Thesis Title: Investigating the association between source-specific fine particulate matter and hospitalization due to myocardial infarction in New York City

Student (UNI): Rachel H. Tao (rht2112)

First Reader (UNI): Andrew G. Rundle, DrPH (agr3)

Second Reader (UNI): Marianthi-Anna Kioumourtzoglou, ScD (mk3961)

May 12, 2021

P9419 Master's Essay in Epidemiology II

Instructor (UNI): Dr. Larkin S. McReynolds (LSM34)

Table of Contents

Thesis Item **Page**

Supplemental Materials

PERMISSION TO USE MY ASSIGNMENT AS AN EXAMPLE FOR FUTURE CLASSES

___ I prefer not

___ You have my permission, but please make the file anonymous

_X_ You have my permission and I would like my name listed

Acknowledgments

I am so grateful to Dr. Marianthi-Anna Kioumourtzoglou, whose incredible guidance on this research made it possible for me to learn new methods and understand environmental epidemiology concepts in a deeper and more complete way. I would like to thank Dr. Andrew Rundle, whose insightful comments have brought this project to a new level and forced me to describe my thought process in a way that both improved the final product and solidified my own understanding of the work. This project would not have been possible without Dr. Amelia Boehme's permission to use data from the New York Department of Health Statewide Planning and Research Cooperative System (SPARCS). I want to thank Dr. McReynolds, who provided useful guidance on how to write a Master Thesis as instructor of the Master Essay course, as well as all my professors in this program who have improved my understanding of Epidemiology, Biostatistics and Data Science. I am grateful to Dr. Yanelli Nunez who provided invaluable advice and collaboration. I also want to thank Sebastian Rowland and Lawrence Chillrud for their contributions to data preprocessing, and for their collaboration throughout this project. I would like to thank all the members of makLab and my classmates in my program. Finally, I want to thank all the non-epidemiologists in my life who tolerated my thinking aloud as I tried to make sense of what I was learning, and for the encouragement of all those who have supported me during my MPH.

Abstract:

Background and Aims: The association between fine particulate matter ($PM_{2.5}$) air pollution and cardiovascular outcomes is well-established. $PM_{2.5}$ is a heterogeneous mixture of chemical constituents and its composition can vary by air pollution source. Different sources of $PM_{2.5}$ can be differentially associated with temperature, depending on the chemical composition of $PM_{2.5}$ produced by different sources and the chemical properties of these underlying chemical mixtures. To evaluate whether $PM_{2.5}$ from certain sources may be differentially associated with cardiovascular disease, we examined the association between same-day exposure to source-specific $PM_{2.5}$ and risk of hospital admission for myocardial infarction (MI) in New York City (NYC) and evaluated potential effect modification by same-day temperature.

Methods: We applied Absolute Principal Component Analysis to identify sources of $PM_{2.5}$ pollution using data from monitors in three different locations in NYC. We used data from the New York Department of Health Statewide Planning and Research Cooperative System on daily city-wide counts of MI admissions (2007–2015). We examined associations between same day exposure to source-specific $PM_{2.5}$ and MI admissions in a time-series analysis, using a quasi-Poisson regression model and adjusting for same-day temperature and relative humidity, lagged 3-day average temperature and relative humidity, day of week, and seasonal and long-term time trends. We then assessed for effect modification by temperature by categorizing temperature into quartiles; we used interaction terms between source-specific $PM_{2.5}$ and the temperature quartiles as indicators in the model.

Results: We identified six sources of $PM_{2.5}$ pollution: 1) nitrate, 2) salt, 3) crustal dust, 4) secondary/regional sulfate, 5) traffic and road dust, and 6) industrial emissions. In adjusted models, we observed a 0.96% (95% confidence interval [CI]: -0.18, 2.11%) increase in MI rates

per one interquartile range (IQR) increase in $PM_{2.5}$ from nitrate sources, a 0.43% (95% CI: -0.13, 0.99%) increase in MI rates per one IQR increase in crustal $PM_{2.5}$, and a 0.35% (95% CI: -0.21, 0.91%) increase in MI rates per one IQR increase in industrial-related $PM_{2.5}$, on average. We observed effect modification by temperature in the crustal $PM_{2.5}$ – MI association. We observed a -0.80% (95% CI: -2.03, 0.45%), 1.10% (95% CI: -0.14, 2.23%), 0.71% (95% CI: -0.45, 1.79%) change in MI rate per one IQR increase in crustal dust at the lowest, second, and third quartile of temperature, respectively; the association between crustal $PM_{2.5}$ and MI at the highest quartile of temperature was null.

Conclusions: Identifying particularly toxic sources of $PM_{2.5}$ can maximize efficiency in air pollution policies. In our NYC study we identified nitrate, crustal dust, and industrial $PM_{2.5}$ as potentially toxic sources for cardiovascular disease. We also observed non-significant differences in effect estimates for the association between crustal dust $PM_{2.5}$ and MI by quartile of temperature, which may be attributable to compositional variation in crustal dust by temperature, or exposure measurement error resulting from temperature-related patterns in indoor-outdoor ventilation.

Introduction

Globally, ischemic heart disease was the leading cause of death in both 2000 and 2019, and is responsible for an increase of more than 2 million deaths over the last two decades (1). An estimated 6.67 million deaths worldwide in 2019 were attributable to air pollution, approximately half of which were due to cardiovascular disease (2).

The association between exposure to particulate matter air pollution and cardiovascular outcomes is well-established, with increasing evidence that fine particulate matter ($PM_{2.5}$) is particularly harmful (3–5). Inhalation of particulate matter can cause cardiovascular disease through several biological pathways, including 1) oxidative stress and inflammation, 2) neural reflex arcs and autonomic imbalance, leading to heart rhythm perturbation, 3) increased blood pressure, mediated by effects of $PM_{2.5}$ exposure on systemic vasculature, and 4) translocation of inhaled $PM_{2.5}$ constituents, such as organic compounds or metals, into systemic circulation (6,7) (Figure 1). These pathways can lead to both chronic cardiovascular disease and acute cardiovascular events (6,7). The relationship between $PM_{2.5}$ exposure and acute cardiovascular events, including hospitalizations and deaths, has been assessed primarily through time-series and case-crossover studies (5,8–10). A recent meta-analysis of twenty-six published studies found that each 10 $\mu g/m^3$ increase in $PM_{2.5}$ was associated with 1.02 times higher risk of myocardial infarction (MI) (95% confidence interval [CI]: 1.01, 1.03) (5).

$PM_{2.5}$ is a heterogeneous mixture of solid particles and liquid droplets, and its chemical composition varies depending on the source of the pollution (11–13). Health effects can therefore also vary depending on which sources contribute to $PM_{2.5}$ pollution in a given region (14–17). Better understanding of which sources of $PM_{2.5}$ pollution are most strongly associated with

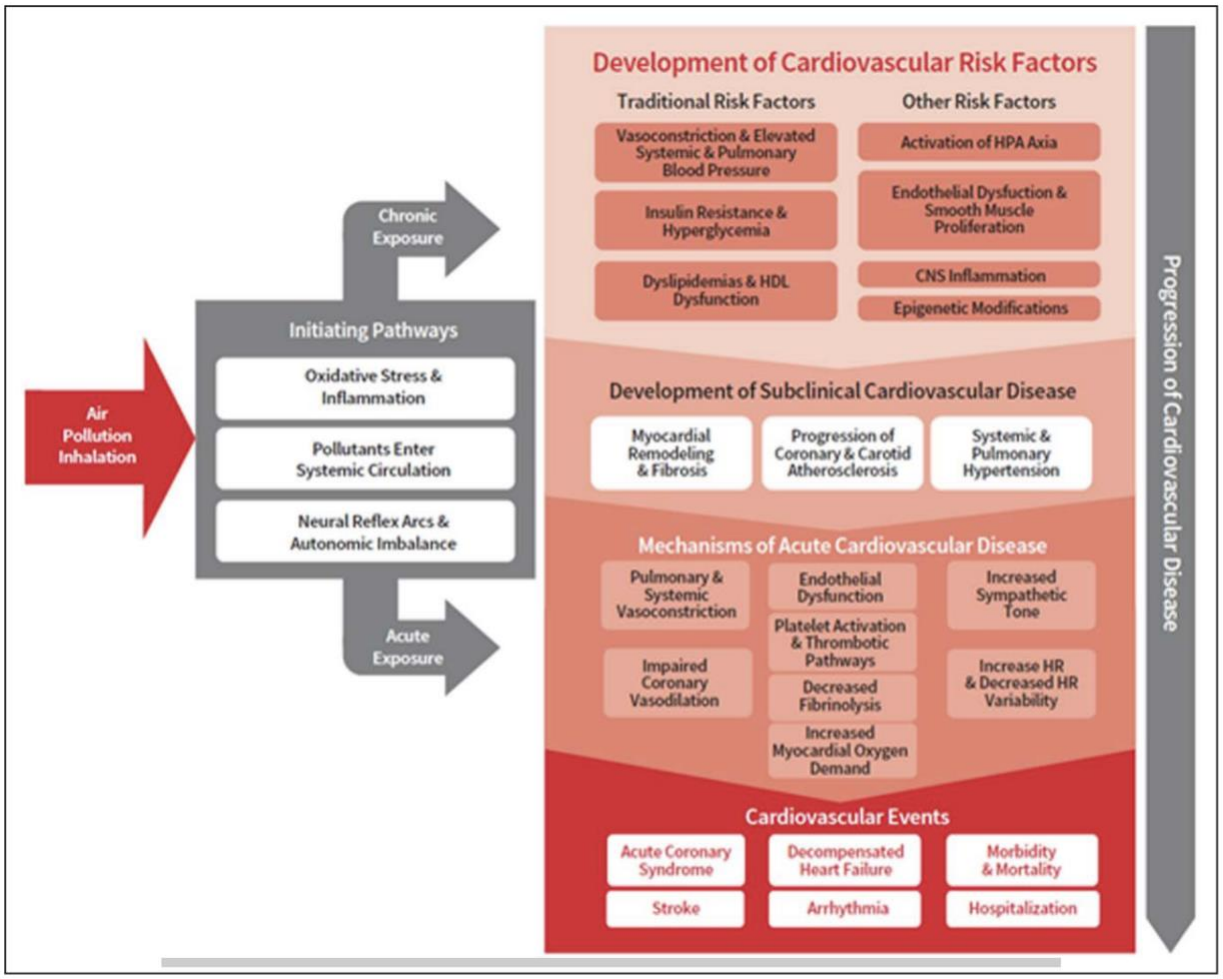


Figure 1. Biological pathways linking air pollution with cardiovascular disease. Mechanisms of cardiovascular disease attributable to air pollution exposures (Adapted from Rajesh Vedanthan and Michael Hadley, 2019 in Brauer et al. 2021)

cardiovascular risk could inform local, regional, and national policy by identifying particularly toxic sources as targets for $PM_{2.5}$ reduction efforts.

Source apportionment is the process of identifying sources of particle pollution using particle compositional data. Source apportionment methods usually describe the variability among pollutants in terms of unobserved variables derived from dimensionality-reduction methods such as Principal Component Analysis (PCA) (18–20). These new variables can be used in subsequent health models, such as time-series analyses, to assess the potential association between specific sources of pollution and adverse health outcomes (13,14,20). For this study we

used Absolute Principal Component Analysis (APCA), an extension of PCA where after component scores are calculated, total $PM_{2.5}$ concentrations are regressed on component scores to estimate the contribution of each source (18). The primary difference between APCA and PCA is the use of linear regression on component scores to compute daily source concentrations in the same units as measured $PM_{2.5}$ and its chemical constituents. APCA improves interpretability by preserving units of measurement and allows for comparison with other source apportionment studies, as it has been extensively used in $PM_{2.5}$ source apportionment analyses (13,17,21,22).

Several studies have explored the relationships between different sources of $PM_{2.5}$ pollution and cardiovascular hospital admissions (12,14,17). The results of previous studies of cities on the East Coast of the United States suggest that $PM_{2.5}$ from residual oil and traffic pollution may be associated with cardiovascular hospital admissions (14,17). A recent study examining source-specific $PM_{2.5}$ and cardiovascular hospital admissions across New York State (NYS) suggested that traffic-related $PM_{2.5}$ may be associated with MI (23). To our knowledge, source-specific analysis of the association between $PM_{2.5}$ and cardiovascular risk in New York City (NYC) has not been conducted since a 2011 paper (14), which used data from 2001-2002. Since 2002, several policies aimed at improving air quality have been implemented on the federal, state, and local levels, including those aimed at reducing sulfur emissions from diesel fuel, as well as changes in the regulations governing electricity generation. Reduction in electricity generation from coal-fired power plants has led to decreases in $PM_{2.5}$ in NYS (24). These changes and others may have changed the prevalence of different pollution sources in NYC or altered the chemical composition of $PM_{2.5}$ coming from different air pollution sources. Here, we provide an updated analysis of source-specific effects of $PM_{2.5}$ on risk of

cardiovascular events in NYC. By including data from the years 2007 to 2015, we also capture long-term time trends that were outside the range of the previous NYC study.

Different sources of $PM_{2.5}$ can be differentially associated with temperature, depending on the chemical composition of $PM_{2.5}$ produced by different sources and the chemical properties of these underlying chemical mixtures. For example, secondary nitrate has been found to have highest concentrations during the colder months because lower temperatures are more favorable for ammonium nitrate formation (25,26). Heat exposure is associated with MI (27) and could have a synergistic relationship with source-specific $PM_{2.5}$ and MI. It is therefore important to determine whether there is evidence of effect modification by temperature.

We used APCA to identify sources of $PM_{2.5}$ pollution using speciated $PM_{2.5}$ data from three locations in NYC (Figure 2). Once sources of $PM_{2.5}$ were identified, we examined associations between sources of $PM_{2.5}$ and MI admissions in NYC from 2007 to 2015. Identified sources were included as the exposure of interest in a time-series analysis using a quasi-Poisson regression model with MI admissions as the outcome measure. We *a priori* expected that traffic-related $PM_{2.5}$ would be positively associated with hospitalization due to MI. We aimed to assess the association between source-specific $PM_{2.5}$ and MI using more recent data than the existing literature on this topic, by leveraging the New York Department of Health Statewide Planning and Research Cooperative System (SPARCS) for data on Emergency Department visits for MI in NYC. We also examined the potentially modifying role of temperature in the relationship between source-specific $PM_{2.5}$ and MI.

Methods

Study Population:

Data source: Daily hospitalization data for MI in NYC were extracted from the SPARCS dataset (28). We used daily city-wide counts of MI as the dependent variable.

The study population consists of people who received care for MI in acute care facilities in NYC, 2007-2015. The population of interest for this study is the population of NYC, 2007-2015, and the SPARCS dataset was used to assess MI in this population. Since the SPARCS data represent inpatient and outpatient admissions at acute care facilities, our study population excludes 1) those who had MI but were deceased before receiving medical care, 2) those who did not seek medical care after experiencing MI symptoms and recovered, 3) those who sought care in non-acute care facilities, and 4) those who had MI in NYC but sought care outside of NYC (28).

Outcome Assessment:

MI admissions were identified based on ICD, 9th Revision (ICD-9) for years prior to 2015 and based on ICD 10th Revision (ICD-10) for 2015. Admissions were identified as cases if ICD-9 code 410.x1 or ICD-10 code I21 occupied one of the first four diagnostic positions. We additionally excluded "childbirth" or "trauma" admission types. Reinfarctions and recurrent MI admissions were included, except readmissions that took place within two days after a previous MI admission for that patient. Observations with a missing date of admission (n = 1035) were excluded for a final sample size of 444,295 MI admissions.

Exposure Assessment:

Data Source: Air pollution data for NYC were extracted from the Air Quality System (AQS) dataset collected and maintained by the United States Environmental Protection Agency (EPA) (29). This dataset is publicly accessible and includes ambient concentrations of a selection of pollutants. Samples for speciated $PM_{2.5}$ were collected every third or sixth day and concentrations are reported in micrograms per cubic meter ($\mu g/m^3$).

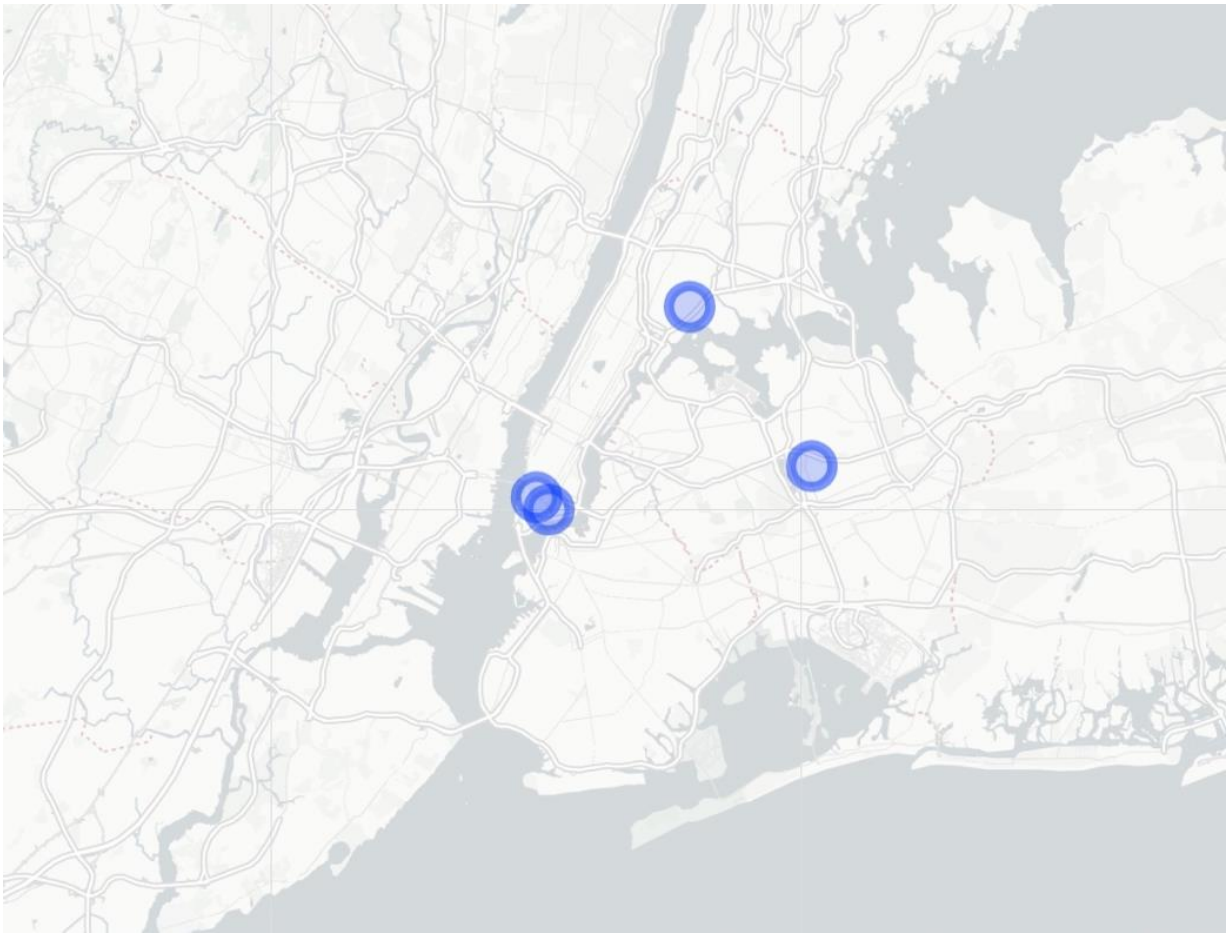


Figure 2. Map of locations of AQS monitors used in this study. Blue dots represent locations of each monitor.

From the AQS dataset, we used data from three locations in NYC: lower Manhattan, southern Bronx, and Queens (Figure 2). Data from two separate monitors in lower Manhattan were used because they were in the same geographic area, and during years when data from one

of these monitors were unavailable, data from the other were available. We used data from the three monitor locations to compute the daily average values for total $PM_{2.5}$ and chemical constituents of $PM_{2.5}$ across NYC. Since daily concentrations were not available from all monitors for every day included in the dataset, average concentrations were calculated for each day from whichever monitors had available data. Inclusion of days where not all monitors had available data allowed for computation of city-wide averages during periods when there were gaps in data availability for certain monitors (Figure S1). We used average daily concentrations of total $PM_{2.5}$ and the following constituents of $PM_{2.5}$: aluminum, ammonium, arsenic, barium, bromine, calcium, cadmium, elemental carbon, organic carbon, chlorine, chromium, copper, iron, lead, magnesium, manganese, nickel, selenium, silicon, sodium, sulfur, titanium, nitrate, vanadium, and zinc.

Covariates:

Data Source: Temperature, pressure, and specific humidity data were extracted from the National American Land Data Assimilation System, NLDAS-2 Forcing (30). NLDAS reports hourly parameter values for 0.125° grids (~11 km × 14 km in NYS).

Since the unit of analysis for our study was days, we took the average of the 24 hours provided in the NLDAS dataset for each day. Since our analysis was on the city-level, and the parameter values from NLDAS are reported in 11 km × 14 km grids, we aggregated daily values for each grid to the geographic extent of NYC via population-weighted averaging at the Census tract level. To accomplish this, we 1) intersected the census tract-level population data from the 2010 US Census with the NLDAS grids, 2) calculated weights based on estimated population in each intersection and 3) multiplied these weights by the NLDAS daily estimates. We used

population-weighting for meteorological variables because these variables are included in the final model as potential confounders of the relationship between source-specific $PM_{2.5}$ and MI, so daily estimates should be as reflective as possible of the actual weather experienced by the average person in NYC. Relative humidity (RH) was calculated from specific humidity, pressure, and temperature.

Statistical Analysis:

Source apportionment:

APCA was used to identify patterns among $PM_{2.5}$ constituents, and a combination of expert knowledge and prior literature was used to identify air pollution sources from the identified patterns. APCA is an extension of PCA, which estimates new variables (called principal components) to explain the total variance using fewer variables than the original number. After PCA is conducted, component scores are rescaled relative to a reference of zero concentration. To rescale component scores, we first calculated the z-score for a reference day for each component by subtracting the mean score of the component from zero and dividing the result by the standard deviation of the component scorers. We then subtracted this value from each component score, to compute the absolute component score (21). Finally, to estimate the mass contribution of each source to chemical constituents of $PM_{2.5}$ and total $PM_{2.5}$ concentration, we regressed $PM_{2.5}$ concentrations on absolute component scores (18). We do this so that when we assess patterns in each component to identify sources of $PM_{2.5}$ pollution, we are looking at source-specific $PM_{2.5}$ and its chemical constituents in the same units of measurement that the original concentrations were measured in. Preserving units of measurement improves interpretability at the point in the process where expert knowledge and prior literature are applied

to identify sources from components. Since there is some degree of subjectivity in this source-identification step, it is important for the inputs to be interpretable, both for improving accuracy in source identification and for ease of comparability with other source apportionment studies. Once sources are identified from components, each component from APCA (now labeled as a source) is evaluated as a predictor in the health model, described in detail below.

For our analysis, we used average daily concentrations for total $PM_{2.5}$ and its chemical constituents, as described in the *Exposure Assessment* section, treating each day as an observation. To identify sources from the principal components, we examined the mass contributions of each component to each of the chemical constituents of $PM_{2.5}$ and to total $PM_{2.5}$ and used a combination of prior literature and expert knowledge to match components to known pollution sources. We excluded dates around 4th of July (7/2 – 7/6) for each year as outlying events due to increased fireworks during this time, as is commonly done in source apportionment studies (12,14).

Time-series health analysis:

Once sources of $PM_{2.5}$ were identified, we conducted a time-series analysis using a Poisson regression model, using quasi-likelihood to account for potential overdispersion in the outcome. To determine which pollution sources are associated with MI, all source contributions from the identified sources were simultaneously included in the regression model as predictors. To test the linearity assumption of our model, we modeled each identified source using a penalized spline in a generalized additive model, while controlling for linear terms of all other sources, along with all covariates. Quasi Akaike's Information Criterion (qAIC) was used to determine whether each source should be modeled linearly or nonlinearly in the final model. We

controlled for several covariates as potential confounders associated with both source-specific $PM_{2.5}$ and MI including: nonlinear terms (natural splines) for same-day (lag 0) ambient temperature (degrees of freedom, df =3), 3-day average (average lag 1-3) ambient temperature (df = 4), same-day relative humidity (df = 4), and 3-day average relative humidity (average lag 1-3, df = 3), dummy variables for day of the week, and a natural spline for seasonal and long-term trends (df = 4 × number of years). We separately controlled for same-day and 3-day average ambient temperature and relative humidity to account for the effects of both heat and cold exposure (27). We assessed potentially nonlinear exposure-response curves using penalized splines in generalized additive models.

To evaluate the potential modifying role of temperature on the association between source-specific $PM_{2.5}$ and MI, we categorized temperature into quartiles, and included interaction terms with indicators for the temperature quartiles in the models. We ran separate models for each source, which included the main effect for the source, the main effect for same-day temperature (modeled non-linearly, df = 3), and an interaction term between the source and categorized temperature, controlling for all other sources and covariates. We used partial F-tests to test for evidence of effect modification at the 5% level of significance. Effect estimates at each quartile of temperature were computed for sources where the partial F-test was statistically significant at the 10% level of significance.

Effect estimates are presented as percent changes in MI hospital admission rate per interquartile range (IQR) increase in source-specific $PM_{2.5}$ if the associations were linear. If nonlinear, we present the full exposure-response curve.

Statistical analyses were conducted using R version 4.0.2 (2020-06-22). Source apportionment code can be publicly accessed via https://github.com/12taor/PM2.5_sources.

Sensitivity Analysis:

To examine the extent to which nonlinearity was attributable to outliers, we removed outliers and assessed linearity using penalized splines in a generalized additive model, as in the main model. Values that were more than 3 standard deviations away from the mean for each source were examined as potential outliers. To evaluate the influence of outliers in linear models, we modeled all sources linearly in a single model adjusting for all other sources and all covariates, and compared the effect estimates for each source when outliers were and were not included.

Results

During the study period (2007-2015), the daily average total $PM_{2.5}$ concentration was 10.32 $\mu g/m^3$ (SD: 5.97), and the daily median number of admissions for MI was 135 (IQR: 27) over the 978 days included in our analysis. MI admissions decreased overall from 2013 to 2015 and followed a seasonal pattern with highest MI rates tending to occur in the winter (Figure S2). Average ambient temperature was 12.33 °C (SD: 9.67) (Table 1). Correlations between $PM_{2.5}$ constituents ranged from -0.1 to 0.9, with highest correlations between ammonium and sulfur, ammonium and nitrate, aluminum and silicon, chlorine and sodium, magnesium and sodium, manganese and nickel, and OC and sulfur (Figure S3).

Table 1. Summary statistics for source-specific $PM_{2.5}$, total $PM_{2.5}$ concentration, MI admission count, and same-day temperature (°C). Source-specific and total $PM_{2.5}$ concentration reported in $\mu g/m^3$ (n = 978).

	Min	25th%ile	Median	75th%ile	Max	Mean (SD)
MI admission count	10	121	135	148	216	135.2 (19.6)
Nitrate ($\mu g/m^3$)	-1.03	0.11	0.59	1.37	11.11	0.90 (1.20)
Salt ($\mu g/m^3$)	-0.09	0.02	0.09	0.22	3.84	0.17 (0.28)
Crustal Dust ($\mu g/m^3$)	-1.52	0.61	1.1	1.71	18.32	1.39 (1.57)
Secondary Sulfate ($\mu g/m^3$)	-15.00	1.26	2.96	5.88	40.44	4.26 (4.98)
Traffic + Road dust ($\mu g/m^3$)	-2.07	1.1	1.77	2.49	9.97	1.87 (1.25)
Industrial ($\mu g/m^3$)	-1.36	0.06	0.12	0.19	2.34	0.13 (0.16)
Total $PM_{2.5}$ ($\mu g/m^3$)	1.62	5.87	8.83	13.03	38.58	10.32 (5.97)
Same-day temperature (°C)	-14.02	4.31	12.67	21.16	31.49	12.33 (9.67)

Table 2. Left panel: average concentration of each chemical constituent per identified source. Right panel: actual mean concentration of each chemical constituent, predicted mean concentration of each chemical constituent using APCA, and percent error of predicted concentrations. All concentrations reported in (ng/m^3).

Chemical Constituent	Nitrate	Salt	Crustal Dust	Secondary Sulfate	Traffic + Road dust	Industrial	Mean Concentration	Predicted Concentration	Percent Error
Aluminum	0.13	0.20	14.69	4.47	5.33	1.68	22.18	26.51	19.51%
Ammonium	205.57	9.02	153.02	799.87	-154.09	40.46	1120.78	1053.86	-5.97%
Arsenic	0.07	0.01	0.03	0.15	0.14	0.00	0.49	0.40	-17.98%
Barium	0.07	0.16	0.52	0.11	2.30	-0.06	1.87	3.10	65.68%
Bromine	0.51	0.31	0.14	0.92	1.10	-0.32	3.06	2.66	-13.12%
Cadmium	0.12	-0.07	0.09	-0.07	0.13	-0.17	1.68	0.03	-98.37%
Calcium	16.07	2.49	13.58	5.46	3.11	0.07	51.85	40.78	-21.35%
Chlorine	11.45	42.81	-4.87	5.40	5.83	-2.02	37.07	58.58	58.04%
Chromium	0.50	-0.09	0.32	-0.46	-0.88	3.19	2.13	2.58	20.89%
Copper	1.07	-0.07	0.63	0.99	2.33	0.69	4.55	5.64	23.87%
Elemental carbon	145.70	-3.45	98.42	133.30	253.47	68.64	707.73	696.09	-1.64%
Iron	11.02	-0.32	24.81	13.89	33.86	25.18	105.50	108.45	2.79%
Lead	0.88	0.04	0.05	0.56	1.16	-0.41	2.02	2.28	13.15%
Magnesium	-0.81	6.30	0.47	-0.02	2.20	-0.30	7.12	7.84	10.22%
Manganese	0.71	0.01	0.40	0.44	0.34	0.38	2.10	2.28	8.74%
Nickel	2.49	0.13	0.34	0.78	-1.73	1.55	4.94	3.55	-28.14%
Nitrate	608.43	101.61	0.31	670.44	103.38	66.42	1613.47	1550.59	-3.90%
Organic carbon	64.84	-46.27	346.91	800.00	882.38	20.93	2693.20	2068.79	-23.18%
Potassium	6.59	3.98	2.97	14.05	20.54	-4.98	36.15	43.14	19.34%
Selenium	0.05	0.01	0.05	0.22	0.01	-0.01	0.40	0.33	-15.71%
Silicon	3.59	1.09	46.19	8.85	7.53	-3.03	61.01	64.22	5.26%
Sodium	5.13	59.40	7.82	4.95	-8.24	5.69	95.65	74.77	-21.83%
Sulfur	-20.40	2.28	155.69	477.01	-10.82	10.88	788.05	614.64	-22.01%
Titanium	0.02	-0.02	1.23	0.44	0.86	0.12	2.37	2.65	11.89%
Vanadium	0.76	0.11	0.72	1.38	0.33	0.32	2.86	3.61	26.55%
Zinc	15.40	0.54	-1.79	3.20	7.76	-3.75	26.09	21.36	-18.14%
Total $PM_{2.5}$	*901.08*	*170.76*	*1386.50*	*4257.78*	*1867.95*	*129.58*	*10322.11*	*8713.65*	*-15.58%*

Source Apportionment

Six sources of $PM_{2.5}$ were identified based on average concentration of each chemical constituent per absolute principal component: 1) nitrate, 2) salt, 3) crustal dust, 4) regional/secondary sulfate, 5) traffic and road dust, and 6) industrial. The component characterized by high levels nitrate and ammonium, as well as elemental carbon, nickel, zinc, calcium, lead, and manganese, was identified as nitrate from local or regional sources, and accounts for 10.34% of total predicted $PM_{2.5}$. The component primarily composed of chlorine, sodium, and magnesium was identified as salt and constituted 1.96% of total $PM_{2.5}$ concentration. The component characterized by high levels of silicon and aluminum, along with moderately high levels of titanium and calcium, was identified as crustal dust, and accounts for 15.91% of total predicted $PM_{2.5}$ concentration. The majority of sulfate appeared to load on a single component, which accounted for 48.86% of total $PM_{2.5}$ and was identified as secondary/regional sulfate pollution. This component also included moderately high levels of potassium, selenium, organic carbon and vanadium. A component characterized by a diverse mixture of organic carbon, barium, lead, nitrate, potassium, copper, zinc, titanium, arsenic, and iron was identified as traffic and road dust, and accounted for 21.44% of total $PM_{2.5}$. A component consisting of high levels of chromium, nickel, and iron, as well as copper, elemental carbon and vanadium was identified as $PM_{2.5}$ resulting from industrial emissions. This component constituted 1.49% of total $PM_{2.5}$ (Table 2, Figure S4).

Comparing the predicted concentration of each constituent from APCA to the actual concentrations, percent error for total $PM_{2.5}$ was -15.58%. Predicted average cadmium concentration should be interpreted with caution, as percent error was -98.37%. Error was also greater than 25% for barium (65.68%), chlorine (58.04%), nickel (-28.14%), and vanadium

(26.55%) concentrations (Table 2). As expected given orthogonal rotation of APCA components, APCA scores for each component were not correlated with one another—the highest correlation between two components was 0.13, between the traffic and road dust component and the industrial component (Figure S5). The component for regional/secondary sulfate was highly correlated with total $PM_{2.5}$ concentration ($r = 0.83$).

We examined the seasonal, long-term, and weekly patterns in predicted concentration for each detected source of $PM_{2.5}$ (Figs. S6, S7). Nitrate appeared to have higher concentrations on weekdays and tended to reach peak concentrations during winter and lowest concentrations during summer. Nitrate also appeared to decrease slightly during the study period 2007-2015. Salt did not appear to show weekly patterns but seems to have a seasonal cycle with peaks in the spring and troughs in the autumn. There were particularly high concentrations of salt in 2014 and 2015, but otherwise concentrations remained constant over the study period. Other than a plateau in 2012, crustal dust concentrations appeared to peak in the summer. Crustal dust does not have strong weekly patterns, though weekdays may have slightly higher concentrations than weekends. Regional/secondary sulfate decreased over the course of the study period and did not follow a clear seasonal or week pattern. Traffic and road dust appeared to increase over the course of the study period. Concentrations seemed to increase during the summer and decrease during the winter, and concentrations may be slightly higher on weekdays than weekends. Industrial $PM_{2.5}$ showed strong weekly trends, with concentrations higher on weekdays than weekends. Industrial $PM_{2.5}$ concentrations did not appear to have clear seasonal trends and seem to have slightly decreased over the course of the study period.

Time-series health analysis:

We observed a 0.96% (95% CI: -0.18, 2.11%) increase in MI rates per one IQR increase in $PM_{2.5}$ from nitrate sources, a 0.43% (95% CI: -0.13, 0.99%) increase in MI rates per one IQR increase in crustal $PM_{2.5}$, and a 0.35% (95% CI: -0.21, 0.91%) increase in MI rates per one IQR increase in industrial-related $PM_{2.5}$, on average, adjusting for covariates (Figure 3). For all other sources (salt, regional/secondary, and traffic/road dust), the association was null (Table 3). For all sources, qAIC was larger for the model including the non-linear term compared to the linear term. Therefore, all sources were included as linear terms in the final model.

A sensitivity analysis removing outliers confirmed that it was appropriate to model all sources as linear terms, as qAIC for the model including all sources as linear terms was less than or equal to qAIC for models with each source modeled non-linearly. When the qAIC for the non-linearly modeled source was equal to the qAIC for the model with all sources linear, the estimated degrees of freedom for the penalized spline of the non-linearly modeled source were examined. For all sources for which this was the case (salt, crustal dust, traffic/road dust), estimated degrees of freedom (edf) = 1.001, and we therefore concluded these sources should be modeled linearly. When outliers were removed and all sources were modeled linearly, the percent change in MI rate per one IQR increase in crustal dust and industrial $PM_{2.5}$ increased slightly to 0.37% (95%: -0.83, 1.59) and 0.50% (95% CI: -0.67, 1.69), respectively. The point estimate for the percent change in MI rate per one IQR increase in nitrate $PM_{2.5}$ increased from 0.96% (95% CI: -0.18, 2.11) to 1.79% (95% CI: -0.20, 3.81) (Figure S8).

Table 3. Percent change in MI rate for each IQR increase in source-specific $PM_{2.5}$, adjusting for same-day and 3-day average temperature, same-day and 3-day average relative humidity, day of the week, and seasonal and long-term trends.

Source	Percent change in MI Rate (95% CI)
Nitrate	0.96 (-0.18, 2.11)
Salt	0.01 (-0.48, 0.49)
Crustal Dust	0.43 (-0.13, 0.99)
Regional/Secondary	0.26 (-0.47, 1.01)
Traffic + Road Dust	0.07 (-0.90, 1.05)
Industrial	0.35 (-0.21, 0.91)

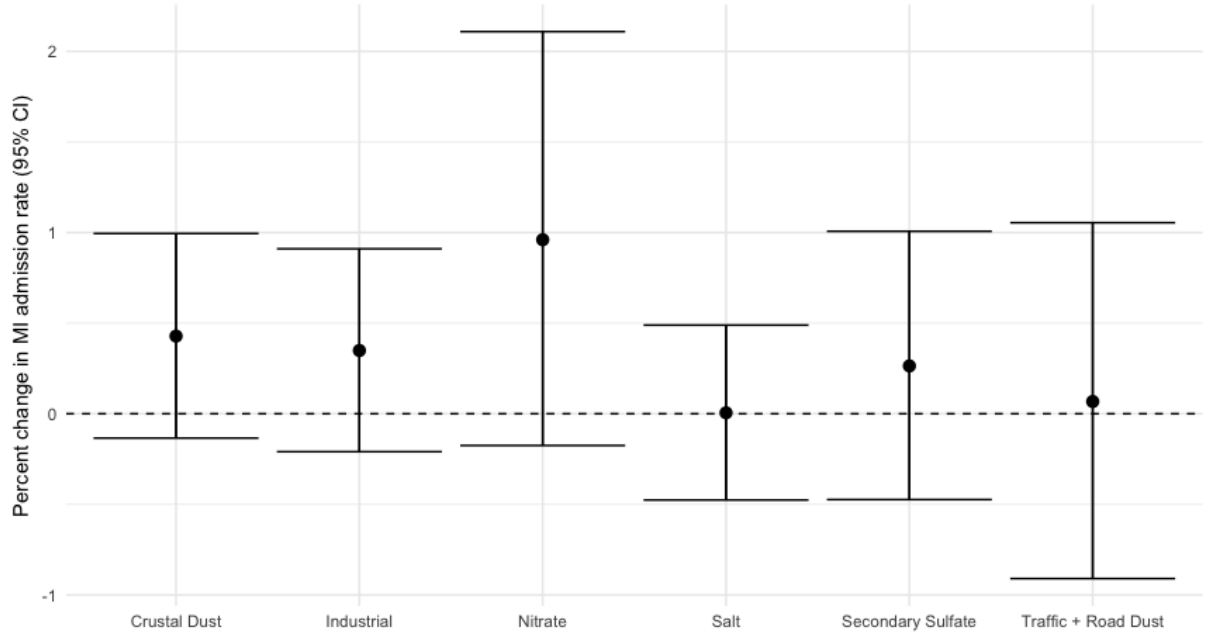


Figure 3. Forest plot of percent change in rate of MI admission for an IQR increase in source-specific $PM_{2.5}$, adjusting for same-day and 3-day average temperature, same-day and 3-day average relative humidity, day of the week, and seasonal and long-term trends.

Based on partial F-tests comparing models with an interaction term to the model with no interaction term, none of the interaction terms between source-specific $PM_{2.5}$ and temperature were statistically significant at the 5% level of significance. The interaction term between crustal

dust and same-day ambient temperature was statistically significant at the 10% level of statistical significance ($F_3 = 2.0795$, $p = 0.0991$), and effect estimates at each quartile of ambient temperature were therefore computed for this source. The percent change in MI rate per one IQR increase in crustal dust appeared to be negative for the lowest quartile of ambient temperature at -0.80% (95% CI: -2.03, 0.45%), and positive for the second and third temperature quartiles at 1.10% (-0.14, 2.23%) and 0.71% (-0.47, 1.79%), respectively, though none of these relationships were statistically significant at the 5% level of significance (Table 4, Figure 4). The relationship between crustal dust and MI admission rate at the highest ambient temperature quartile was null (Table 4, Figure 4).

Table 4. Percent change in MI rate for each IQR increase in crustal $PM_{2.5}$ at each quartile of same-day ambient temperature, adjusting all other sources of $PM_{2.5}$, 3-day average temperature, same-day and 3-day average relative humidity, day of the week, and seasonal and long-term trends.

Quartile of Temperature	Percent change in MI Rate per IQR increase in crustal dust (95% CI)
0^{th}-25^{th}%ile	-0.80 (-2.03, 0.45)
25^{th}-50^{th}%ile	1.10 (-0.14, 2.23)
50^{th}-75^{th}%ile	0.71 (-0.47, 1.79)
75^{th}-100^{th}%ile	0.52 (-2.06, 2.92)

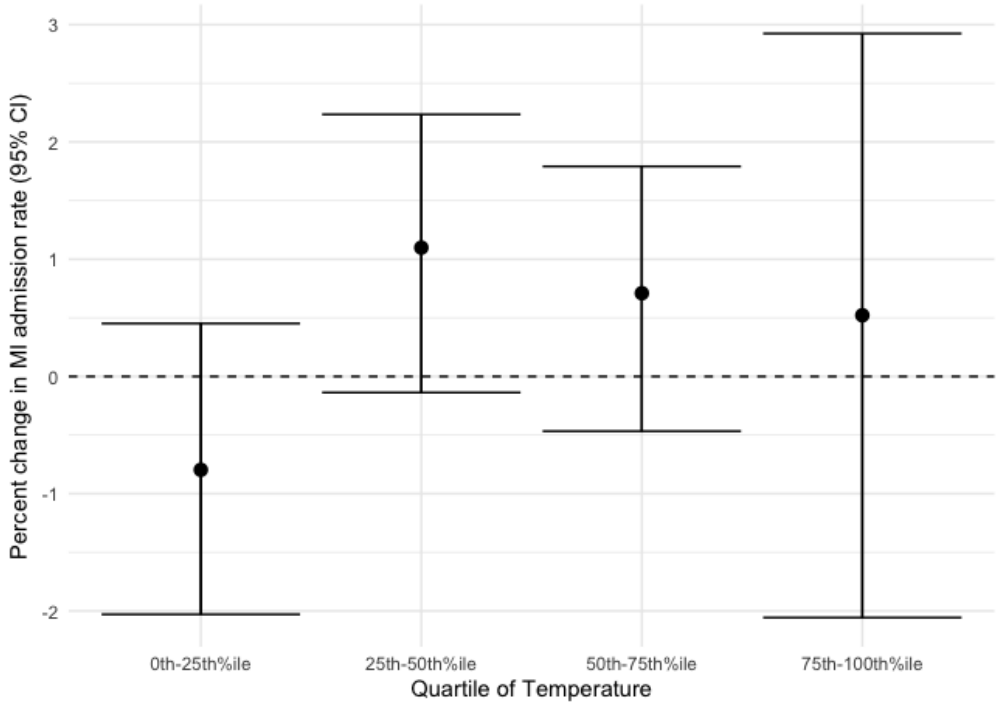


Figure 4. Forest plot of percent change in rate of MI admission for an IQR increase in crustal $PM_{2.5}$ at each quartile of same-day ambient temperature, adjusting for all other sources of $PM_{2.5}$, 3-day average temperature, same-day and 3-day average relative humidity, day of the week, and seasonal and long-term trends.

Discussion

Using data from the EPA's publicly available AQS database, we identified six sources of $PM_{2.5}$ pollution: 1) nitrate, 2) salt, 3) crustal dust, 4) secondary/regional sulfate, 5) traffic and road dust, and 6) industrial. Leveraging data from SPARCS, we observed increased rates of MI admission with increased concentrations of nitrate, crustal dust, and industrial $PM_{2.5}$, but not for salt, secondary/regional sulfate, or traffic/road dust (Table 3, Figure 3). Although the observed associations were not statistically significant at the 5% level of significance, we consider these associations to be clinically meaningful based on the range of the 95% confidence intervals. Our findings were inconsistent with our hypothesis that traffic-related $PM_{2.5}$ would be positively associated with MI admission rate. The relationship between crustal $PM_{2.5}$ and MI admission rates appeared to be negative at the lowest quartile of same-day ambient temperature and positive at the middle quartiles of same-day ambient temperature, suggesting that there may be evidence of effect modification by temperature (Figure 4). Since the interaction term for crustal dust and temperature was only statistically significant at the 10% level of significance, and not the 5% level of significance, interaction results should be interpreted with caution.

Source Apportionment:

Nitrate

The nitrate component, which accounts for 10.34% of total $PM_{2.5}$ concentration, primarily consists of nitrate and ammonium, but also includes elemental carbon, as well as nickel, zinc, calcium, lead, and manganese (Table 2). Urban ammonium nitrate often originates from transported agricultural sources, but can also be the result of local air pollution from traffic exhaust (both diesel and gasoline) or garbage and sewage accumulation due to higher population density (31–34). The seasonal trends we detected in nitrate concentration are consistent with the

chemical properties of the constituents of this source, as ammonium nitrate is thermodynamically favored at lower temperatures and higher relative humidity (25,26). Squizzato et al. (2018) found that seasonal trends in secondary nitrate were slightly different in Bronx and Manhattan compared with other sites in New York state, and suggested that these differences might relate to traffic patterns in New York City compared with the rest of New York State (35). Secondary nitrate components were detected in two recent $PM_{2.5}$ source apportionment analyses in New York, and in both cases this source accounted for 10-24% of total $PM_{2.5}$ (19,35).

Salt

The salt component had high levels of chlorine, sodium, and magnesium, and appeared to increase in the spring and decrease in the autumn (Table 2, Figure S7). Other recent source apportionment studies have identified more than one salt component, differentiating between fresh and aged sea salt (19,35). In an NYC-based study, Masiol et al. (2017) identified coastal areas in the northeastern US as the potential geographic origin for their fresh sea salt component, and coastal areas in the southeastern US as the potential origin of their aged sea salt component (19). The seasonal patterns we detected in $PM_{2.5}$ from sea salt may therefore represent weather patterns from across the East Coast of the US.

Crustal Dust

The crustal dust component, accounting for 15.91% of total $PM_{2.5}$ concentration, was characterized by high levels of silicon and aluminum, and also included relatively high levels of titanium and calcium. Crustal dust can come from natural sources such as soil, and can also mix with suspended road dust and construction dust (14,19,35). Our results are consistent with other studies, which also found higher concentrations of crustal dust during the summer months (19).

We qualitatively observed a change in the seasonal pattern of crustal dust concentrations during 2011-2013 (Figure S7), which coincided partially with the 2011-2014 period during which speciated $PM_{2.5}$ data were unavailable at the Bronx monitor (Figure S1). It is possible that reliance on speciated $PM_{2.5}$ data from the Queens and Manhattan monitors during this period led to measurement error in our source apportionment analysis, resulting in altered seasonal patterns in crustal dust when data from the Bronx monitor were missing.

Secondary Sulfate

Secondary sulfate, which accounted for 48.86% of total $PM_{2.5}$ concentration, was characterized by high levels of sulfate, ammonium, and organic carbon (Table 2). From 2007-2012, secondary sulfate has a decreasing overall trend with clear summer peaks, which becomes less discernible after 2012 when levels of secondary sulfate reach a plateau. Sulfate typically originates from SO_2 emissions by coal-fired power plants in the upper Ohio River Valley (36,37). Our results are consistent with other recent $PM_{2.5}$ source apportionment analyses in New York, and the overall decrease in secondary sulfate concentrations since 2007 is likely attributable to decreased use of coal for power generation (35).

Traffic and Road Dust

$PM_{2.5}$ from traffic and road dust loaded on a single component accounting for 21.44% of total $PM_{2.5}$, which consisted of organic carbon, elemental carbon, barium, lead, nitrate, potassium, copper, zinc, titanium, arsenic, and iron (Table 2). $PM_{2.5}$ from traffic exhaust tends to consist of high levels of elemental and organic carbon, as well as ammonium nitrate, iron, copper, and zinc (19,31,35). Given that ammonium was not present in our traffic component (though nitrate was), it is possible that there was some admixture with the nitrate component,

which included high levels of both ammonium and nitrate, as well as moderately high levels of elemental and organic carbon. Road dust consists of a mixture of particles from road wear, brake and tire wear, and particles deposited on or near roads from other sources (38,39). Iron, copper, lead, barium and zinc, all of which load highly on this component, are commonly used in brake lining materials, which likely contribute to road dust (39). Though previous studies have identified separate patterns for $PM_{2.5}$ from tailpipe emissions and $PM_{2.5}$ from road dust (17,35), here both appeared to load on a single component. In a source apportionment analysis of NYC using speciated $PM_{2.5}$ data from a single monitor in midtown Manhattan, Lall et al. (2011) also found a single component representing what appeared to be a mixture of emissions exhaust, re-suspended road dust and tire/brake wear (14).

$PM_{2.5}$ concentrations from traffic and road dust appeared to increase over the course of the study period (Figure S7). This trend is consistent with patterns in registered vehicles in New York City during the period from 2007-2015. There was a steady increase in registered vehicles in New York, Bronx, and Queens counties from 2011-2016, preceded by a drop in registered vehicles from 2008-2011 in New York county (vehicle registration remained constant in Bronx and Queens counties during the 2008-2011 period) (35).

Industrial

Industrial $PM_{2.5}$ consisted of high levels of chromium, nickel, and iron, as well as copper, elemental carbon and vanadium (Table 2). Concentrations were higher on weekdays than weekends and seem to have slightly decreased over the study period (Figs. S6, S7). Our industrial component was similar to that identified in Squizzato et al. (2018), which included lead, iron, manganese, copper, and zinc (35). Industrial emissions likely originate from chemical and metal processing, coke production, and metal recycling in New York State (35).

Other Sources

We expected to detect residual heating oil as a distinct source of $PM_{2.5}$ air pollution, but none of the components from our analysis were consistent with chemical patterns associated with this source. Residual heating oil was identified as a source of $PM_{2.5}$ in both of the most recent source apportionment analyses in New York, and also has been identified as an established source of $PM_{2.5}$ in previous studies of urban centers in the East Coast region of the US (14,17,19,35). Typically this source consists primarily of nickel and vanadium, and can also include sulfur, manganese, and zinc (19). Residual heating oil is of interest because it has been found to be associated with cardiovascular outcomes, and is considered to be particularly toxic due to high nickel content (40). Nickel had a prediction error of -28.14% and vanadium had a prediction error of 26.55% in APCA, both among the highest prediction errors for chemical constituents (Table 2). Although based on prior literature, residual heating oil is likely to be present as a source of $PM_{2.5}$ in New York City, our analysis may not have been accurate enough for these key markers of residual heating oil to capture it as a distinct source. Since $PM_{2.5}$ from residual heating oil comes from buildings that use residual heating oil boilers, which tend to be localized in particular neighborhoods (Upper East Side, northern Manhattan, and southern Bronx), not all of which are close to AQS monitors, it is possible that the signal for this source was diluted in our city-wide analysis using average concentrations of speciated $PM_{2.5}$ across locations.

Health Model:

Nitrate

We observed a 0.96% (95% CI: -0.17, 2.11%) increase in MI rates per one IQR increase in same-day $PM_{2.5}$ from nitrate sources, and the percent change increased to 1.79% (95% CI: -

0.20, 3.81%) when outliers were removed. Our results are consistent with other studies, which have linked exposure to nitrate as a constituent of $PM_{2.5}$ to all-cause and cardiovascular mortality (23,41,42). Although Ostro et al. (2007) did not find a statistically significant association between nitrate and cardiovascular mortality at lag 0, they did observe a 1.5% (95% CI: -0.2, 3.3%) per IQR increase in cardiovascular mortality at lag 3 (41). Cao et al. (2012) found positive associations between nitrate and cardiovascular mortality at lags 0-3, and also found a positive association between ammonium and cardiovascular mortality at lags 0 and 1. In a NYS-based study, Rich et al. (2019) found a 0.5% (95% CI: -0.1, 1.1%) increase in same-day acute cardiovascular admission rate with each IQR increase in secondary nitrate $PM_{2.5}$, as well as a 1.7% (95% CI: 0.4, 3.0%) increase for lags 0-3 in a source apportionment analysis using Positive Matrix Factorization (PMF), another commonly used source apportionment algorithm. They suggested that secondary nitrate might be a surrogate for advection of condensed reactive oxidative species (ROS) by locally formed nitrate particles, as ammonium nitrate is a particle on which carbonaceous species can condense (23). Since exposure to reactive oxidative species can lead to oxidative stress (43), a pathway by which $PM_{2.5}$ pollution can cause cardiovascular disease, advection of condensed ROS may be a mechanism by which nitrate $PM_{2.5}$ could be associated with MI.

The increase in the effect estimate for the association between nitrate and MI rate after removing outliers suggests that the outliers that were removed were driving the association downward in the main analysis. Given the mixture of local and regional sources that can coexist within the nitrate $PM_{2.5}$ source, it is possible that the chemical composition of the nitrate source on days when nitrate concentration was particularly high or particularly low differed slightly from the chemical composition on days with nitrate concentrations closer to the mean. If future

source apportionment analyses detect more than one compositionally distinct source of nitrate from $PM_{2.5}$, it would be interesting to determine if one nitrate-related source is more associated with MI than another.

Crustal Dust

We observed a 0.43% (95% CI: -0.13, 0.99%) increase in MI admission rates per one IQR increase in crustal $PM_{2.5}$, which is predominantly composed of silicon and aluminum. Silicon as a chemical constituent of $PM_{2.5}$ has been linked to cardiovascular mortality, and could therefore contribute to the observed association (44,45). Mortality related to $PM_{2.5}$ has also been found to be modified by increased proportion of aluminum, a key constituent of crustal dust (46). Aluminum and silicon together as chemical constituents of $PM_{2.5}$ have also been linked to inflammation and oxidative stress (47). Prior literature on the potential association between crustal $PM_{2.5}$ and cardiovascular disease has been mixed. In a Boston-based study, Kioumourtzoglou et al. (2014) found that crustal dust was strongly associated with hospital admissions for cardiovascular disease when APCA was used for source apportionment analysis, but not when PMF was used (17). An NYC-based study using data from 2001 found no association between their equivalent component (soil) and cardiovascular admissions (14). On the other hand, an NYS source apportionment study identified a road dust component that had similar composition to our crustal dust component (35), and in the subsequent health analysis this component was associated with a 0.8% (95% CI: 0.0, 1.7%) increase in same-day hospital admission rates for MI (23).

The association between crustal dust and MI admission rate appears to be negative at the lowest quartile of temperature and positive at the middle quartiles of temperature (Figure 4). These results should be interpreted with caution, as the interaction term between crustal dust

$PM_{2.5}$ and temperature was only statistically significant at the 10% level of significance, and not the 5% level of significance. However, we choose to report these findings because it is likely that we had reduced ability to detect statistically significant interactions due to insufficient statistical power, and the associations we were able to detect at the 10% level of significance may be informative for future studies with larger sample sizes. Our results are consistent with previous studies which have found most positive effect estimates for the association between total $PM_{2.5}$ and mortality at moderate temperatures (46).

There are several potential mechanisms that could lead to differences in the effect of crustal dust on MI rate at different temperatures, including seasonal changes in crustal dust composition and exposure measurement error related to infiltration of outdoor pollutants indoors at different ambient temperatures (46). Chemical composition of crustal $PM_{2.5}$ may vary seasonally, as a function of prevailing winds and the geographic origin of $PM_{2.5}$. Masiol et al. (2017) found that increased crustal material concentrations in NYC were associated with moderate winds from the south, and that concentrations were highest in the spring and summer (19). Transported crustal material related to southern winds may have slightly different composition and therefore different toxicity compared with crustal dust originating from other areas. Crustal dust mixed with construction or road dust also likely has a different chemical composition than crustal dust from soil, and the degree to which the crustal dust source detected here includes other dust sources may vary seasonally. Temperature-related variation in ventilation likely plays a role in measurement of the effect of $PM_{2.5}$ on MI, as people are most likely to keep their windows open while indoors at moderate temperatures and least likely to keep their windows open at extreme cold and warm temperatures. In a nationwide study of total $PM_{2.5}$ and selected $PM_{2.5}$ constituents, Franklin et al. (2008) found that effect estimates for the

association between total $PM_{2.5}$ and overall mortality were highest at moderate temperatures and lowest at cold and warm temperatures—the authors used this finding to argue that temperature can be used as a surrogate for ventilation (46).

The apparent protective effect of crustal dust $PM_{2.5}$ at low temperatures is unexpected. One possible, though somewhat unlikely, explanation for our result may be residual confounding from effects of wind. Wind speed can be associated with risk of MI (48), and also could be associated with source-specific $PM_{2.5}$. If at low temperatures, high wind speed is positively associated with MI and negatively associated with crustal dust, this could lead to an apparent protective effect of crustal dust at low temperatures, without necessarily reflecting a true protective effect. Perhaps high wind velocity at low temperatures contributes to wind chill, worsening the experience of cold at high wind velocities and therefore contributing to increased cold-induced coronary vasoconstriction and subsequent MI (48). If cold-weather winds in NYC tend to originate from a direction that carries less crustal dust, and the composition of crustal dust from this direction makes it non-toxic, these pathways could lead to a negative association between crustal dust and MI, through windspeed as a backdoor path.

Industrial

We observed a 0.35% (95% CI: -0.21, 0.91%) increase in MI hospital admission rate per one IQR increase in industrial $PM_{2.5}$. The industrial component contains high levels of nickel, which has been associated with cardiovascular disease in several studies, and is considered to be a chemical constituent of $PM_{2.5}$ of particular concern for health outcomes (49,50). Several of the other chemical constituents of industrial $PM_{2.5}$, including zinc, chromium, and lead, were also found in a study based in Xi'an, China to be associated with increased risk for cardiovascular

disease (50). Previous New York based studies have not found evidence of an association between industrial $PM_{2.5}$ and MI or cardiovascular disease hospital admissions (14,23).

Traffic and Road Dust

Although we expected to find an association between traffic-related $PM_{2.5}$ and MI hospital admission rate, we did not find evidence of this association. Traffic-related $PM_{2.5}$ has been associated with cardiovascular disease in prior source apportionment studies, and several common chemical constituents of traffic emissions are known to be associated with systemic inflammation, a biological pathway by which $PM_{2.5}$ can be associated with MI (14,17,23,49,51,52). Over the course of our study period, there were several policy changes aimed at reducing the toxicity of traffic-related air pollution, including the requirement for NO_x control for heavy-duty diesel trucks and buses (2010) and the requirement in NYS that all distillate fuels sold in NYS be ultralow sulfur by July 1, 2012 (35). It is possible that previously-recorded associations between traffic-related $PM_{2.5}$ and cardiovascular disease have been reduced as a result of these policy changes, rendering them undetectable in our analysis. Of note, Rich et al. (2019) detected a 2.3% (95% CI: 0.1, 4.5%) increase in MI hospital admission rate per IQR increase in spark-ignition emissions, a 0.4% (95% CI: -0.5, 1.2%) increase in MI hospital admission rate per IQR increase in diesel, and a 0.8% (95% CI: 0.0, 1.7%) increase in MI hospital admission rate per IQR increase in road dust (23). It is possible that policy changes aimed at reducing the toxicity of diesel fuel for buses were particularly effective in NYC, as public transportation may account for a particularly large proportion of traffic-related $PM_{2.5}$ air pollution in NYC. However, it is also important to note that since Rich et al. were using data from the entire population of NYS, they likely had more statistical power to detect changes in MI rates from spark-ignition emissions and road dust that may have appeared null in our analysis.

Secondary Sulfate

We did not find evidence of an association between secondary sulfate and MI hospital admission rate. We might have expected to find an association, since sulfate as a constituent of $PM_{2.5}$ has previously been found to be associated with MI and overall cardiovascular disease (44,53,54). However, secondary sulfate as an identified source of $PM_{2.5}$ from source apportionment analysis has not consistently been associated with cardiovascular hospital admissions (14,23).

Strengths:

This study had several strengths, including the leveraging of the SPARCS dataset, reproducibility of source apportionment analysis, and examination of interaction by temperature. The SPARCS dataset and the EPA's AQS database allowed for a long study period during which it was possible to detect long-term trends. The public availability of the AQS database also improves reproducibility of our source apportionment analysis, which is available on GitHub. We also were able to examine potential effect modification by temperature, using an easily interpretable method: quantile indicator terms.

Limitations:

This study also had several limitations, including decreased power due to limited sample size and exposure measurement error. Since data on $PM_{2.5}$ constituents were only available once every 3 or 6 days, our final dataset included only 978 days, despite spanning 9 years. Since our models had multiple covariates, non-linear terms, and interaction terms, a sample size of 978 may not have allowed for sufficient power to detect all associations that were present. Lack of daily speciated $PM_{2.5}$ data also prevented examination of exposure windows of interest other

than same-day exposure to source-specific $PM_{2.5}$. Several studies have found evidence of associations between source-specific $PM_{2.5}$ and cardiovascular disease at lag 1 or lag 2 for sources where no association was detectable at lag 0, so it would have been useful to be able to examine exposure windows other than same-day exposure (14,17).

Our results are likely subject to exposure measurement error, as we did not have speciated $PM_{2.5}$ data available from all three monitors for each day included in the dataset. We expect that $PM_{2.5}$ sources, and consequently the concentrations of $PM_{2.5}$ and its chemical constituents, vary at different geographic locations within NYC, and we aimed to capture the city-wide values by taking the average of values measured at three separate locations. However, from 2011-2014 there were missing data in the AQS dataset for the Bronx monitor, and for most of 2007 there were missing data for the Manhattan monitor (Figure S1). Our computed city-wide averages on days with missing data only represent the values at the locations with available data, and therefore may not be comparable to computed city-wide averages on days with full data. This could have influenced our source apportionment solution as it may have influenced the correlations among city-wide averaged constituents over time. In a source apportionment analysis using $PM_{2.5}$ chemical constituent data from the same AQS dataset as the current study, Squizzato et al. (2018) found that the same sources were identifiable at the Queens, Bronx, and Manhattan locations, but source contributions differed by site (35). Their results suggest that aggregating all available data on days where data were missing from one or more monitors could lead to biased estimates for city-wide daily contributions of source-specific $PM_{2.5}$.

For these missing data to result in differential error in effect estimates for the health analysis, there would need to be bias not only in the predicted concentration of each source, but also in toxicity. For example, Squizzato et al. (2018) found that concentrations for secondary

nitrate were lowest at the Queens site and highest at the Manhattan site (35). If the composition of $PM_{2.5}$ from secondary nitrate is identical across NYC, but has different concentrations in different areas, these differences are not necessarily a cause for concern in our effect estimates for MI. However, if secondary nitrate were higher in Manhattan because of a local source of nitrate that had a slightly different composition from regional sources of nitrate and was therefore more toxic, this could mean that our effect estimates were biased away from the null—this hypothetical local Manhattan source would be more influential in our health analysis when data from the Bronx site are not available, and would overestimate the degree which $PM_{2.5}$ from secondary nitrate is associated with MI for the average person in NYC. This is a completely hypothetical example, not based on any known local toxic sources of nitrate in Manhattan compared with the Bronx, and similar examples could be considered for other identified sources. Missing data could lead to bias either toward or away from the null depending on the existence of a toxic local source near a monitoring site with missing data or a site with complete data.

Future studies should explore improved practices for handling missing data to account for geographic variation in $PM_{2.5}$ chemical constituents. Such methods might include simulation models for $PM_{2.5}$ constituents, similar to those used in the EPA's Community Multiscale Air Quality Model (CMAQ). Since these kinds of models rely on the availability of air quality data from monitors, development of such approaches would need to be accompanied by increased availability of speciated $PM_{2.5}$ data at air quality monitors. Increased number of monitors tracking speciated $PM_{2.5}$ and use of simulation models for estimating exposure to chemical constituents of $PM_{2.5}$ could also allow for population-weighted averaging for $PM_{2.5}$ constituents, and therefore source-specific $PM_{2.5}$. If speciated $PM_{2.5}$ simulation models with sufficiently high

geographic accuracy were available, population-weighted averaging using gridded surfaces would reduce exposure measurement error for health analyses.

Other sources of measurement error include outdoor versus indoor exposure, as well as interpretation issues in source apportionment analysis. We did not account for differences between outdoor and indoor exposure to source-specific $PM_{2.5}$ and used only data from outdoor monitors, even though most people spend the majority of their time indoors and are therefore primarily exposed to indoor air. We also were not able to identify all key sources of $PM_{2.5}$ and did not have complete separation for all sources examined. This limitation stems from more general limitations of APCA and other dimensionality-reduction methods for source apportionment, which include issues with interpretability, as well as subjectivity in detecting patterns that are identifiable as sources of $PM_{2.5}$. Finally, our results may not be generalizable to other locations, which may be characterized by a different mixture of $PM_{2.5}$ pollution sources.

Conclusions:

Leveraging data from SPARCS and from the EPA's AQS database, we found non-significant increased rates of hospital admissions for MI with increased same-day nitrate, crustal dust, and industrial $PM_{2.5}$ in NYC from 2007-2015. We observed a non-significant negative association between crustal dust and MI admissions rate at low temperatures and a non-significant positive association between crustal dust at moderate temperatures. Continued assessment of effect modification by temperature in studies examining the relationship between source-specific $PM_{2.5}$ and cardiovascular outcomes is warranted. Exposure measurement error from temperature-dependent changes in ventilation may have a meaningful effect on effect

estimates for health outcomes with certain sources of $PM_{2.5}$ and should be taken into account more regularly in source apportionment health studies.

Supplemental Figures and Tables

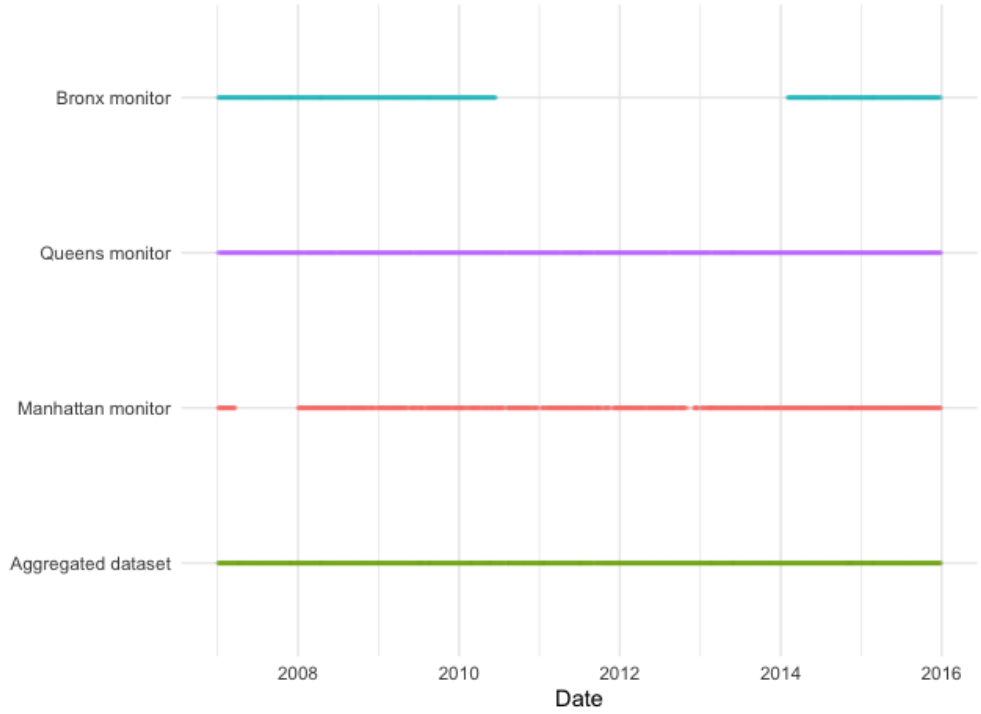


Figure S1. Data availability at each monitor, and for the full aggregated dataset used in the main analysis.

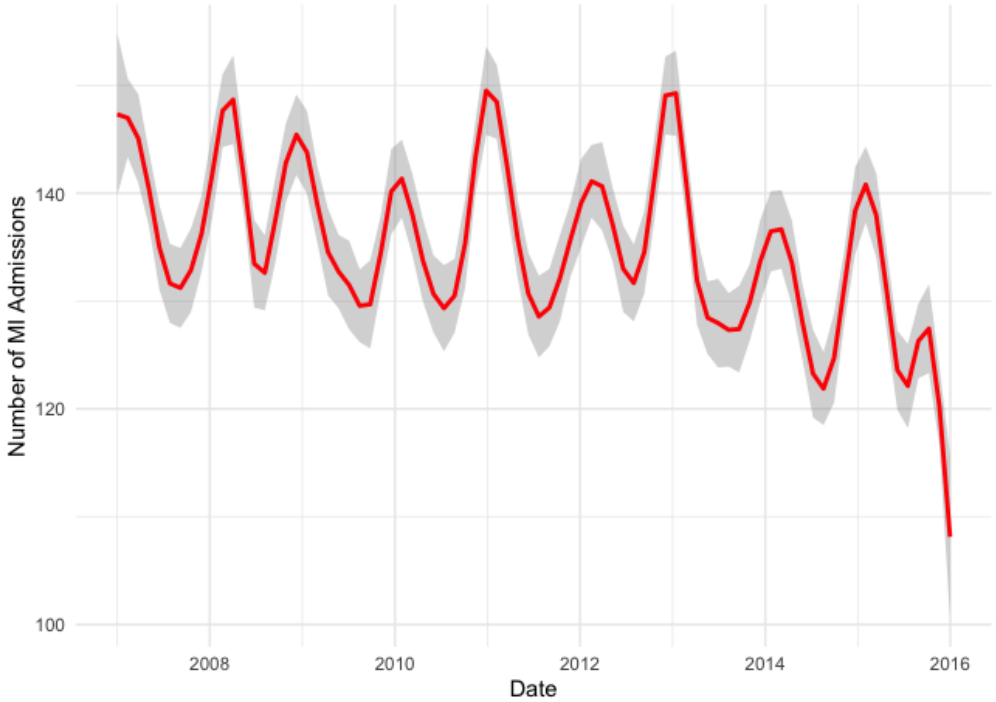


Figure S2. Seasonal and long-term trends for MI admission count

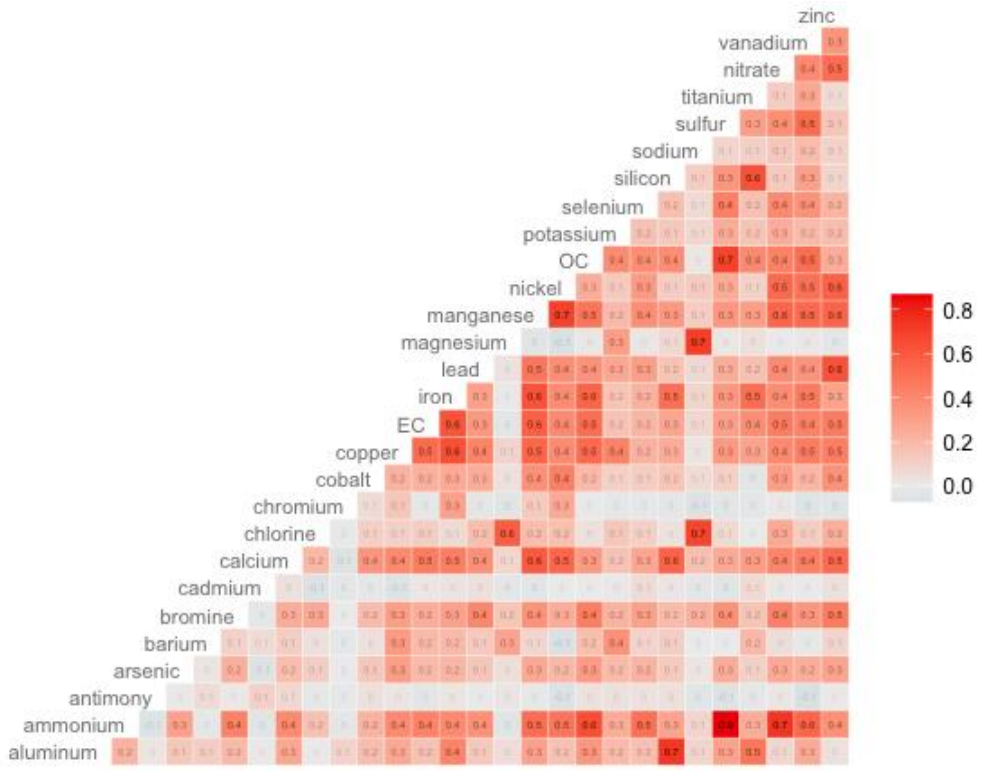


Figure S3. Correlation matrix for chemical constituents of $PM_{2.5}$

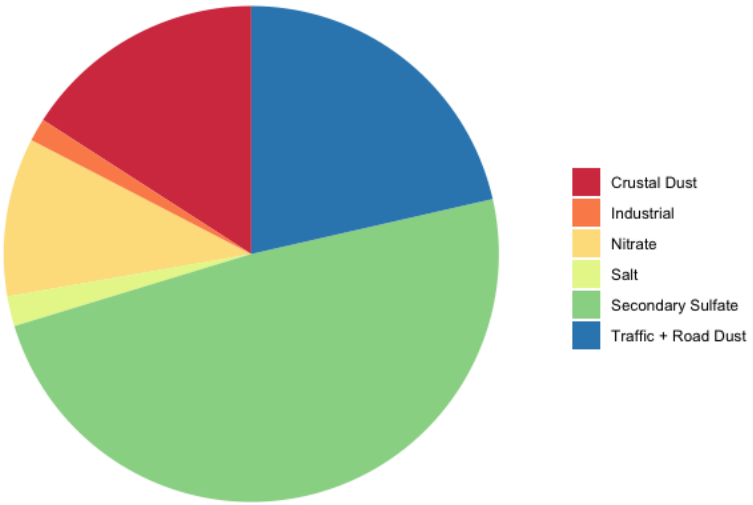


Figure S4. Contributions of each source to total $PM_{2.5}$ concentration. 1) nitrate, 2) salt, 3) crustal dust, 4) secondary/regional sulfate, 5) traffic and road dust, 6) industrial

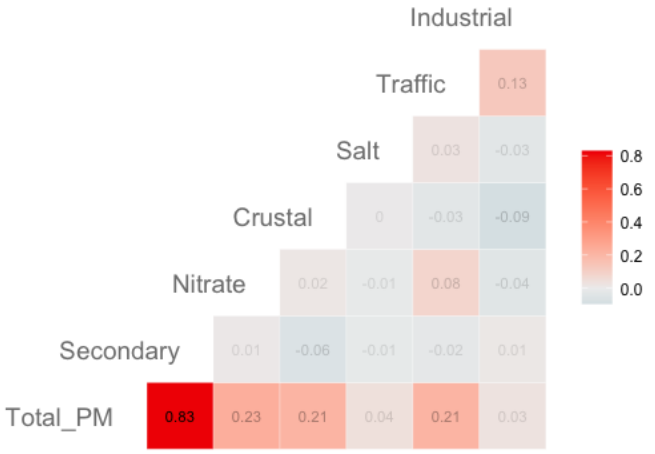


Figure S5. Correlation matrix for each identified source from APCA, compared with actual total $PM_{2.5}$ concentration.

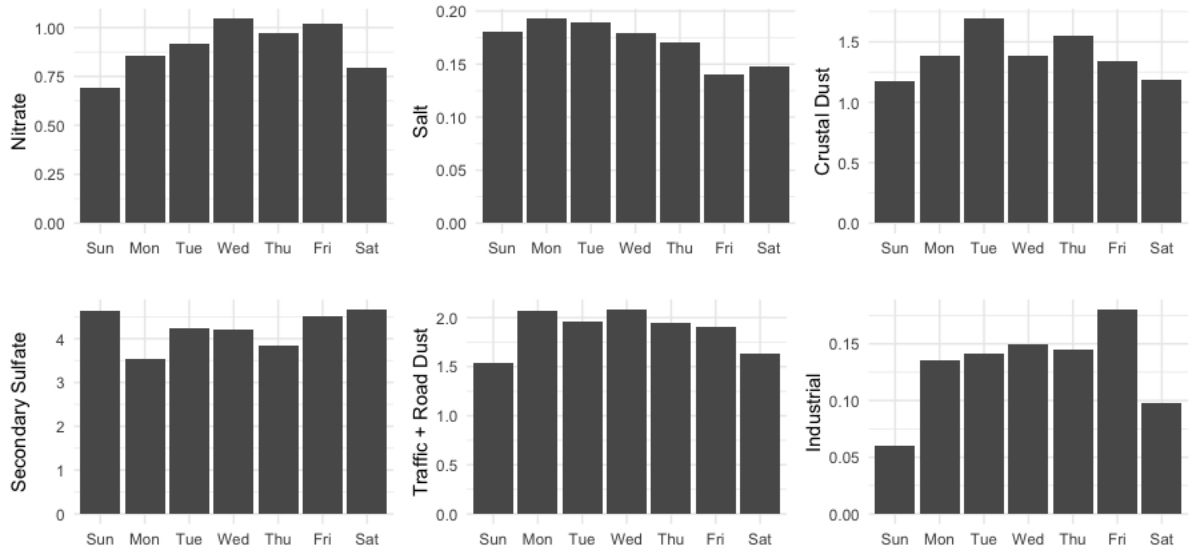


Figure S6. Mean source-specific $PM_{2.5}$ ($\mu g/m^3$) per day of week.

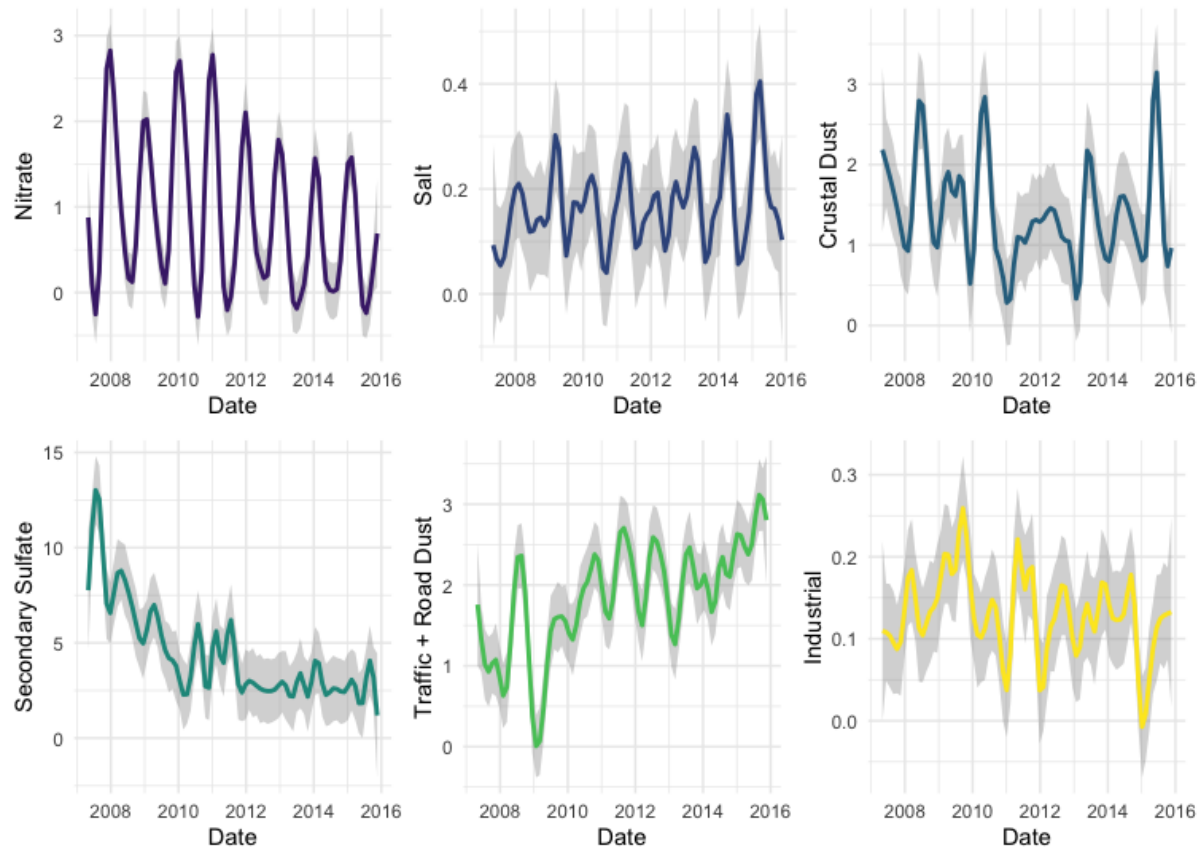


Figure S7. Seasonal and long-term trends for each source, reported in $\mu g/\mathrm{m}^3$

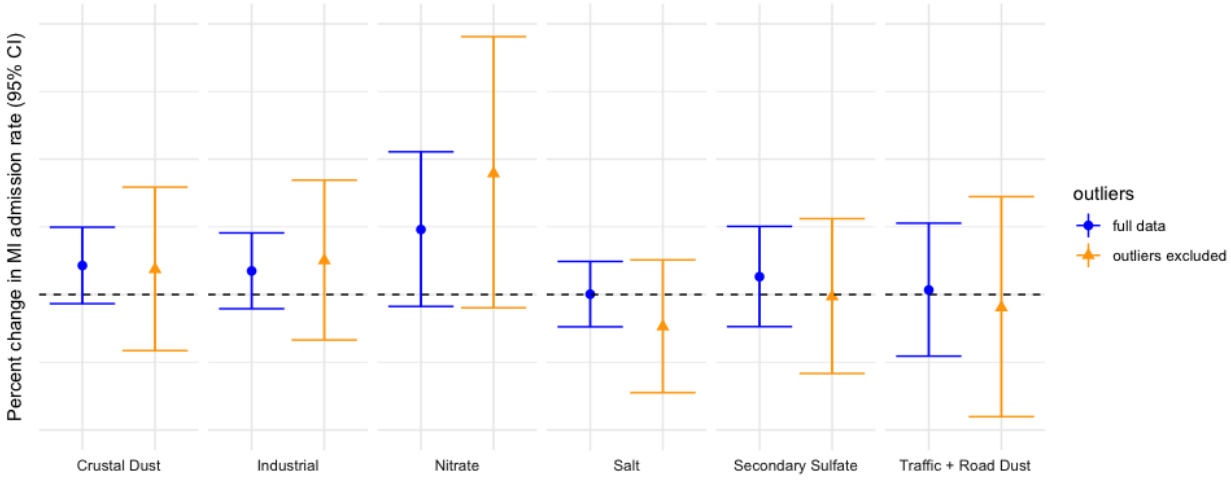


Figure S8. Forest plot comparing percent change in MI admission rate per IQR change in source-specific $PM_{2.5}$ using full data (blue circles as point values) and estimates when outliers more than 3SD from the mean are excluded (orange triangles as point values)

References:

1. Global Health Estimates 2020: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2019. Geneva: World Health Organization; 2020.

2. Murray CJL, Aravkin AY, Zheng P, Abbafati C, Abbas KM, Abbasi-Kangevari M, et al. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet. 2020 Oct;396(10258):1223–49.

3. Hoek G, Krishnan RM, Beelen R, Peters A, Ostro B, Brunekreef B, et al. Long-term air pollution exposure and cardio- respiratory mortality: a review. Environ Health. 2013 Dec;12(1):43.

4. Liu C, Chen R, Sera F, Vicedo-Cabrera AM, Guo Y, Tong S, et al. Ambient Particulate Air Pollution and Daily Mortality in 652 Cities. N Engl J Med. 2019 Aug 22;381(8):705–15.

5. Farhadi Z, Abulghasem Gorgi H, Shabaninejad H, Aghajani Delavar M, Torani S. Association between PM2.5 and risk of hospitalization for myocardial infarction: a systematic review and a meta-analysis. BMC Public Health. 2020 Dec;20(1):314.

6. Brook RD, Rajagopalan S, Pope CA, Brook JR, Bhatnagar A, Diez-Roux AV, et al. Particulate Matter Air Pollution and Cardiovascular Disease: An Update to the Scientific Statement From the American Heart Association. Circulation. 2010 Jun;121(21):2331–78.

7. Newby DE, Mannucci PM, Tell GS, Baccarelli AA, Brook RD, Donaldson K, et al. Expert position paper on air pollution and cardiovascular disease. Eur Heart J. 2015 Jan 7;36(2):83–93b.

8. Weichenthal S, Lavigne E, Evans G, Pollitt K, Burnett RT. Ambient PM2.5 and risk of emergency room visits for myocardial infarction: impact of regional PM2.5 oxidative potential: a case-crossover study. Environ Health. 2016 Dec;15(1):46.

9. Dai L, Zanobetti A, Koutrakis P, Schwartz JD. Associations of Fine Particulate Matter Species with Mortality in the United States: A Multicity Time-Series Analysis. Environ Health Perspect. 2014 Aug;122(8):837–42.

10. Davoodabadi Z, Soleimani A, Pourmoghaddas A, Hosseini SM, Jafari-Koshki T, Rahimi M, et al. Correlation between air pollution and hospitalization due to myocardial infarction. ARYA Atheroscler [Internet]. 2019 Aug 11 [cited 2020 Nov 16];15(4). Available from: https://doi.org/10.22122/arya.v15i4.1834

11. Thurston GD, Ito K, Lall R. A source apportionment of U.S. fine particulate matter air pollution. Atmos Environ. 2011 Aug;45(24):3924–36.

12. Ito K, Xue N, Thurston G. Spatial variation of PM2.5 chemical species and source-apportioned mass concentrations in New York City. Atmos Environ. 2004 Oct;38(31):5269–82.

13. Hopke PK, Gladney ES, Gordon GE, Zoller WH, Jones AG. The use of multivariate analysis to identify sources of selected elements in the Boston urban aerosol. Atmospheric Environ 1967. 1976 Jan;10(11):1015–25.

14. Lall R, Ito K, Thurston GD. Distributed Lag Analyses of Daily Hospital Admissions and Source-Apportioned Fine Particle Air Pollution. Environ Health Perspect. 2011 Apr;119(4):455–60.

15. Wurth R, Kioumourtzoglou M-A, Tucker KL, Griffith J, Manjourides J, Suh H. Fine particle sources and cognitive function in an older Puerto Rican cohort in Greater Boston: Environ Epidemiol. 2018 Sep;2(3):e022.

16. Hopke PK, Ito K, Mar T, Christensen WF, Eatough DJ, Henry RC, et al. PM source apportionment and health effects: 1. Intercomparison of source apportionment results. J Expo Sci Environ Epidemiol. 2006 May;16(3):275–86.

17. Kioumourtzoglou M-A, Coull BA, Dominici F, Koutrakis P, Schwartz J, Suh H. The impact of source contribution uncertainty on the effects of source-specific PM2.5 on hospital admissions: A case study in Boston, MA. J Expo Sci Environ Epidemiol. 2014 Jul;24(4):365–71.

18. Thurston GD, Spengler JD. A quantitative assessment of source contributions to inhalable particulate matter pollution in metropolitan Boston. Atmospheric Environ 1967. 1985 Jan;19(1):9–25.

19. Masiol M, Hopke PK, Felton HD, Frank BP, Rattigan OV, Wurth MJ, et al. Source apportionment of PM2.5 chemically speciated mass and particle number concentrations in New York City. Atmos Environ. 2017 Jan;148:215–29.

20. Gibson EA, Nunez Y, Abuawad A, Zota AR, Renzetti S, Devick KL, et al. An overview of methods to address distinct research questions on environmental mixtures: an application to persistent organic pollutants and leukocyte telomere length. Environ Health. 2019 Dec;18(1):76.

21. Kavouras IG, Koutrakis P, Cereceda-Balic F, Oyola P. Source Apportionment of PM_{10} and PM_{25} in Five Chilean Cities Using Factor Analysis. J Air Waste Manag Assoc. 2001 Mar;51(3):451–64.

22. Ito K, Christensen WF, Eatough DJ, Henry RC, Kim E, Laden F, et al. 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. J Expo Sci Environ Epidemiol. 2006 Jul;16(4):300–10.

23. Rich DQ, Zhang W, Lin S, Squizzato S, Thurston SW, van Wijngaarden E, et al. Triggering of cardiovascular hospital admissions by source specific fine particle concentrations in urban centers of New York State. Environ Int. 2019 May;126:387–94.

24. Squizzato S, Masiol M, Rich DQ, Hopke PK. PM2.5 and gaseous pollutants in New York State during 2005–2016: Spatial variability, temporal trends, and economic influences. Atmos Environ. 2018 Jun;183:209–24.

25. Schaap M, Spindler G, Schulz M, Acker K, Maenhaut W, Berner A, et al. Artefacts in the sampling of nitrate studied in the "INTERCOMP" campaigns of EUROTRAC-AEROSOL. Atmos Environ. 2004 Dec;38(38):6487–96.

26. Pinder RW, Dennis RL, Bhave PV. Observable indicators of the sensitivity of PM2.5 nitrate to emission reductions—Part I: Derivation of the adjusted gas ratio and applicability at regulatory-relevant time scales. Atmos Environ. 2008 Feb;42(6):1275–86.

27. Sun Z, Chen C, Xu D, Li T. Effects of ambient temperature on myocardial infarction: A systematic review and meta-analysis. Environ Pollut. 2018 Oct;241:1106–14.

28. Statewide Planning and Research Cooperative System (SPARCS) [Internet]. Available from: https://www.health.ny.gov/statistics/sparcs/

29. US Environmental Protection Agency. Air Quality System Data Mart, Daily Summary Data: Particulates. Available from: https://www.epa.gov/airdata

30. Cosgrove BA, Lohmann D, Mitchell KE, Houser PR, Wood EF, Schaake JC, et al. Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. J Geophys Res Atmospheres. 2003 Nov 27;108(D22):2002JD003118.

31. Zhou C, Zhou H, Holsen TM, Hopke PK, Edgerton ES, Schwab JJ. Ambient Ammonia Concentrations Across New York State. J Geophys Res Atmospheres. 2019 Jul 27;124(14):8287–302.

32. Thiruvengadam A, Besch M, Carder D, Oshinuga A, Pasek R, Hogo H, et al. Unregulated greenhouse gas and ammonia emissions from current technology heavy-duty vehicles. J Air Waste Manag Assoc. 2016 Nov 1;66(11):1045–60.

33. Galán Madruga D, Fernández Patier R, Sintes Puertas MA, Romero García MD, Cristóbal López A. Characterization and Local Emission Sources for Ammonia in an Urban Environment. Bull Environ Contam Toxicol. 2018 Apr;100(4):593–9.

34. Sun K, Tao L, Miller DJ, Khan MA, Zondlo MA. On-Road Ammonia Emissions Characterized by Mobile, Open-Path Measurements. Environ Sci Technol. 2014 Apr;48(7):3943–50.

35. Squizzato S, Masiol M, Rich DQ, Hopke PK. A long-term source apportionment of PM2.5 in New York State during 2005–2016. Atmos Environ. 2018 Nov;192:35–47.

36. Dutkiewicz VA, Qureshi S, Khan AR, Ferraro V, Schwab J, Demerjian K, et al. Sources of fine particulate sulfate in New York. Atmos Environ. 2004 Jun;38(20):3179–89.

37. Hopke PK, Zhou L, Poirot RL. Reconciling Trajectory Ensemble Receptor Model Results with Emissions. Environ Sci Technol. 2005 Oct;39(20):7980–3.

38. Amato F, Moreno T, Pandolfi M, Querol X, Alastuey A, Delgado A, et al. Concentrations, sources and geochemistry of airborne particulate matter at a major European airport. J Environ Monit. 2010;12(4):854.

39. Thorpe A, Harrison RM. Sources and properties of non-exhaust particulate matter from road traffic: A review. Sci Total Environ. 2008 Aug;400(1–3):270–82.

40. Peltier RE, Hsu S-I, Lall R, Lippmann M. Residual oil combustion: a major source of airborne nickel in New York City. J Expo Sci Environ Epidemiol. 2009 Sep;19(6):603–12.

41. Ostro B, Feng W-Y, Broadwin R, Green S, Lipsett M. The Effects of Components of Fine Particulate Air Pollution on Mortality in California: Results from CALFINE. Environ Health Perspect. 2007 Jan;115(1):13–9.

42. Cao J, Xu H, Xu Q, Chen B, Kan H. Fine Particulate Matter Constituents and Cardiopulmonary Mortality in a Heavily Polluted Chinese City. Environ Health Perspect. 2012 Mar;120(3):373–8.

43. Zhang W, Lin S, Hopke PK, Thurston SW, van Wijngaarden E, Croft D, et al. Triggering of cardiovascular hospital admissions by fine particle concentrations in New York state: Before, during, and after implementation of multiple environmental policies and a recession. Environ Pollut. 2018 Nov;242:1404–16.

44. Ostro B, Lipsett M, Reynolds P, Goldberg D, Hertz A, Garcia C, et al. Long-Term Exposure to Constituents of Fine Particulate Air Pollution and Mortality: Results from the California Teachers Study. Environ Health Perspect. 2010 Mar;118(3):363–9.

45. Badaloni C, Cesaroni G, Cerza F, Davoli M, Brunekreef B, Forastiere F. Effects of long-term exposure to particulate matter and metal components on mortality in the Rome longitudinal study. Environ Int. 2017 Dec;109:146–54.

46. Franklin M, Koutrakis P, Schwartz J. The Role of Particle Composition on the Association Between PM2.5 and Mortality. Epidemiology. 2008 Sep;19(5):680–9.

47. Becker S, Dailey LA, Soukup JM, Grambow SC, Devlin RB, Huang Y-CT. Seasonal variations in air pollution particle-induced inflammatory mediator release and oxidative stress. Environ Health Perspect. 2005 Aug;113(8):1032–8.

48. Mohammad MA, Koul S, Rylance R, Fröbert O, Alfredsson J, Sahlén A, et al. Association of Weather With Day-to-Day Incidence of Myocardial Infarction: A SWEDEHEART Nationwide Observational Study. JAMA Cardiol. 2018 Nov 1;3(11):1081.

49. Thurston GD, Burnett RT, Turner MC, Shi Y, Krewski D, Lall R, et al. Ischemic Heart Disease Mortality and Long-Term Exposure to Source-Related Components of U.S. Fine Particle Air Pollution. Environ Health Perspect. 2016 Jun;124(6):785–94.

50. Huang W, Cao J, Tao Y, Dai L, Lu S-E, Hou B, et al. Seasonal Variation of Chemical Species Associated With Short-Term Mortality Effects of PM2.5 in Xi'an, a Central City in China. Am J Epidemiol. 2012 Mar 15;175(6):556–66.

51. Mills NL, Törnqvist H, Gonzalez MC, Vink E, Robinson SD, Söderberg S, et al. Ischemic and Thrombotic Effects of Dilute Diesel-Exhaust Inhalation in Men with Coronary Heart Disease. N Engl J Med. 2007 Sep 13;357(11):1075–82.

52. Mills NL, Törnqvist H, Robinson SD, Gonzalez M, Darnley K, MacNee W, et al. Diesel Exhaust Inhalation Causes Vascular Dysfunction and Impaired Endogenous Fibrinolysis. Circulation. 2005 Dec 20;112(25):3930–6.

53. Yang J, Zhou M, Li M, Yin P, Hu J, Zhang C, et al. Fine particulate matter constituents and cause-specific mortality in China: A nationwide modelling study. Environ Int. 2020 Oct;143:105927.

54. Peterson GCL, Hogrefe C, Corrigan AE, Neas LM, Mathur R, Rappold AG. Impact of Reductions in Emissions from Major Source Sectors on Fine Particulate Matter–Related Cardiovascular Mortality. Environ Health Perspect. 2020 Jan;128(1):017005.

SUPPLEMENTAL MATERIALS

S1. Structured Thesis Plan

S1a. Narrative*

* Some elements of the thesis plan have changed since it was originally written

Title: Assessing the effect of air pollution on hospitalization due to myocardial infarction using novel source apportionment methods

Student: Rachel Tao, Applied Biostatistics

Introduction

The association between exposure to air pollution and cardiovascular outcomes is well-established, with increasing evidence that fine particulate matter ($PM_{2.5}$) may be particularly harmful (1). Better understanding of which sources of $PM_{2.5}$ pollution are most associated with cardiovascular risk in specific areas could inform local policy on efforts to reduce $PM_{2.5}$ pollution.

Source apportionment is the process of identifying sources of pollution from pollution data. Source apportionment methods usually describe the variability among pollutants in terms of unobserved variables derived from dimensionality-reduction methods such as Principal Components Analysis (2–4) These new variables can be used in subsequent analyses, such as time-series models, to assess the potential association between specific sources of pollution and health outcomes (4–6). All existing source apportionment methods require some degree of subjective decision-making with respect to selecting how many potential sources to include in the final model, so novel methods that reduce this subjectivity are needed (4). Principal Component Pursuit (PCP) is a developing method for source apportionment that would reduce the degree of subjectivity for selecting sources (see methods).

To our knowledge, source apportionment analysis of the association between fine particulate matter and cardiovascular risk in New York City has not been done since 2011(6), using data from 2001-2002. For the current project, we plan to provide an update to this previous analysis, and compare previously used source apportionment methods, Positive Matrix Factorization (PMF) and Absolute Principal Component Analysis (APCA), with our novel method, PCP.

Study Aims

Aim 1: Apply PCP to identify $PM_{2.5}$ sources in New York City. Compare PCP with previously used source apportionment methods, Positive Matrix Factorization (EPA) and/or Absolute Principal Component Analysis.

Aim 2: Identify sources of $PM_{2.5}$ pollution using PCP and investigate the association with hospitalization due to myocardial infarction.

Hypotheses

We expect the same pollution sources to be identified using PMF, APCA, and PCP source apportionment methods, and that the time trends of a given identified source (e.g., traffic) will be correlated across the different methods. We will perform time-series analyses with hospitalization due to MI as the outcome, and factors/components identified as pollution sources as predictors of interest. We expect the following sources of $PM_{2.5}$ to be positively associated with hospitalization due to myocardial infarction: traffic and residual oil. We will test for effect modification by temperature, and expect that the relationship between $PM_{2.5}$ and myocardial infarction will be stronger at higher temperatures.

Methods

Air pollution data:

Air pollution data for New York City have been extracted from the Air Quality System (AQS) dataset collected and maintained by the United States Environmental Protection Agency (7). This dataset is publicly accessible and includes ambient concentrations of a selection of certain pollutants. We will use daily concentrations of total $PM_{2.5}$ and the following chemical components of $PM_{2.5}$: aluminum, arsenic, bromine, calcium, chlorine, chromium, copper, iron, lead, magnesium, manganese, nickel, selenium, silicon, sulfur, titanium, nitrates, vanadium, and zinc.

Meteorological data will be extracted from the National Oceanic and Atmospheric Administration (NOAA) National Climatic Data Center (8), and we will use the values from the monitor in Central Park for temperature and humidity.

Hospitalization data:
Daily hospitalization data for myocardial infarction in New York City has been extracted from the Statewide Planning and Research Cooperative System dataset. I will use daily city-wide counts of myocardial infarctions. Because this dataset is not considered human subjects, this project does not require IRB approval. I have been given permission to use these data by Drs. Boehme and Kioumourtzoglou.

Source apportionment:
Two previously used approaches (PMF, APCA) and one novel approach (PCP) to source apportionment will be used. All three methods will describe variability among the observed $PM_{2.5}$ components in terms of unobserved variables, which may be referred to as factors or components depending on the method. Expert knowledge will be used to assign factors/components to sources of $PM_{2.5}$ (e.g., traffic, residual heating oil).

Positive Matrix Factorization (PMF) is a source apportionment method developed by the US EPA (9), which takes into account the uncertainty associated with each observation and imposes the restriction that all chemical component contributions are non-negative. For PMF, the number of expected factors (i.e., number of sources) must be provided *a priori*. We will use expert knowledge to determine how many sources of $PM_{2.5}$ we should expect and perform PMF using the EPA's software. We will use loadings on factors to determine if the sources we expected to see are indeed present, and then use factors identifiable as specific sources in the time-series health analysis (10).

Absolute Principal Component Analysis (APCA) is an extension of Principal Components Analysis, which creates new unobserved variables called components or principal components to explain the total variance in fewer variables than the original number. In PCA, principal components are uncorrelated with, or orthogonal to, one another. For both PCA and APCA, it is necessary to either specify *a priori* to only select principal components that together account for more than a certain proportion of the total variance (e.g. > 80%) or visualize the results to determine an appropriate cut-off point (11). After PCA is conducted, APCA will involve rotating the principal components determined from PCA and rescaling the component scores relative to a reference of zero pollution, so that total $PM_{2.5}$ mass concentrations can be regressed on the component scores (2).

Principal Component Pursuit (PCP) is a novel method used primarily in computer vision applications, such as face recognition, that can be understood as a robust form of Principal Components Analysis (12). PCP is in the process of being adapted as a method of source apportionment. Although expert knowledge provides us with reasonable estimates for the number of sources to expect, it would be preferable to use a method that allows for identification of potential $PM_{2.5}$ sources without specifying the expected number of sources or total proportion of variance explained a priori. PCP aims to deconstruct the data matrix into a low-rank matrix, which we can use to identify distinct patterns in the data such as pollution sources, and a sparse matrix, which includes unique events that cannot be explained by the identified consistent

patterns. Once fully adapted for use in environmental epidemiology, principal components from the low-rank matrix that explain a non-zero percent of the total variance will be considered signals of potential sources of $PM_{2.5}$. I work with Dr. Marianthi-Anna Kioumourtzoglou as part of the multidisciplinary development team adapting PCP to environmental epidemiology applications. To identify sources, we will examine the loadings of each of the chemical components of $PM_{2.5}$ on each principal component and use expert knowledge to identify sources from the principal components.

Time-series health analysis:
Once sources of $PM_{2.5}$ have been identified, we will conduct a time-series analysis using a Poisson regression model, which will control for temperature, humidity, day of the week, season, and long-term trends. To determine which pollution sources are associated with MI, each factor or component identified as a pollution source using source apportionment analysis will be included in the regression model as a predictor. We will create separate models using the sources identified by each source apportionment method (APCA, PMF, and PCP).

Temperature and humidity will be included in the model as a combined measure (either apparent temperature or wet bulb temperature), which will be an indicator of temperature as it is perceived by humans. Combining these measures will also allow for assessment of potential effect modification without necessitating the inclusion of multiple interaction terms in the model.

To compare source apportionment methods, we will compare time trends for each identified air pollution source across source apportionment methods. For example, if the component from APCA identified as traffic is higher on weekdays, we expect that the component from PCP identified as traffic will also be higher on weekdays. We will examine the correlations between components/factors identified as the same air pollution source by different source apportionment methods over temporal patterns, including weekdays/weekends and season.

References to date:

1. Hoek G, Krishnan RM, Belen R, Peters A, Ostro B, Brunekreef B, et al. Long-term air pollution exposure and cardio- respiratory mortality: a review. Environ Health. 2013 Dec;12(1):43.

2. Thurston GD, Spengler JD. A quantitative assessment of source contributions to inhalable particulate matter pollution in metropolitan Boston. Atmospheric Environ 1967. 1985 Jan;19(1):9–25.

3. Masiol M, Hopke PK, Felton HD, Frank BP, Rattigan OV, Wurth MJ, et al. Source apportionment of PM2.5 chemically speciated mass and particle number concentrations in New York City. Atmos Environ. 2017 Jan;148:215–29.

4. Gibson EA, Nunez Y, Abuawad A, Zota AR, Renzetti S, Devick KL, et al. An overview of methods to address distinct research questions on environmental mixtures: an application to persistent organic pollutants and leukocyte telomere length. Environ Health. 2019 Dec;18(1):76.

5. Hopke PK, Gladney ES, Gordon GE, Zoller WH, Jones AG. The use of multivariate analysis to identify sources of selected elements in the Boston urban aerosol. Atmospheric Environ 1967. 1976 Jan;10(11):1015–25.

6. Lall R, Ito K, Thurston GD. Distributed Lag Analyses of Daily Hospital Admissions and Source-Apportioned Fine Particle Air Pollution. Environ Health Perspect. 2011 Apr;119(4):455–60.

7. US Environmental Protection Agency. Air Quality System Data Mart, Daily Summary Data: Particulates. Available from: https://www.epa.gov/airdata

8. Menne MJ, Durre I, Korzeniewski B, McNeill S, Thomas K, Yin X, et al. Global Historical Climatology Network - Daily (GHCN-Daily), Version 3 [Internet]. NOAA National Centers for Environmental Information; 2012 [cited 2020 Nov 18]. Available from: https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ncdc:C00861

9. Gary Norris. Positive Matrix Factorization Model for environmental data analyses [Internet]. Available from: https://www.epa.gov/air-research/positive-matrix-factorization-model-environmental-data-analyses

10. Paatero P, Tapper U. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. Environmetrics. 1994 Jun;5(2):111–26.

11. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning [Internet]. New York, NY: Springer New York; 2013 [cited 2020 Nov 16]. (Springer Texts in Statistics; vol. 103). Available from: http://link.springer.com/10.1007/978-1-4614-7138-7

12. Candès EJ, Li X, Ma Y, Wright J. Robust principal component analysis? J ACM. 2011 May;58(3):1–37.

S1b. Signed Structured Thesis Plan Form

MASTER'S THESIS IN EPIDEMIOLOGY
Structured Thesis Plan
IRB, Data Use and Publication Agreement

Section A: Student information (signature needed on last page) Date submitted 11 / 23 / 20

Name: Rachel Tao UNI: rht2112 Phone: 917-817-4729

Academic Advisor: Andrew Rundle, DrPH Advisor UNI agr3

Section B: Thesis Reader Information (First Readers must be EPI Faculty, Second Readers are optional)

First Reader: Andrew Rundle, DrPH UNI: agr3

Second Reader: Marianthi-Anna Kioumourtzoglou, ScD Organizational affiliation MSPH, EHS Dept

Section C: Thesis Data Source Information

Is your thesis related to your practicum ____ yes _x_ no Comments (please describe how) _____

Data is publicly available __Y _x_N If no, Name of data set owner **Amelia Boehme, PhD (Marianthi-Anna Kioumourtoglou, ScD is authorized user)**

Owner Organizational affiliation MSPH Email akb2188@cumc.columbia.edu Phone: _____

Name of data set: Statewide Planning and Research Cooperative System (SPARCS)

Source of data set: NYS Department of Health

Data type (see page 2): ___ data has identifiers ___ de-identified ___ public use data with data use agreement _____ public use data without data use agreement __x__other: Not human subjects (daily city-wide counts)

If international data, country where collected: _____ Collected under local country IRB: ___yes ___no

Section D1: CU IRB information (If data has an IRB, complete next 3 lines and skip to Section E)

Data has existing CU IRB: _x__ yes ___ no Scope of student thesis can be added to existing IRB: ___yes ___no
_x_ not needed as data have been fully aggregated (city-wide daily counts) and data owner has provided verbal permission for use

Title of existing IRB N/A

Existing CU IRB PI Amelia Boehme, PhD IRB number: N/A IRB Expiration date N/A

Section D2 If no existing CU IRB: (If data has an IRB, skip to Section E)

IRB Pre-Screen submitted: ___yes ___no ___Future submission planned Date _____

New IRB protocol submission needed: ___yes ___no ___unknown Date of planned submission _____

If continuing to use data from your practicum, describe IRB ruling for your prescreen: _____

PI on planned IRB submission: _____ PI UNI_____ Phone _____

Need to locate a PI: ___ yes ___ no ___ unknown Comments: _____

DEPARTMENT OF EPIDEMIOLOGY MASTER'S THESIS DATA DESCRIPTION (Skip this section if you have an IRB)

SECTION D3 (continued): Description of data with which the student will be working
(Students with an IRB # may skip this section, all others including those with a prescreen, please complete this section)

Does any member of the research team (Student, first reader, second reader or data set owner) have any of the following information on observations in the data set? If "Yes" or "Unknown", please specify below.

Yes	No	Unknown	
	x		Participant names
	x		Telephone numbers, fax numbers, email addresses
	x		Social security numbers
	x		Medical record numbers, health plan beneficiary numbers, account numbers, certificate/license numbers
	x		Vehicle identifiers and serial numbers, including license plate numbers
	x		Device identifiers and serial numbers
	x		Web universal resource locators (URLs) or Internet protocol (IP) address numbers
	x		Biometric identifiers, including fingerprints and voiceprints, or full-face photographic images or comparable images
	x		Any other unique identifying number, characteristic/code, unless otherwise permitted by the Privacy Rule for re-identification
	x		Any geographic subdivisions smaller than a state, including street address, city, county, precinct, ZIP code, and their equivalent geographical codes, except for the initial three digits of a ZIP code if: -The geographic unit formed by combining all ZIP codes with the same three initial digits contains more than 20,000 people -The initial three digits of a ZIP code for all such geographic units containing 20,000 or fewer people are changed to 000
	x		Any elements of dates (except year) for dates directly related to an individual, including birth date, admissions date, discharge date, date of death; and all ages over 89 and all elements of dates (including year) indicative of such age, except that such age and elements may be aggregated into a single category of age 90 or older

If you answered "Yes" or "Unknown" to any of the above questions, please specify and describe:

The version of the SPARCS dataset I will be using is aggregated for total MI admissions per day in NYC, and thus no longer considered human subjects. I will also be using AQS data from the EPA and meteorological data from NOAA, all of which is fully publicly available and includes no health or human-related information

Section E: Attach your structured thesis plan narrative (2-3 pages, single spaced, 11 or 12 point font):

See attached

Section F: Data Use

As a ***data set owner/authorized user***, I have the authority to grant or verify permission to use this data for the purposes described in this proposal. The student has permission to analyze the data, to prepare a master's thesis as described in the thesis plan and to prepare manuscript(s) for publication based on the analyses described in the structured thesis plan, subject to the review and revisions of the reader(s). If requested, students must submit analytic computer code to the first reader.

Please detail other/additional data use conditions *(add additional pages as needed)*_____

Section G: Publication

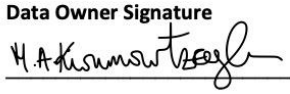
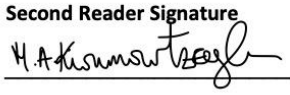
As a ***reader,*** I agree to supervise this student's work on their Department of Epidemiology master's thesis. ***I consider the project described to be valid as master's level work and to utilize methods consistent with the students training or I will be responsible for supervising the student in methods I recommend that is beyond the student's masters level training .*** With regard to publication of the thesis, unless otherwise stipulated below, the student will be listed as first author with readers/data set owners that contribute to the analytic strategy or interpretation of findings listed as coauthors[1]. The corresponding author is the senior author unless otherwise stipulated below.

Readers may set a date during which the student must have made satisfactory revisions or become a co-author rather than first author of any manuscript resulting from the thesis work. This date is considered unlimited if not specified.

From the beginning date of May 15, 2021, the student has: __3 mos _x_ 6 mos __1 year __ other timeframe to complete and submit a manuscript to a journal selected by mutual agreement.

Please detail other publication/authorship conditions: *(add additional pages as needed)* Otherwise, a collaborator may be assigned to completes and submits the manuscript, and this person will be listed as an additional first author on any publication resulting from this work.

I agree to abide by the stipulations described above.

Student signature	Date
	11/19/20
First Reader Signature	Date
	11/23/2020
Data Owner Signature	Date
	11.19.2020
Second Reader Signature	Date
	11.19.2020

1 The thesis team does not enter into discussions on authorship arrangements except to suggest that they be specified apriori. Form Version 09/21/2017

S2. Quasi-Systematic Literature Review

Author (pub. year/data years)	Title	Journal	Population of interest / Sample characteristics	Location	Study design	Statistical methods	Outcome measure(s)	Exposure(s)	Covariates	Major findings
GBD 2019 Risk Factors Collaborators (2019/1990-2019)	Global burden of Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019	*Lancet*	global population	international	meta-analysis	Bayesian meta-regression methods	mortality, years of life lost, years of life lived with disability, disability-adjusted life-years	87 risk factors and combinations of risk factors		6.67M global deaths attributable to air pollution
Hoek et al. 2013	Long-term air pollution exposure and cardio-respiratory mortality: a review	*Environmental Health*	N/A	USA	meta-analysis	N/A	all-cause, cardiovascular, and respiratory mortality	PM2.5	smoking, sex, education (BMI, diabetes, road traffic noise)	identify relevant covariates - spatial issues - period issues
Liu et al. 2019	Ambient Particulate Air Pollution and Daily Mortality in 652 Cities	*NEJM*	global	USA	meta-analysis	overdispersed generalized additive models with random-effects meta-analysis	all-cause mortality, cardiovascular mortality, respiratory mortality	PM2.5, PM10		independent associations between short-term exposure to PM10 and PM2.5 and daily all-cause, cardiovascular, and respiratory mortality

Farhadi et al. 2020	Association between PM2.5 and risk of hospitalization for myocardial infarction: a systematic review and a meta-analysis	*BMC Public Health*	global	global	meta-analysis	I^2	MI hospitalization	PM2.5		there is an association between PM2.5 and MI hospitalization
Brook et al. 2010 (American Heart Association)	Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the American Heart Association	*Circulation*	N/A	USA	review	N/A	cardiovascular disease	PM2.5		exposure to PM2.5 over a few hours to weeks can trigger cardiovascular disease-related events, longer-term exposure increases risk for cardiovascular mortality, reductions in PM levels are associated with decreases in cardiovascular mortality within a time frame as short as a few years
Newby et al. (2015)	Expert position paper on air pollution and cardiovascular disease	*European Heart Journal*	N/A	internation al	review	N/A	cardiovascular disease	air pollution		Biological mechanisms for effects of air pollution on cardiovascular disease include inflammation

Weichenthal et al. (2016/2004-2011)	Ambient PM2.5 and risk of emergency room visits for myocardial infarction: impact of regional PM2.5 oxidative potential: a case-crossover study	*Environmental Health*	population of Ontario, CA	16 cities in Ontario, CA	case-crossover	conditional logistic regression	MI hospitalization	regional PM2.5	oxidative potential	regional PM2.5 oxidative potential may modify the relationship between PM2.5 and MI
Dai et al. (2014/2000-2006)	Associations of Fine Particulate Matter Species with Mortality in the United States: A Multicity Time-Series Analysis	*Environmental Health Perspectives*	population of US	75 US cities	time-series	Poisson regression	all-cause, cardiovascular, MI, stroke, and respiratory mortality	speciated PM2.5	infiltration rate, county-level smoking, alcohol	increased risk of mortality with PM2.5, varies by season and species
Davoodabadi et al. (2019/2010-2012)	Correlation between air pollution and hospitalization due to myocardial infarction	*ARYA Atherosclerosis*	MI patients	Isfahan, Iran	case-crossover	conditional logistic regression	MI hospitalization	PM10, PM2.5, NO2, SO2, CO, O3	temperature, windspeed, humidity	elevated PM2.5 48 hours before admission
Thurston et al. (2011/2000-2005)	A source apportionment of U.S. fine particulate matter air pollution	*Atmos Environ*	N/A	USA	source apportionment	factor analysis, APCA	N/A	speciated PM2.5		Identified metals industry, crustal/soil particles, traffic, steel industry, coal combustion, oil combustion, salt and biomass burning sources

Ito et al. (2004/2001-2002)	Spatial variation of PM2.5 chemical species and source-apportioned mass concentrations in New York City	*Atmos Environ*	N/A	NYC	source apportionment	APCA	N/A	speciated PM2.5		sources: secondary aerosols, soil, traffic-related, residual oil
Hopke et al. (1976/1970)	The use of multivariate analysis to identify sources of selected elements in the Boston urban aerosol	*Atmos Environ*	N/A	Boston	source apportionment	factor analysis	N/A	speciated PM2.5		sources: crustal dust, sea salt, residual fuel burning, traffic, refuse incineration
Lall et al. (2011/2001-2002)	Distributed Lag Analyses of Daily Hospital Admissions and Source-Apportioned Fine Particle Air Pollution	*Environmental Health Perspectives*	adults 65+	NYC	time-series	distributed-lag GLM, PMF	all-cause, respiratory and cardiovascular Medicare admissions	source-specific PM2.5	same-day and lagged temperature and relative humidity, season, winter flu, day of week	traffic-related PM2.5 associated with CVD admissions
Wurth et al. (2018/2004-2012)	Fine particle sources and cognitive function in an older Puerto Rican cohort in Greater Boston	*Environmental Epidemiology*	older Puerto Rican adults	Boston	longitudinal study	linear mixed model with random intercept	cognitive function	PM2.5	age, sex, season, physical activity, education, income-to-poverty ratio	long term exposures to BC and nickel, tracers of traffic and oil combustion associated with decreased cognitive function

Hopke et al. (2006/1995-1998)	PM source apportionment and health effects: 1. Intercomparison of source apportionment results	*Journal of Exposure Science and Environmental Epidemiology*	N/A	Washington DC, Phoenix, AZ	source apportionment	PCA, Unmix, PMF	N/A	speciated PM2.5		overall good agreement between statistical methods, less good correlation with respect to different traffic-related components
Kioumourtzoglou M-A, et al. (2014/2003-2010)	The impact of source contribution uncertainty on the effects of source-specific PM2.5 on hospital admissions: A case study in Boston, MA	*Journal of Exposure Science and Environmental Epidemiology*	Medicare enrollees	Boston, MA	case-crossover using time-stratified approach	PMF, APCA, time-series	CVD hospital admissions	source-specific PM2.5	same-day temp, same-day dew point, 2-day moving average temp, PM2.5	short-term assoc with mobile and regional, long-term assoc with residual oil
Thurston and Spengler (1985)	A quantitative assessment of source contributions to inhalable particulate matter pollution in metropolitan Boston	*Atmospheric Environment*	N/A	Boston	source apportionment	PCA	N/A	source-specific PM2.5		sources: sulfur, coal, crustal
Masiol et al. (2017/2005-2016)	Source apportionment of PM2.5 chemically speciated mass and particle number concentrations in NYC	*Atmospheric Environment*	NYC metro area (queens college monitor) June 2009 - July 2010	NYC	source apportionment	PMF	N/A	speciated PM2.5	wind direction	9 factor solution

Gibson E, Nunez Y, et al. (2019/2001-2002)	An overview of methods to address distinct research questions on environmental mixtures: an application to persistent organic pollutants and leukocyte telomere length	*Environmental Health*	US adults, NHANES	USA	cross-sectional	clustering, PCA, EFA, variable selection, weighted quantile sum regression, BKMR	leukocyte telomere length	persistent organic pollutants	age, sex	suitable statistical method depends on research question (methods paper primarily)
Rich et al. (2019/2005-2016)	Triggering of cardiovascular hospital admissions by source specific fine particle concentrations in urban centers of New York State	*Environ Int*	population of NYS	New York State	time-stratified case-crossover design	PMF, conditional logistic regression	cardiac arrhythmia, ischemic stroke, congestive heart failure, ischemic heart disease, MI	source-specific PM2.5	temperature, relative humidity	traffic-related, road dust associated with MI
Squizzato et al. (2018/2005-2016)	PM2.5 and gaseous pollutants in New York State during 2005–2016: Spatial variability, temporal trends, and economic influences	Atmospheric Environment	N/A	NYS	looked at overall trends in emissions compared with economic trends	inverse distance weighting for spatial interpolation	N/A	N/A		tends in PM2.5 and gaseous pollutants line up with economic trends
Schaap et al. (2004)	Artefacts in the sampling of nitrate studied in the "INTERCOMP" campaigns of	*Atmospheric Environment*	N/A	N/A	bench research		N/A	N/A		stability of nitrate at low temperatures

	EUROTRAC-AEROSOL									
Pinder et al. (2008)	Observable indicators of the sensitivity of PM2.5 nitrate to emission reductions—Part I: Derivation of the adjusted gas ratio and applicability at regulatory-relevant time scales	*Atmospheric Environment*	N/A	Eastern US		chemical transport model	PM2.5 nitrate	NH3, SO2, NOx emissions	temperature	successfully identified robust indicators for estimating sensitiv9ity of PM2.5 nitrate to SO2 and NH3 emission changes
Sun et al. (2018)	Effects of ambient temperature on myocardial infarction: A systematic review and meta-analysis	*Environmental Pollution*	studies of MI and temperature	international	meta-analysis	I2, Cochran's Q test	MI	ambient temperature		heat exposure and cold exposure both associated with MI. same-day heat exposure, lagged cold exposure
Zhou et al. (2019/2016-2017)	Ambient Ammonia Concentrations Across New York State	*JGR Atmospheres*	N/A	NYS		conditional bivariate probability function		ammonia		rural sites more affected by transported ammonia, urban sites more affected by ammonia from traffic
Thiruvengadam et al. (2016)	Unregulated greenhouse gas and ammonia emissions from current technology heavy-duty vehicles	*J Air Waste Manag Assoc*	N/A	N/A						ammonia and nitrous oxide can be byproducts of modern heavy-duty diesel and natural gas engines

Galan Madruga et al. (2018/2014-2015)	Characterization and Local Emission Sources for Ammonia in an Urban Environment	*Bulletin of Environmental Contamination and Toxicology*	N/A	Madrid		k-means clustering		ammonia		traffic, garbage, and sewage sites are sources of ammonia in this urban study
Sun et al. (2014/2012)	On-Road Ammonia Emissions Characterized by Mobile, Open-Path Measurements	*Environmental Science & Technology*	N/A	NJ and California	measure MH3 and CO using car sensors					on-road NH3 emissions may be underestimated in many studies
Squizzato et al. (2018/2005-2016)	A long-term source apportionment of PM2.5 in New York State 2005-2016	*Atmospheric Environment*	New York State	New York State	source apportionment	PMF	N/A	speciated PM2.5		sources: secondary sulfate, secondary nitrate, OP-rich, gas, diesel vehicles, road dust, biomass burning, fresh, aged sea salt, residual oil
Dutkiewiicz et al. (2004/2001-2002)	Sources of fine particulate sulfate in New York	*Atmospheric Environment*	N/A	NYS	trends in sulfate	backward air trajectories				highest concentrations were associated with air masses that pass through the Ohio River Valley and Great Lakes Basin
Hopke et al. (2005)	Reconciling Trajectory Ensemble Receptor Model Results with Emissions	*Environmental Science & Technology*	N/A	Underhill, VT, Brigantine, NJ		residence time analysis				identified the Ohio River Valley and Great Lakes Basin as areas of highest probability for coal-fired power plants
Amato et al. (2010)	Concentrations, sources and geochemistry of airborne particulate matter at a	*J Environ. Mont.*	N/A	Barcelona	source apportionment	PCA				sources: traffic/industrial, mineral/works, sea salt, secondary, biomass

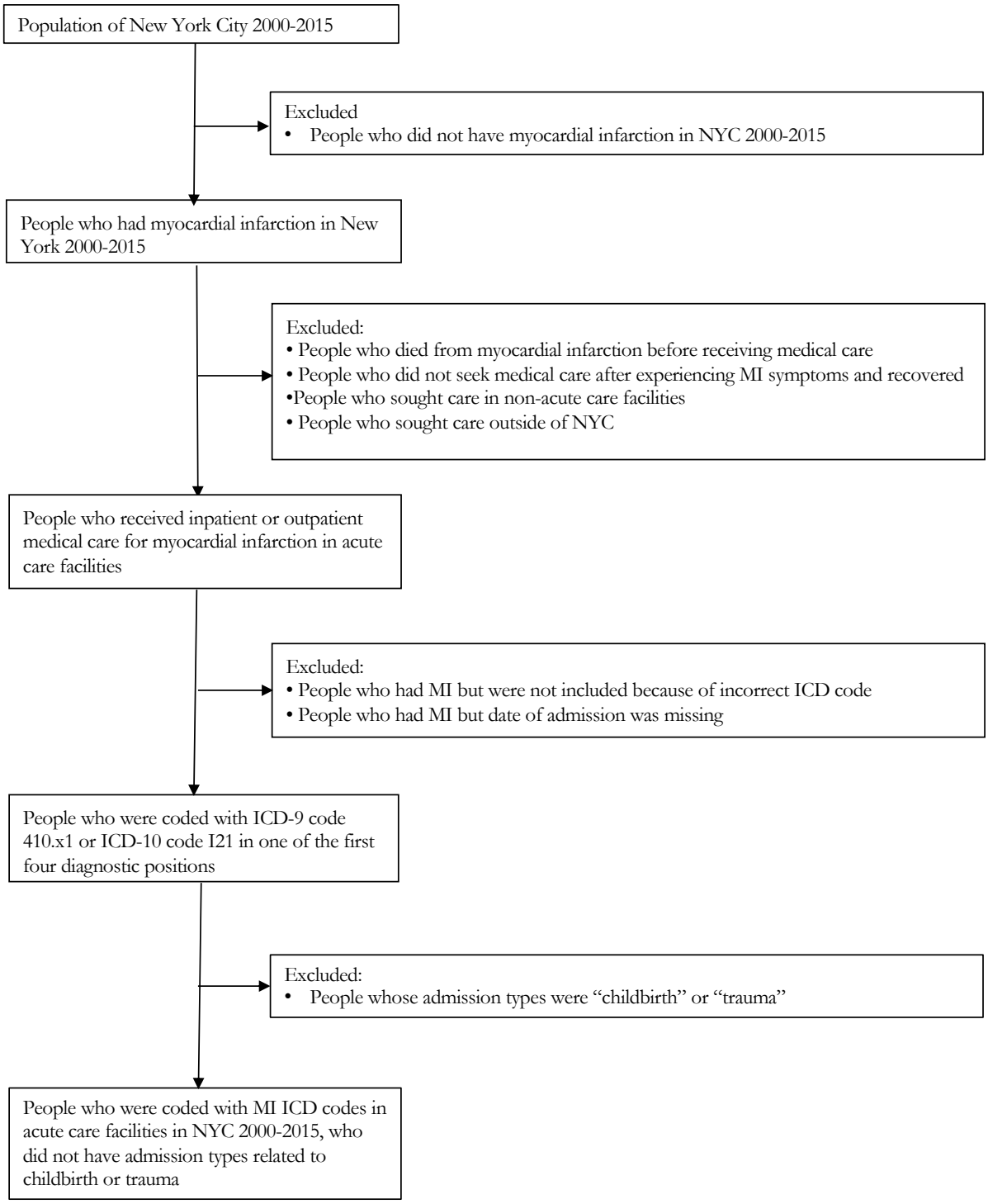
	major European airport									
Thorpe and Harrison (2008)	Sources and properties of non-exhaust particulate matter from road traffic: A review	*Science of the Total Environment*	N/A	N/A	review	N/A				road dust can have many different sources (road wear, break wear, etc.)
Peltier et al. (2008)	Residual oil combustion: a major source of airborne nickel in New York City	*Journal of Exposure Science and Environmental Epidemiology*	N/A	NYC	source tracers	time series at different sites				residual oil pollution is highest in NYC, particularly in winter
Ostro et al. (2007/2000-2003)	he Effects of Components of Fine Particulate Air Pollution on Mortality in California: Results from CALFINE	*Environmental Health Perspectives*	adults 65+	California	time-series	Poisson regression with natural splines	all-cause mortality, respiratory mortality, cardiovascular mortality	PM2.5 and 19 other species	temperature, day of week, relative humidity	PM2.5 mass, EC, OC, and nitrate all associated with cardiovascular deaths at different lags
Cao et al. (2012/2004-2008)	Fine Particulate Matter Constituents and Cardiopulmonary Mortality in a Heavily Polluted Chinese City	*Environmental Health Perspectives*	population of Xi'an	Xi'an, China	time-series	Poisson regression with natural splines	total, cardiovascular, and respiratory mortality	PM2.5 constituents	day of week, temporal trend, temperature, SO2 and NO2 concentrations	positive associations with OC, EC, ammonium, nitrate, chlorine, and nickel for at least 1 lag day with mortality
Zhang et al. (2018/2007-2016)	Triggering of cardiovascular hospital admissions by fine particle concentrations in New York state: Before,	*Environmental Pollution*	Residents living within 15 miles of $PM_{2.5}$ monitoring sites in 5 major urban	NYS	Case-crossover	Conditional logistic regression	Total CVD, nine specific subtypes	$PM_{2.5}$	Temp, RH	Pollutant concentration and CVD admission rates decreased after admission changes, but $PM_{2.5}$ mass was still associated

	during, and after implementation of multiple environmental policies and a recession		centers in NYS							with higher rate of IHD events
Ostro et al. (2010/2002-2007)	Long-Term Exposure to Constituents of Fine Particulate Air Pollution and Mortality: Results from the California Teachers Study	*Environmental Health Perspectives*	former female school professionals in California	California	cohort study	Cox	all-cause, cardiopulmonary, and IHD mortality	PM2.5 constituents	16 individual-level covariates (risk factors)	long-term exposure to PM2.5 constituents increases mortality risk in several areas, particularly those associated with fossil fuels and crustal origin
Badaloni et al. (2017)	Effects of long-term exposure to particulate matter and metal components on mortality in the Rome longitudinal study	*Environ Int*	adults 30+	Rome	cohort study	Cox	non-accidental, CVD, and IHD mortality	PM10, PM2.5, PM2.5 constituents	sex, DOB, marital status, place of birth, education, occupation, SEP index	vehicular exhaust, non-tailpipe emissions and mixed oil burning/industry play a role in mortality
Franklin et al. (2008/2000-2005)	The role of particle composition on the association between $PM_{2.5}$ and mortality	*Epidemiology*	Residents of 25 US communities	US	Time-series	Poisson	Non-accidental death	Total $PM_{2.5}$, select constituents	Temperature, season	Temperature is a reasonable surrogate for ventilation, aluminum and silicon are associated with mortality
Mohammad et al. (2018/1998-2013)	Association of weather with day-to-day incidence of myocardial infarction: a SWEDEHEART nationwide	*JAMA Cardiol*	Population of Sweden	Sweden	Time-series	Poisson	MI	Air temperature, wind velocity, sunshine duration, air pressure,		Low air temp, low air pressure, high wind velocity, shorter sunshine duration associated with risk of MI

	observational study							air humidity, snow precip, rain precip, change in air temp		
Thurston GD,…Lall R…Arden Pope C (2016, 1982-2004)	Ischemic heart disease mortality and long-term exposure to source-related components of US fine particulate air pollution	*Environmental Health Perspectives*	adults in 100 US metropolitan areas	USA	nationwide cohort study	APCA (orthogonal), cox proportional hazards	cardiovascular disease	PM2.5	42, incl: smoking, BMI, 'occupational dirtiness index', marital status, education	coal combustion IHD HR = 1.5 (CI = 1.02, 1.08), traffic borderline, crustal, soil and biomass not associated with IHD
Huang et al. (2012/2004-2008)	Seasonal Variation of Chemical Species Associated with Short-Term Mortality Effects of PM2.5 in Xi'an, a Central City in China	*Am J Epidemiol*	population of Xi'an	Xi'an, China	time-series	Poisson regression	all-cause and cause-specific mortality	PM2.5 and PM2.5 constituents	temperature and relative humidity	secondary components (sulfate/ammonium), combustion species (EC, sulfur chlorine) and transition metals (chromium, lead, nickel, zinc) appeared responsible for most increased risk in mortality
Mills et al. (2007)	Ischemic and Thrombotic Effects of Dilute Diesel-Exhaust Inhalation in Men with Coronary Heart Disease.	*NEJM*	men with coronary heart disease	N/A	randomized, double-blind, crossover study	t-tests	myocardial, vascular, and fibrinolytic function	diesel exhaust		greater increase in ischemic burden during exposure to diesel exhaust compared to just exercise
Mills et al. (2005)	Diesel Exhaust Inhalation Causes Vascular Dysfunction	*Circulation*	healthy men	N/A	double-blind, randomized, cross-over study	ANOVA, t-test	bilateral forearm blood flow and inflammatory factors	diesel exhaust		inhalation of dilute diesel exhaust impairs 2 important and complementary

	and Impaired Endogenous Fibrinolysis									aspects of vascular function in humans: the regulation of vascular tone and endogenous fibrinolysis
Yang et al. (2020/2011-2013)	Fine particulate matter constituents and cause-specific mortality in China: A nationwide modelling study	*Environ Int*	population of China	China	time-series	quasi-Poisson regression with polynomial distributed lags	cause-specific mortality	PM2.5 constituents: OC, EC, sulfate, nitrate, ammonium	stratification by region, gender, age group, education level	EC, OC, sulfate, nitrate, and ammonium all associated with mortality at lag 0-3, particularly CVD and MI
Peterson et al. (2020/1990-2010)	Impact of Reductions in Emissions from Major Source Sectors on Fine Particulate Matter–Related Cardiovascular Mortality	Environmental Health Perspectives	population of USA	USA	difference in difference	linear regression models	CVD mortality	PM2.5 and PM22.5 constituents	median household income, percent nonwhite pop, population	reductions in sulfur-dioxide emissions from large point sources and nitrates and EC emissions from mobile sources contributed to largest reduction in PM2.5 related mortality rates respectively

S3. Study Population Flow Diagram



Note: Reinfarctions and recurrent MI admissions were included, except readmissions that took place within two days after a previous MI admission for that patient. This means that the final sample size of 964,606 MI admissions (1044 excluded because of missing date of admission) does not reflect the sample population, but the sample size for MI count.

S4. Thesis Reader and Data Description Form

P9419 Master's Essay in Epi
Thesis Reader and Data Description Form
Writable Version 9/21/2017

Student Name: Rachel Tao UNI: rht2112

You will need to save this document to your computer before completing

==========READER INFORMATION==========

First Reader: Andrew Rundle UNI: agr3

Second Reader (Optional): Marianthi-Anna Kioumourtzoglou Email: mk3961@cumc.columbia.edu

Academic advisor: Andrew Rundle UNI: agr3

==========THESIS TOPIC==========

Research topic to be investigated: Principal Component Pursuit as a novel approach to assessing environmental mixtures as a health exposure for myocardial infarction – this analysis will use publicly available Air Quality System (AQS) air pollution data for NYC from the EPA, as well as SPARCS data for myocardial infarction outcome data
Comments: This is an update from my previous topic about CSO and green infrastructure, which I will no longer be doing.

==========DATA SET==========

I have a data set: ☒ Yes ☐ No ☐ Possibly

Dataset name: SPARCS, AQS (Air Quality Systems)

My dataset is: ☒US only ☐International (Specify country: Your response here)

Approximate sample size: unknown (Write "unknown" if dataset is not in hand and you do not know)

My dataset is from:

☐ My first reader: (Name) Your response here
☒ My second reader: (Name) Marianthi-Anna Kioumourtzoglou
☒ Another CUMC/Mailman professor: (Name) Amelia Boehme
☐ Another Columbia University source: Your response here
☒ A publicly available database (Specify source): US EPA
☐ NYC DOHMH
☒ Other (Specify source): Not yet known

The thesis dataset is:

☒ In my possession now
☐ Not yet in my possession
☐ Expected to be in my possession by: Your response here
☐ Other (please describe): Your response here

1

If you do not have a dataset, describe what kind of data you are looking for:
I am already able to access publicly available information on air pollution in NYC through the AQS database on the EPA website. I gained access to the SPARCS dataset through my 2nd reader, who previously had access to it. I was given a fully deidentified version of the dataset with no geographic identifiers by another member of my 2nd reader's lab (Sebastian Rowland) who uses the SPARCS data for the outcome I want to look at (myocardial infarction).

Continued

For additional comments, please use the space provided on page 3.

==========IRB INFORMATION==========

Does your data have and of the following attributes? (Please circle all that apply—Please do not leave any blank)

Y	N	U	(Y=Yes N=No U=Unknown)
☐	☒	☐	Participant names
☐	☒	☐	Telephone numbers
☐	☒	☐	Facsimile numbers
☐	☒	☐	Electronic mail addresses
☐	☒	☐	Social security numbers
☐	☒	☐	Medical record numbers
☐	☒	☐	Health plan beneficiary numbers
☐	☒	☐	Account numbers
☐	☒	☐	Certificate/license numbers
☐	☒	☐	Vehicle identifiers and serial numbers, including license plate numbers
☐	☒	☐	Device identifiers and serial numbers
☐	☒	☐	Web universal resource locators (URLs)
☐	☒	☐	Internet protocol (IP) address numbers
☐	☒	☐	Biometric identifiers, including fingerprints and voiceprints
☐	☒	☐	Full-face photographic images or comparable images
☐	☒	☐	Any other unique identifying number, characteristic/code, unless otherwise permitted by the Privacy Rule for re-identification

☒	☐	☐	Any geographic subdivisions smaller than a state, including street address, city, county, precinct, ZIP code, and their equivalent geographical codes, except for the initial three digits of a ZIP code if: ▪ The geographic unit formed by combining all ZIP codes with the same three initial digits contains more than 20,000 people ▪ The initial three digits of a ZIP code for all such geographic units containing 20,000 or fewer people are changed to 000
☒	☐	☐	Any elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death; and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of 90 or older

Dataset has existing CUMC IRB: ☐Yes ☐No ☒Don't know

IRB number: Your response here

Data was used during/is from my practicum ☐Yes ☒No

Practicum went to IRB prescreen: ☐Yes ☐No

Prescreen ruling: Your response here

I do not have an existing IRB, but believe that my thesis will be is:

☒Exempt from IRB

☐IRB Modification

☐Expedited IRB

☐Standard IRB

☐Don't Know

OTHER: Please list the types of bibliographic software that you use:

	Never	Basic	Intermediate	Advanced
EndNote	☐	☐	☒	☐
Reference Manager	☒	☐	☐	☐
RefWorks	☒	☐	☐	☐
Zotero	☐	☐	☒	☐

Others: (please specify): Your response here

Please list anything else you would like us to know: I plan to do a sub-analysis of ongoing research that I am doing with Professor Kioumourtzoglou, on the topic of using

Principal Component Pursuit as a novel approach to assessing environmental mixtures as a health exposure – this analysis will use publicly available Air Quality System (AQS) air pollution data for NYC from the EPA, as well as SPARCS data for myocardial infarction outcome data. I was hoping to already have the data in-hand by today, but because of meeting scheduling with collaborators, I will need to clarify the IRB and data privacy questions next week.

Continued
Continued
Continued
Continued
Continued
Continued
Continued

Student Name: Rachel Tao

Student Signature: Rachel Tao