EC	0.19	0.03	0.81	-0.18	-0.05	0.03	0.12	0.15	0.80	-0.05		
Fe	0.08	0.36	0.74	0.21	0.04	0.73	0.05	0.32	0.09	0.21		
Mn	0.26	0.41	0.29	0.11	-0.16	0.89	-0.02	0.07	0.04	0.13		
Ni	0.00	0.00	0.11	0.06	-0.04	0.01	0.05	0.14	-0.20	0.58		
NO3	0.09	0.20	-0.04	-0.60	0.14	-0.17	-0.27	-0.13	0.34	0.39		
OC	0.45	0.12	0.54	0.10	0.03	0.03	0.27	0.10	0.77	0.06		
Κ	-0.04	0.95	-0.21	-0.11	-0.11	-0.15	-0.12	0.61	0.45	-0.15		
Se	0.62	-0.08	-0.03	-0.07	-0.01	0.03	0.27	-0.08	0.52	-0.14		
Si	-0.18	0.35	0.28	0.64	0.23	0.32	0.09	0.74	0.05	0.12		
Na	-0.13	0.00	-0.15	0.08	0.78	0.66	0.01	-0.04	-0.13	-0.22		
SO4	0.94	0.07	0.00	0.00	-0.04	-0.02	0.92	0.00	0.15	0.09		
S	0.94	0.06	0.02	0.02	-0.02	0.00	0.92	0.02	0.16	0.09		
Ti	0.03	0.84	0.07	0.29	-0.05	0.11	0.11	0.83	0.00	0.03		
V	0.70	0.62	-0.07	-0.04	0.27	0.06	-0.13	0.26	0.37	-0.48		
Zn	0.25	0.23	0.45	-0.47	-0.10	0.93	-0.03	-0.01	-0.03	0.09		
CO	-0.20	0.00	0.63	-0.16	-0.16	-0.23	-0.07	-0.02	0.03	0.11		
O3	-0.04	-0.01	0.07	-0.02	-0.02	-0.05	0.08	-0.05	0.10	0.37		
SO2	0.44	-0.01	-0.18	0.54	-0.09	0.06	0.68	0.11	-0.17	-0.08		

Table 5. Source factor loadings from principal components analysis for Memphis, TN and Birmingham, AL (bold indicates factor loadings over 0.50)

		Milv	vaukee, WI	Detroit, MI						
	Industrial/			Salt/			Industrial/	1		
	Traffic Related	Coal	Unknown	Nitrate	Crustal	Coal	Salt	Crustal	Traffic	Unknown
Al	-0.02	-0.16	-0.02	-0.08	0.80	-0.17	0.03	0.88	-0.06	0.01
NH_4	0.10	0.84	0.00	0.33	0.02	0.90	0.13	0.02	0.11	-0.03
Br	0.15	0.30	0.12	-0.31	0.25	0.60	0.20	0.07	0.25	-0.03
Ca	0.54	0.27	0.07	-0.27	0.36	-0.07	0.24	0.55	0.30	0.27
Cl	-0.14	0.08	0.04	0.51	0.23	0.10	0.60	-0.17	0.01	-0.29
Cu	0.16	-0.02	0.73	-0.03	-0.12	0.11	0.02	-0.18	0.24	0.71
EC	0.61	0.26	0.23	0.10	0.07	0.01	0.06	0.09	0.79	0.15
Fe	0.82	0.15	0.12	-0.01	0.07	-0.16	0.72	0.33	0.20	0.21
Mn	0.82	-0.06	0.03	-0.06	0.11	-0.20	0.83	0.02	0.17	-0.03
Ni	0.41	0.07	0.02	0.05	-0.15	0.06	0.35	0.17	-0.30	0.09
NO_3	0.13	0.39	-0.06	0.72	0.02	0.44	0.40	-0.04	0.12	-0.41
OC	0.32	0.35	0.28	-0.08	0.12	0.33	0.04	0.05	0.55	0.29
Κ	-0.25	-0.11	0.95	0.07	-0.04	0.03	-0.02	0.09	0.04	0.66
Se	-0.04	0.59	-0.05	0.08	-0.04	0.70	-0.05	0.11	-0.02	-0.16
Si	0.38	0.01	-0.06	-0.15	0.70	0.10	0.11	0.83	0.09	-0.05
Na	-0.13	-0.03	-0.03	0.22	0.59	0.14	-0.11	0.47	-0.10	-0.17
SO_4	-0.02	0.88	0.12	-0.15	-0.02	0.88	0.01	0.02	0.07	0.15
S	0.02	0.89	0.14	-0.17	0.00	0.88	0.01	0.01	0.07	0.17
Ti	-0.02	0.01	0.82	-0.10	0.11	0.09	0.09	0.53	-0.01	0.30
V	0.01	0.06	0.12	0.10	0.32	0.12	0.37	-0.02	-0.11	0.31
Zn	0.70	0.11	0.06	0.08	0.18	-0.05	0.89	-0.02	0.08	0.05
CO	0.49	-0.16	0.16	0.34	-0.20	-0.10	0.09	-0.10	0.70	0.06
O ₃	0.60	0.10	0.14	0.43	0.10	0.27	-0.01	0.24	0.54	-0.20
NO_2	0.37	0.31	0.11	0.19	-0.11	0.17	-0.20	0.45	0.26	0.02
SO ₂	-0.26	0.39	0.15	-0.71	0.12	0.65	-0.16	0.12	-0.18	0.37

Table 6. Source factor loadings from principal components analysis for Milwaukee, WI and Detroit, MI (bold indicates factor loadings over 0.50)

			San Diego, CA				Riverside, CA				
	Crustal/					Coal/	Crustal/				
	Industrial	Coal	Traffic Related	Salt	Unknown	Residual Oil	Industrial	Traffic Related	Unknown	Salt	
Al	0.94	-0.08	-0.03	0.11	-0.11	-0.18	0.76	-0.16	0.32	0.16	
NH_4	-0.05	0.93	0.15	-0.13	-0.06	0.92	-0.19	0.06	-0.03	0.11	
Br	-0.03	0.55	0.62	0.24	-0.05	0.73	0.18	0.22	0.08	0.20	
Ca	0.96	-0.04	0.00	0.13	-0.10	-0.10	0.85	0.17	-0.01	-0.03	
Cl	-0.02	-0.18	0.26	0.82	0.03	0.15	-0.32	0.43	0.08	0.45	
Cu	0.12	-0.12	0.27	-0.03	0.39	-0.04	-0.14	0.02	100.0	-0.06	
EC	0.11	-0.09	0.73	-0.28	0.21	0.14	0.39	0.73	0.08	-0.11	
Fe	0.66	0.00	0.20	-0.28	0.34	0.11	0.83	0.29	0.06	-0.07	
Mn	0.78	0.12	-0.01	-0.09	-0.04	0.09	0.65	0.24	0.14	0.00	
Ni	0.04	0.37	-0.27	-0.13	0.59	0.67	0.12	-0.02	-0.08	-0.09	
NO_3	-0.07	0.76	0.41	-0.13	-0.08	0.88	-0.20	0.19	0.02	0.10	
OC	0.07	-0.04	0.82	-0.19	0.07	0.50	0.24	0.51	0.13	0.05	
Κ	0.30	-0.08	0.69	0.15	-0.04	-0.09	-0.18	-0.02	102.00	0.00	
Se	0.02	0.02	-0.14	0.22	0.33	-0.01	0.13	0.10	-0.04	0.65	
Si	0.92	0.00	0.01	-0.10	-0.01	-0.03	0.95	-0.08	0.01	0.06	
Na	0.05	0.06	-0.06	0.80	0.27	-0.02	0.10	-0.30	-0.02	0.72	
SO_4	0.06	0.90	-0.27	0.20	0.04	0.87	-0.08	-0.30	0.09	0.15	
S	0.04	0.89	-0.25	0.06	0.04	0.88	-0.06	-0.30	0.05	0.14	
Ti	0.72	0.00	0.07	-0.12	0.24	0.01	0.47	-0.02	0.71	0.02	
V	0.03	0.60	0.01	-0.03	0.39	0.70	0.18	-0.10	0.30	0.00	
Zn	0.12	0.23	0.55	-0.16	0.18	-0.06	0.26	0.58	0.09	0.02	
CO	0.03	-0.15	0.62	-0.32	0.21	0.19	0.25	0.74	0.02	-0.03	
O_3	0.15	0.00	0.46	-0.44	0.41	0.28	0.49	0.55	0.03	-0.11	
NO_2	-0.10	-0.05	0.14	0.12	0.48	0.21	0.16	-0.43	0.07	-0.16	
SO_2	0.00	0.33	-0.44	0.42	-0.30	0.30	0.08	-0.78	0.03	0.17	

Table 7. Source factor loadings from principal components analysis for San Diego, CA and Riverside, CA (bold indicates factor loadings over 0.50)