Pollutant composition modification of the effect of air pollution on progression of coronary artery calcium

The Multi-Ethnic Study of Atherosclerosis

Joshua P. Keller*, Timothy V. Larsonb, Elena Austinc, R. Graham Barred, Lianne Shepparde, Sverre Vedalc, Joel D. Kaufmanc, Adam A. Szpirof

Background: Differences in traffic-related air pollution (TRAP) composition may cause heterogeneity in associations between air pollution exposure and cardiovascular health outcomes. Clustering multipollutant measurements allows investigation of effect modification by TRAP profiles.

Methods: We measured TRAP components with fixed-site and on-road instruments for two 2-week periods in Baltimore, Maryland. We created representative TRAP profiles for cold and warm seasons using predictive k-means clustering. We predicted cluster membership for 1005 participants in the Multi-Ethnic Study of Atherosclerosis and Air Pollution with follow-up between 2000 and 2012. We estimated cluster-specific relationships between coronary artery calcification (CAC) progression and long-term exposure to fine particulate matter ($PM_{2.5}$) and oxides of nitrogen (NO$_x$).

Results: We identified two clusters in the cold season, notable for higher ratios of gases and ultrafine particles, respectively. A 5-μg/m$^3$ difference in $PM_{2.5}$ was associated with 17.0 (95% confidence interval [CI] = 7.2, 26.7) and 42.6 (95% CI = 25.7, 59.4) Agatston units/year CAC progression among participants in clusters 1 and 2, respectively (effect modification $P = 0.006$). A 40 ppb difference in NO$_x$ was associated with 22.2 (95% CI = 7.7, 36.7) and 41.9 (95% CI = 23.7, 60.2) Agatston units/year CAC progression in clusters 1 and 2, respectively ($P = 0.08$). Similar trends occurred using clusters identified from warm season measurements. Clusters correlated highly with baseline pollution level.

Conclusions: Clustering TRAP measurements identified spatial differences in composition. We found evidence of greater CAC progression rates per unit $PM_{2.5}$ exposures among people living in areas characterized by high ratios of ultrafine particle counts relative to NO$_x$ concentrations.

Introduction

The relationship between exposure to traffic-related air pollution (TRAP) and cardiovascular morbidity and mortality has been well established. TRAP is a complex mixture of many different particulate and gaseous components that can vary across fine spatial scales and is of particular interest in urban areas. Our focus in this study is on directly emitted TRAP whose composition and concentration levels vary across an urban area with traffic intensity, distance to roadway, fuel type (heavy duty diesel versus gasoline), age and condition of vehicles, and urban area with traffic intensity.

Copyright © 2018 The Authors. Published by Wolters Kluwer Health, Inc. on behalf of Environmental Epidemiology. All rights reserved. This is an open access article distributed under the Creative Commons Attribution License 4.0 (CCBY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Environmental Epidemiology (2018) 2:e0024

Received: 28 January 2018; Accepted: 1 June 2018

Published online 9 July 2018

DOI: 10.1097/EE9.0000000000000024

What this study adds

This article presents a novel investigation of how differences in traffic-related air pollution may modify the relationship between long-term particulate matter exposure and cardiovascular health. A distinctive feature is the use of near- and on-road pollutant measurements to inform predictions of pollution profiles using modern statistical clustering methods. This study provides evidence of greater health impact of exposure to pollution, with higher proportions of ultrafine particles within metropolitan regions.
Complex multipollutant datasets are often analyzed using dimension reduction techniques, which simplify the complex variability of the data into a smaller set of features. Clustering methods, which provide a promising approach for understanding multipollutant health effects, partition multipollutant observations into a prespecified number of groups or clusters. This provides a categorical division of the data based on pollutant profile that simplifies the interpretation of simultaneous exposure to multiple pollutants. In the popular “k-means” algorithm, clusters are selected to minimize the (Euclidean) distance between each observation and the center, or representative exposure vector, of its assigned cluster. For analyses of administrative data, records can be assigned to a cluster based on city. For cohort studies, a classification model can be used to predict cluster membership at subject residences. Cluster membership for each subject or record can then serve as an effect modifier for an association between a single exposure and membership for each subject or record can then serve as an effect modifier for an association between a single exposure and outcome. This allows for heterogeneity in the association between a single composite pollutant (e.g., NOX, fine particulate matter [PM2.5]) and health outcomes to be identified across groups distinguished by predicted differences in pollution composition at subject locations.

Prior modeling of spatial variation in TRAP across cities has focused on separate land-use regression models for ultrafine particles and volatile organic compounds (VOCs). While these models can provide estimates of associations with differences in level of a single TRAP component, the single-component approach cannot capture the mixture features identifiable from considering variation in multiple TRAP components simultaneously as can be done through clustering.

The Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air) investigated the association between long-term air pollution exposure and progression of subclinical measures of atherosclerosis, including coronary artery calcification (CAC). Using spatiotemporal predictions of individual pollutants, Kaufman et al. found an association between CAC progression and higher levels of NOX and PM2.5 exposure. Recent supplemental monitoring campaigns have obtained an expansive suite of multipollutant measurements in MESA Air cities. These observations provide the opportunity to explore variation in CAC progression due to differences in TRAP composition.

Other socioeconomic factors include income, education, and smoking status.

### Table 1

Baseline characteristics of the study population together and stratified by predicted membership in cold and warm season clusters

| Characteristics          | Full cohort | Cluster 1 | Cluster 2 | Cluster 3 |
|--------------------------|-------------|-----------|-----------|-----------|
| Age (y)                  | 63 ± 10     | 63 ± 9.8  | 64 ± 10   | 63 ± 9.8  |
| Male                     | 471 (47)    | 357 (48)  | 114 (44)  | 357 (48)  |
| Ethnicity                |             |           |           |           |
| White                    | 498 (50)    | 373 (50)  | 125 (48)  | 379 (54)  |
| Black                    | 507 (50)    | 370 (50)  | 137 (52)  | 322 (46)  |
| Income ($1,000)          | 53 ± 35     | 58 ± 34   | 39 ± 34   | 59 ± 34   |
| Normalized SES indexa    | 0.16 ± 0.73 | -0.29 ± 0.30 | 0.83 ± 0.78 | -0.36 ± 0.87 |
| General health           |             |           |           |           |
| BMI                      | 29 ± 5.8    | 29 ± 5.7  | 30 ± 6.0  | 29 ± 5.7  |
| Cholesterol (mmol/L)     | 192 ± 36    | 193 ± 35  | 192 ± 37  | 193 ± 35  |
| HDL (mmol/L)             | 51 ± 15     | 52 ± 15   | 50 ± 16   | 51 ± 14   |
| Triglycerides (mmol/L)   | 118 ± 72    | 117 ± 72  | 122 ± 72  | 119 ± 74  |
| CAC score                | 186 ± 462   | 170 ± 455 | 232 ± 478 | 176 ± 463 |
| Pollutant exposure       |             |           |           |           |
| NOX (ppb)                | 42 ± 12     | 36 ± 7.4  | 58 ± 7.2  | 36 ± 7.6  |
| PM2.5 (µg/m³)            | 15.9 ± 0.80 | 15.6 ± 0.52 | 16.8 ± 0.72 | 15.5 ± 0.51 |

Values given as mean ± standard deviation or n (%).
aUnitless scale. Lower values indicate greater affluence.

BMI indicates body mass index; HDL, high-density lipoprotein; SES, socioeconomic status.
participants at a 2-week time scale. Cross-validated R² for the models indicated good out-of-sample prediction accuracy (R² of 0.92 and 0.84, for NOX and PM₁.₅, respectively).²³ For modeling CAC, NOX and PM₁.₅ exposure predictions were aggregated into long-term averages from recruitment through exam visit, based on participant residential history.

**Multipollutant TRAP data**

Pollution measurements were made at 43 intersections within the Baltimore, Maryland, metropolitan area during 2-week periods in February 2012 and June 2012, which we refer to as the cold and warm seasons, respectively. Measurements were made in two seasons to capture differences in TRAP composition due to changes in sources and meteorology throughout the year. Most of the intersections were in residential areas and not on major roads, given the goal to characterize TRAP profiles relevant to subjects’ ambient exposures. Measurements of nitrogen dioxide (NO₂), oxides of nitrogen (NOX), ozone (O₃), and specific VOCs were made using stationary badge monitors (see Table 2). Carbon monoxide (CO) concentrations and particle number (PN) counts for different-size bins were measured using an on-road mobile platform at a collection of locations in and near the intersection by traversing the blocks bordering the intersection of interest. The smallest-size bin for PN counts captured particles 25 to 400 nm in diameter. Riley et al.²⁶ provide a detailed description of the sample collection procedures. The mobile measurements were made during the afternoon commuting period and were adjusted for day-to-day variability by subtracting the fifth percentile of each pollutant and taking the median value of all adjusted observations at each location.²⁶

We scaled observations from both the badge and mobile platforms by the measured badge NO₂ concentration at the respective location. Standardization by NO₂ level allows clusters derived from the data to be informed by relative pollution composition and not solely by overall level. We then standardized these pollution fractions to have mean zero and unit standard deviation.

**Clustering**

Using the predictive k-means method,¹⁵ we clustered the component species fractions separately by season. The predictive k-means method identifies cluster centers by simultaneously optimizing the deviation of the multipollutant observation from its assigned cluster center and the assignment of each monitor location to a cluster based on prediction variables, which are typically a function of spatial location. This results in clusters whose membership can be better predicted at subject locations compared to clusters from traditional k-means clustering, which does not incorporate prediction variables when identifying cluster centers. The predictive k-means method is implemented in the publicly available R package “kmeans,” and additional technical details are provided in eAppendix 1; http://links.lww.com/EE/A17.

We used a large collection of geographic variables containing land use information, distance to roadways, emissions inventories, and other values derived from Geographic Information Systems (GIS). These covariates, listed in eTable 1; http://links.lww.com/EE/A17, have been used previously to develop spatial and spatiotemporal prediction models.²⁹⁻³² We reduced the values of these geographic variables at monitor locations to a small set of principal component analysis (PCA) scores, which we included as prediction variables for modeling cluster assignment predictive k-means.

We chose the number of clusters and the number of PCA scores by 10-fold cross-validation (CV). We limited the models to between 2 and 5 clusters and 1 and 6 PCA scores. Models were compared according to their mean squared prediction error (MSPE), which is the sum of the squared distances between the observed pollutant fractions (after standardization) and their predicted cluster centers.¹⁵ This provides an aggregate estimate of the error in using predicted cluster membership relative to the observed value for each measured TRAP component. eAppendix 2; http://links.lww.com/EE/A17 provides additional detail for calculating this CV metric.

We predicted cluster membership at each participant residence using a multinomial logistic regression model (see eAppendix 1; http://links.lww.com/EE/A17). The covariates for prediction were PCA scores evaluated using geographic covariates at participant residence locations and based upon the relationship between variables and scores at monitor locations. We assigned cluster membership for each participant record according to residence at the time of the exam.

**CAC progression analyses**

Following the approach of the primary MESA Air analyses,⁶ we estimated the association between CAC progression and pollution exposure (PM₂.₅ or NOX) via a mixed model. The model includes a cross-sectional component that models baseline CAC, a longitudinal component that accommodates time-varying confounders, and a time-varying component that includes transient factors affecting measurements.⁶,³¹ Variables included are age, sex, race/ethnicity, site, scanner type, adiposity, physical activity level, smoking and second-hand smoke exposure status, employment outside the home, total cholesterol level, high-density lipoprotein level, triglyceride level, statin use, an index of neighborhood socioeconomic status,³⁴ education, and income. We excluded all data from participants after a coronary revascularization procedure.

We included cluster membership as an effect modifier for the longitudinal associations between CAC and PM₂.₅ (or NOX). In total, we fit four separate models, corresponding to the two pollutants of interest and the two groupings of the cohort based on the cold- and warm-season clusters. The coefficients from the cluster–pollutant–time interaction provide cluster-specific estimates of the association between pollution exposure (PM₂.₅ or NOX) and CAC progression. Statistical significance of the effect modification was assessed using a likelihood ratio test.

**Sensitivity analyses**

Because we found strong correlation between membership in the identified clusters and baseline exposure levels for
participants, we conducted sensitivity analyses that estimated cluster-specific CAC progression associations for alternative cluster definitions. To compare against clusters not derived from the TRAP measurements, we split the cohort into those with baseline pollution exposure (NOX or PM$_{2.5}$) above and below the cohort-wide median level. As a second sensitivity analysis, we orthogonalized the GIS covariates against baseline NOX level (using year 2000 annual averages for monitor locations) and then computed new PCA scores from these modified covariates and clustered the pollutant measurements via predictive $k$-means. This sensitivity analysis was designed to identify clusters that were less correlated with the regional trend in baseline exposure levels.

**Results**

**TRAP data**

The locations of the 43 monitoring sites are provided in Figure 1. Three (different) locations in each of the cold and warm seasons were removed due to instrument error in the processing of badge measurements, leaving 40 sites for each season. Figure 2 shows a heatmap of the correlations between the pollutants in the cold season, after scaling by NOX. eFigure 1; http://links.lww.com/EE/A17 provides the analogous plot for the warm-season data.

**Clustering results**

In the cold season, the predictive $k$-means model with three clusters and two PCA scores performed the best in CV (MSPE=16.82; eTable 1; http://links.lww.com/EE/A17). However, parameters in this model were not fully identifiable because one cluster included only two locations but had three coefficients for classification. Therefore, we chose the model with three clusters and one PCA score for the cold season because it was not overdetermined and had the second-best CV performance (MSPE = 16.94). In the warm season, the best CV results were for the model with three clusters and two PCA scores (MSPE = 15.32).

The cluster centers from the cold season are depicted in Figure 3. Most monitors were assigned to Cluster 1 (21 locations) or Cluster 2 (17 locations). Cluster 1 was characterized by above average ratios of NO$_2$ and ozone relative to NOX, while the profile for Cluster 2 showed the opposite trend, with lower fractions of gases but higher ratios of ultrafine (25–400 nm) and accumulation mode (0.05–1 μm) particle counts relative to NOX. Locations west and north of downtown were primarily assigned to Cluster 1, while those assigned to Cluster 2 were located closer to downtown (see Figure 1). Cold-season Cluster 3, which comprised two locations, had high ratios of almost all gases and particle sizes. A summary of cluster attributes is provided in Table 3.

We predicted that, at baseline, 743 and 262 participants belonged to cold-season Clusters 1 and 2, respectively, and none to Cluster 3. Table 1 summarizes baseline characteristics of the study cohort, stratified by cold-season and warm-season cluster membership. In addition to the clear geographic patterns between clusters, participants in Cluster 1 tended to have higher levels of education and socioeconomic status than those in Cluster 2, while baseline NOX and PM$_{2.5}$ exposure levels were higher in Cluster 2 (Figure 4).

The centers for the clusters identified in the warm season are depicted in eFigure 2; http://links.lww.com/EE/A17. The first cluster (to which 19 locations were assigned) was characterized by lower ratios of VOCs and higher ratios of NO2, O3, and PN$_{1-3}$. The second cluster (12 locations) had the highest ratios of all VOCs and CO but low ratios of particle counts. The third cluster (9 locations) was notable for its low PN$_{1-3}$ counts and NO2 and O3 concentrations.

Warm-season cluster membership also showed a spatial pattern, with monitors located in the urban center primarily assigned to Cluster 1 and locations furthest from downtown generally assigned to Cluster 3 (eFigure 3; http://links.lww.com/EE/A17).
Figure 2. Heatmap of the correlation between measurements in the cold season. NO₂ indicates nitrogen dioxide; O₃, ozone; PN, particle number.

Figure 3. Cold-season cluster centers. NO₂ indicates nitrogen dioxide; NOₓ, oxides of nitrogen; O₃, ozone; PN, particle number.
Most participants (701 at baseline) were predicted to belong to Cluster 1 in the warm season. A total of 151 and 153 participants were predicted to belong to Clusters 2 and 3, respectively. Participants in Cluster 2 tended to have lower levels of income, education, and neighborhood socioeconomic status (Table 1). The warm-season clusters also showed patterns in baseline NOX and PM2.5 concentrations at residences (Figure 4).

In the sensitivity analysis using clusters based on GIS covariates with correlation with baseline NOX removed, the best models (according to CV MSPE) had three clusters and two PCA scores in the cold season and two clusters and four PCA scores in the warm season (eTable 1; http://links.lww.com/EE/A17). Similar to cold-season Cluster 1 from the primary analysis, cold-season Cluster 1 from the sensitivity analysis had below-average ratios for all TRAP components relative to NOX except for ultrafine and accumulation mode particle counts (eFigure 4; http://links.lww.com/EE/A17). Cold-season Cluster 2 from the sensitivity analysis had low ratios of ultrafine and accumulation mode particle counts, similar to Cluster 1 from the primary analysis. Cluster 3 had high ratios of almost all TRAP components relative to NOX but only included four monitors. In the warm season, Cluster 1 broadly resembled a combination of Cluster 1 and 3 from the primary analysis and had high particle count fractions and low VOC fractions, while Cluster 2 showed the opposite trend (eFigure 6; http://links.lww.com/EE/A17). In both seasons, the baseline NOX exposures for participants did not show a strong correlation with cluster membership (Figure 4).

**CAC progression estimates**

In a model without effect modification by cluster, a difference of 5 μg/m³ in PM2.5 was associated with 23.0 Agatston units per year CAC progression (95% confidence interval [CI] = 14.2, 31.7). When cold-season cluster membership was used as an effect modifier, the estimated association with a 5 μg/m³ difference in PM2.5 was 17.0 (95% CI = 7.2, 26.7) units/year for participants belonging to Cluster 1 and 42.6 (95% CI = 25.7, 59.4) units/year for participants in Cluster 2 (see Table 4). The model with cluster-specific progression terms was significantly improved over the model without effect modification.

**Table 3**

Descriptive summary of clusters

| Season     | Clustering covariates | Cluster | No. of monitors | No. of subjects | Predominant location(s) | Distinctive high-ratio features | Distinctive low-ratio features |
|------------|-----------------------|--------|----------------|----------------|-------------------------|---------------------------------|--------------------------------|
| Cold       | Original              | 1      | 21             | 743            | Northern and western suburban area | NO2, O3, some VOCs | PN, PN |
|            |                       | 2      | 17             | 262            | Downtown area           | PN, NO2, O3, some VOCs | PN, PN |
|            |                       | 3      | 2              | 0              | Singletons              | VOCs, BC, PN1, PN1-3 | PN, NO2, O3, BC, PN1-3 |
|            | Decorrelated with NOX | 1      | 19             | 664            | Downtown area, with some suburban sites | PN1 | VOCs, NO2, O3, BC, PN1-3 |
|            |                       | 2      | 17             | 341            | Western and northern suburbs | Some VOCs | Particle counts |
|            |                       | 3      | 4              | 0              | Singletons              | VOCs and particles | VOCs, NO2, O3, some VOCs |
| Warm       | Original              | 1      | 19             | 701            | Western suburban area  | NO2, O3, PM1-3 | VOCs |
|            |                       | 2      | 12             | 151            | Between downtown and suburbs | VOCs, CO | PN, PN |
|            |                       | 3      | 9              | 153            | Downtown area           | PN1, PN1 | NO2, O3, PM1-3 |
|            | Decorrelated with NOX | 1      | 26             | 722            | Downtown area and western suburbs | NO2, PN1, PN1-3 | VOCs, CO |
|            |                       | 2      | 14             | 283            | Areas just north and west of downtown | VOCs, CO | NO2, PN1, PN1-3 |

Counts are numbers of subjects predicted to belong to the cluster at baseline. Ratios are relative to NOX concentration. BC indicates black carbon; CO, carbon monoxide; NO2, nitrogen dioxide; NOX, oxides of nitrogen; O3, ozone; PM, particulate matter; PN, particle number; VOC, volatile organic compound.

**CAC progression estimates**

In a model without effect modification by cluster, a difference of 5 μg/m³ in PM2.5 was associated with 23.0 Agatston units per year CAC progression (95% confidence interval [CI] = 14.2, 31.7). When cold-season cluster membership was used as an effect modifier, the estimated association with a 5 μg/m³ difference in PM2.5 was 17.0 (95% CI = 7.2, 26.7) units/year for participants belonging to Cluster 1 and 42.6 (95% CI = 25.7, 59.4) units/year for participants in Cluster 2 (see Table 4). The model with cluster-specific progression terms was significantly improved over the model without effect modification.

**Figure 4.** Baseline NOX exposure by cluster membership. NOX indicates oxides of nitrogen.
different from the city-wide model without cluster interactions ($P = 0.006$). Effect modification by warm season cluster was not statistically significant ($P = 0.10$) but showed a similar trend: 17.1 (95% CI = 7.1, 27.0), 24.7 (95% CI = -0.1, 49.6), and 39.8 (95% CI = 20.0, 59.6) Agatston units/year CAC progression associated with 5 μg/m$^3$ difference in PM$_{2.5}$ among participants in Clusters 1, 2, and 3, respectively. The median baseline PM$_{2.5}$ concentration was 15.77 μg/m$^3$. The estimated associations (per 5 μg/m$^3$ difference in PM$_{2.5}$) were 13.7 (95% CI = 2.1, 25.4) and 33.1 (95% CI = 21.2, 44.9) units/year for participants with baseline exposure below and above this value, respectively.

The estimated associations with NO$_X$ exposure showed a similar trend across clusters (see Table 5). A difference of 40 ppb in NO$_X$ exposure was associated with 22.2 (95% CI = 7.7, 36.7) and 41.9 (95% CI = 23.7, 60.2) Agatston units/year CAC progression among participants in cold-season Clusters 1 and 2, respectively. These results were not significantly different ($P = 0.076$) from the estimate for all participants pooled together (28.2 units/year, 95% CI = 17.1, 39.3). Estimates for the warm-season clusters followed the same pattern as for PM$_{2.5}$: 20.7 (5.9, 35.6), 25.7 (-2.0, 53.4), and 38.1 (16.8, 59.5) for Clusters 1, 2, and 3, respectively. There was no evidence for effect modification by baseline NO$_X$ ($P = 0.64$), although the point estimates had a similar trend to results from the model with effect modification by cold-season cluster.

In the sensitivity analysis using covariates that had been orthogonalized with respect to NO$_X$, there was little difference in the estimated association between pollution exposure and CAC progression between clusters from the warm season (eTable 2; http://links.lww.com/EE/A17). When clustering subjects using the cold-season data, the estimates from Cluster 1 vs. 2 were significantly different ($P = 0.023$ and $P = 0.017$ for PM$_{2.5}$ and NO$_X$, respectively). The confidence intervals for the progression estimates in Cluster 2 were notably broad and included zero for NO$_X$ and PM$_{2.5}$. In cold-season Cluster 1, the estimated association between CAC progression and both PM$_{2.5}$ and NO$_X$ exposure was similar to, but greater than, the estimates from the city-wide analysis without clustering. Similar to the trend observed in the original analysis, the largest associations with CAC progression were observed among people predicted to be exposed to pollution with higher ratios of ultrafine and accumulation-mode particles relative to NO$_X$.

### Discussion

We have used a novel collection of near- and on-road multi-pollutant measurements to assess how spatial differences in pollution composition affect the relationship of PM$_{2.5}$ and NO$_X$ exposure with CAC. By reducing dimension of the measurements using clustering and categorizing the cohort by these clusters, we were able to incorporate multipollutant information into a longitudinal model for CAC progression.

In our primary analysis, we found significant differences in the association between CAC progression and PM$_{2.5}$ exposure when grouping subjects by predicted cold-season TRAP profile. The association was strongest among participants in Cluster 2, which was notable for its locations being primarily in the downtown area and its above average ratios of ultrafine and accumulation-mode particle counts relative to NO$_X$. Freshly emitted vehicle exhaust is one likely source for these higher particle counts in the urban center. When grouping by predicted warm season TRAP profiles, the estimated association between CAC progression and PM$_{2.5}$ exposure was weaker in the cluster with lower ratios of ultrafine particle counts (Cluster 2) and strongest in the cluster with high ratios of ultrafine particle counts (Cluster 3).

Our results suggest that overall PM$_{2.5}$ exposure among individuals whose ambient residential TRAP concentrations have high levels of ultrafine particles relative to NO$_X$ has greater adverse cardiovascular impacts than exposure among participants whose ambient residential TRAP concentrations have different composition. Evidence for increased risk of atherosclerosis due to exposure to ultrafine particles, relative to larger particle sizes, has been found in mice. Ultrafine particles have also been linked to a broader set of inflammation- and endothelial-related outcomes. Although a defining feature of the identified TRAP profiles were the relative fractions of ultrafine particle counts, the moderately high correlation between ultrafine and accumulation-mode particle (Figure 2 and eFigure 1; http://links.lww.com/EE/A17) mean that the differences between clusters cannot be attributed to ultrafine particles alone. The lower fractions of VOCs and gases in cold-season Cluster 2 and warm-season Cluster 1 may also play a role in the observed differences.

A striking feature of the clustering results is the strong correlation between cluster membership, geographic location, and overall PM$_{2.5}$ and NO$_X$ levels. The predictive $k$-means clustering procedure incorporated an aggregation of geographic covariates via the PCA scores, which have a strong gradient from downtown out to the suburbs. These covariates are similar to those used for predicting exposures in the MESA Air cohort, which may drive some of the correlation between cluster membership and exposure level. This correlation makes it difficult to determine whether the between-cluster differences identified are due to differences in TRAP composition or differences in baseline exposure. We addressed this concern by conducting a

---

**Table 4**

Cluster-specific estimates of the association between CAC progression, in Agatston units per year, and differences of 5 μg/m$^3$ PM$_{2.5}$

| Source of clusters | Cluster name | Estimate (95% confidence interval) | Effect modification | P value$^a$ |
|--------------------|--------------|-----------------------------------|---------------------|------------|
| None               | City-wide    | 23.0 (14.2, 31.7)                 | 0.06                |            |
| Cold-season data   | Cluster 1    | 17.0 (7.2, 26.7)                 | 0.006               |            |
|                   | Cluster 2    | 42.6 (25.7, 59.4)                 | 0.10                |            |
| Warm-season data   | Cluster 1    | 17.1 (7.1, 27.0)                 | 0.10                |            |
|                   | Cluster 2    | 24.7 (0.1, 49.6)                 | 0.013               |            |
|                   | Cluster 3    | 39.8 (20.0, 59.6)                 | 0.013               |            |
| Baseline PM$_{2.5}$| ≤Median      | 13.7 (2.1, 25.4)                 | 0.013               |            |
|                   | >Median       | 33.1 (21.1, 44.9)                | 0.013               |            |

$^a$P-values are from a likelihood ratio test comparing against the model without cluster-specific progression estimates.

$^b$No subjects were predicted to belong to Cluster 3 in the Cold Season.

CAC indicates coronary artery calcification; PM, particulate matter.

---

**Table 5**

Cluster-specific estimates of the association between CAC progression, in Agatston units per year, and differences of 40 ppb NO$_X$.

| Source of clusters | Cluster name | Estimate (95% confidence interval) | Effect modification | P value$^a$ |
|--------------------|--------------|-----------------------------------|---------------------|------------|
| None               | City-wide    | 28.2 (17.1, 39.3)                 | 0.076               |            |
| Cold-season data   | Cluster 1    | 22.2 (7.7, 36.7)                 | 0.076               |            |
|                   | Cluster 2    | 41.9 (23.7, 60.2)                 | 0.38                |            |
| Warm-season data   | Cluster 1    | 20.7 (5.9, 35.6)                 | 0.38                |            |
|                   | Cluster 2    | 25.7 (0.0, 53.4)                 | 0.38                |            |
|                   | Cluster 3    | 38.1 (16.8, 59.5)                 | 0.38                |            |
| Baseline PM$_{2.5}$| ≤Median      | 25.1 (16.8, 45.5)                | 0.64                |            |
|                   | >Median       | 30.6 (17.6, 43.6)                | 0.64                |            |

$^a$P-values are from a likelihood ratio test comparing against the model without cluster-specific progression estimates.

CAC indicates coronary artery calcification; NO$_X$, oxides of nitrogen; PM, particulate matter.
sensitivity analysis that removed variability from the covariates that is explainable by baseline NO\textsubscript{X} and yielded clusters that did not correlate well with baseline exposure. Using the cold-season TRAP measurements, the sensitivity analysis found larger associations between CAC and PM\textsubscript{2.5} exposure. Participants predicted to have TRAP profiles with higher fractions of ultrafine particles (Cluster 1). This supports the primary results, which found greater rates of progression among people predicted to have TRAP profiles with higher fractions of ultrafine particles.

Predicted cluster membership correlated with socioeconomic status, in addition to baseline exposure and geographic location. Strong relationships between socioeconomic status and pollution exposure have been previously identified in this cohort. The results from our analysis, however, identify differences that are not fully explainable by socioeconomic differences alone. In our primary analyses, participants in cold-season Cluster 1 and warm-season Cluster 1 had higher levels of affluence and weaker, but still non-zero, estimated associations between CAC and PM\textsubscript{2.5} and NO\textsubscript{X} exposure. However, participants in warm-season Cluster 2 had the lowest income and neighborhood-level socioeconomic index, but their estimated association between PM\textsubscript{2.5} and NO\textsubscript{X} exposure and CAC progression was very similar to the city-wide average. Additionally, participants in warm-season Cluster 3 had higher income and were more affluent but also had a stronger estimated association between PM\textsubscript{2.5} and CAC progression than participants in Cluster 2.

We estimated TRAP profiles for the cold and warm season separately and found similar trends in the relative component fractions across seasons. This could be due to similar wide-spread sources of directly emitted TRAP in each location across seasons, in contrast to residential heating sources or secondary pollutants that vary by season. The trend of effect modification by predicted profile was similar as well, although with different levels of statistical evidence. This is likely due in part to the overlap between the predicted cluster membership.

A limitation of this analysis was the relatively small number of locations at which component measurements were made. Although a sample size of 40 locations is far more than the one or two locations per city at which component data is typically available via the Chemical Speciation Network, it is nonetheless a relatively small sample for building a cluster prediction model. Including spatial splines in the prediction model as a mechanism for spatial smoothing is impractical with cluster sizes of 21 and 17 from the cold-season data. This leads to the prediction model being derived from one or two PCA scores, which can capture small-scale variability but can also be dominated by larger trends, as was the case in this study. The results may be impacted by differences in the time period of the data. The geographic variables used for cluster prediction were from the period 2000–2006. The cluster profiles were derived from multipollutant measurements in 2012 but were used to predict representative exposure profiles for the entire study period. It is possible the TRAP profiles and the relationship between geographic covariates and TRAP within the Baltimore region changed over time, although major highway patterns and industrial sources were largely stable.

The city-wide and cluster-specific estimates were all greater than the estimated association in the full MESA Air cohort, which includes participants from five other metropolitan areas. In the full cohort, differences of 5 μg/m\textsuperscript{3} difference in PM\textsubscript{2.5} and 40 ppb in NO\textsubscript{X} were associated with 4.1 (95% CI = 1.4, 6.8) and 4.8 (95% CI = 0.9, 8.7) Agatston units/year CAC progression, respectively.\textsuperscript{6} While the exposure prediction models for Baltimore did have the best overall predictive accuracy,\textsuperscript{22} the difference in progression estimates is not likely due to exposure assessment accuracy alone. Differences in pollution composition between the six MESA Air cities could potentially cause some of this difference; however, such differences are masked by between-city differences in cohort members. The defining feature of the MESA cohort is the oversampling of different ethnicities. The Baltimore sub-cohort, however, only includes white and black participants, while the other cities also have different racial-ethnic groups, which could be one source of the differences from the Baltimore-only results.

We have presented a novel approach to using multipollutant TRAP measurements within a metropolitan area to assess effect modification by pollution composition in longitudinal relationships between pollution exposure and CAC. Our results found that the same difference in PM\textsubscript{2.5} or NO\textsubscript{X} concentration was associated with faster CAC progression among participants living in areas predicted to have higher ratios of ultrafine particle counts relative to NO\textsubscript{X} during the cold season. These results highlight how incorporating multipollutant measurements into health effect analyses can yield insight into heterogeneity in the relationships between air pollution exposure and health.

Conflict of interest statement

The authors declare that they have no conflicts of interest with regard to the content of this report.

References

1. Hoek G, Krishnan RM, Beelen R, et al. Long-term air pollution exposure and cardio-respiratory mortality: a review. Environ Heal 2013;12:43.
2. Brook RD, Rajagopalan S, Pope CA, et al. Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association. Circulation 2010;121:2331–78.
3. Beelen R, Raaschou-Nielsen O, Staafogga M, et al. Effects of long-term exposure to air pollution on natural-cause mortality: an analysis of 22 European cohorts within the multicentre ESCAPE project. Lancet 2014;383(9919):785–95.
4. Dockery DW, Pope CA, Xu X, et al. An association between air pollution and mortality in six U.S. cities. N Engl J Med 1993;329:1753–9.
5. Miller KA, Siscovick DS, Sheppard L, et al. Long-term exposure to air pollution and incidence of cardiovascular events in women. N Engl J Med 2007;356:447–58.
6. Kaufman JD, Adar SD, Barr RG, et al. Association between air pollution and coronary artery calcification within six metropolitan areas in the USA (the Multi-Ethnic Study of Atherosclerosis and Air Pollution): a longitudinal cohort study. Lancet 2016;388:696–704.
7. Karner AA, Eisinger DS, Niemeier DA. Near-roadway air quality: synthesizing the findings from real-world data. Environ Sci Technol 2010;44:5334–44.
8. Fischer PH, Hoek G, Van Reeuwijk H, et al. Traffic-related differences in outdoor and indoor concentrations of particles and volatile organic compounds in Amsterdam. Atmos Environ 2000;34:3713–22.
9. Vu TV, Delgado-saborit JM, Harrison RM. Review : Particle number size distributions from seven major sources and implications for source apportionment studies. Atmos Environ 2015;122:114–32.
10. Oakes M, Baxter L, Long TC. Evaluating the application of multipollutant exposure metrics in air pollution health studies. Environ Int 2014;69:90–9.
11. Greenbaum D, Shaikh R. First steps toward multipollutant science for air quality decisions. Epidemiology 2010;21:195–7.
12. Vedal S, Kaufman JD. What does multi-pollutant air pollution research mean? Am J Respir Crit Care Med 2011;183:3–4.
13. Dominici F, Peng R, Barr CD, Bell ML. Protecting human health from air pollution: shifting from a single-pollutant to a multi-pollutant approach. Epidemiology 2010;21:187–94.
14. Health Effects Institute. HEI Strategic Plan for Understanding the Health Effects of Air Pollution 2015–2020. Boston, MA: 2015.
15. Keller JR, Drton M, Larson T, Kaufman JD, Sandler DP, Sipriu AA. Covariate-adaptive clustering of exposures for air pollution epidemiology cohorts. Ann Appl Stat 2017;11:93–113.
16. Kiooumourtzoglou M, A-A, Austin E, Koutrakis P, Dominici F, Schwartz J, Zanobetti A. PM\textsubscript{2.5} and survival among older adults: Effect modification by race/ethnicity composition. Epidemiology 2015;26:321–7.
17. Hartigan J, Wong M. Algorithm AS 136: A k-means clustering algo- rithm. Appl Stat 1979;28:100–8.
18. Austin E, Coull BA, Zanobetti A, Koutrakis P. A framework to spatially cluster air pollution monitoring sites in US based on the PM2.5 composition. *Environ Int* 2013;59:244–54.

19. Hoek G, Beelen R, Kos G, et al. Land use regression model for ultrafine particles in Amsterdam. *Environ Sci Technol* 2011;45:622–8.

20. Abernethy RC, Allen RW, McKendry IG, Brauer M. A land use regression model for ultrafine particles in Vancouver, Canada. *Environ Sci Technol* 2013;47:5217–25.

21. Rivera M, Basagaña X, Aguilera I, et al. Spatial distribution of ultrafine particles in urban settings: a land use regression model. *Atmos Environ* 2012;54:657–66.

22. Weichenthal S, Van Ryswyk K, Goldstein A, Shekarrizfard M, Hatzopoulou M. Characterizing the spatial distribution of ambient ultrafine particles in Toronto, Canada: A land use regression model. *Environ Pollut* 2016;208:241–8.

23. Mukerjee S, Smith LA, Johnson MM, Neas LM, Stallings CA. Spatial analysis and land use regression of VOCs and NOx from school-based urban air monitoring in Detroit/Dearborn, USA. *Sci Total Environ* 2009;407:4642–51.

24. Kaufman JD, Adar SD, Allen RW, et al. Prospective study of particulate air pollution exposures, subclinical atherosclerosis, and clinical cardiovascular disease: The Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air). *Am J Epidemiol* 2012;176:825–37.

25. Xu W, Riley EA, Austin E, et al. Use of mobile and passive badge air monitoring data for NOx and ozone air pollution spatial exposure prediction models. *J Expo Sci Environ Epidemiol* 2017;27:184–92.

26. Riley EA, Schaal LN, Johnson MM, Neas LM, Stallings CA. Spatial analysis and land use regression of VOCs and NOx from school-based urban air monitoring in Detroit/Dearborn, USA. *Sci Total Environ* 2009;407:4642–51.

27. Bild DE, Bluemke DA, Burke GL, et al. Multi-Ethnic Study of Atherosclerosis: objectives and design. *Am J Epidemiol* 2002;156:871–81.

28. Carr JJ, Nelson JC, Wong ND, et al. Calcified coronary artery plaque measurement with cardiac CT in population-based studies: standardized protocol of Multi-Ethnic Study of Atherosclerosis (MESA) and Coronary Artery Risk Development in Young Adults (CARDIA) Study 1. *Radiology* 2005;234:35–43.

29. Keller JP, Olives C, Kim S-Y, et al. A unified spatiotemporal modeling approach for predicting concentrations of multiple air pollutants in the Multi-Ethnic Study of Atherosclerosis and Air Pollution. *Environ Health Perspect* 2015;123:301–9.

30. Mercer LD, Szpiro AA, Sheppard L, et al. Comparing universal kriging and land-use regression for predicting concentrations of gaseous oxides of nitrogen (NOx) for the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air). *Atmos Environ* 2011;45:4412–20.

31. Wang M, Keller JP, Adar SD, et al. Development of long-term spatiotemporal models for ambient ozone in six metropolitan regions of the United States: The MESA Air study. *Atmos Environ* 2015;123:79–87.

32. Bergen S, Sheppard L, Sampson PD, et al. A national prediction model for PM2.5 component exposures and measurement error-corrected health effect inference. *Environ Health Perspect* 2013;121:1017–25.

33. Gassett AJ, Sheppard L, McClelland RL, et al. Risk factors for long-term coronary artery calcium progression in the Multi-Ethnic Study of Atherosclerosis. *J Am Heart Assoc* 2015;4:e001726.

34. Moore K, Diez Roux A V, Auchincloss A, et al. Home and work neighbourhood environments in relation to body mass index; the Multi-Ethnic Study of Atherosclerosis (MESA). *J Epidemiol Community Health* 2013;67:846–53.

35. Araujo JA, Barajas B, Kleinman M, et al. Ambient particulate pollutants in the ultrafine range promote early atherosclerosis and systemic oxidative stress. *Circ Res* 2008;102:589–96.

36. Nel A, Xia T, Lutz M, Li N. Toxic potential of materials at the nanolevel. *Science (80-)* 2007;311(5726):622–7.

37. Oberdörster G, Oberdörster E, Oberdörster J. Nanotoxicology: an emerging discipline evolving from studies of ultrafine particles. *Environ Health Perspect* 2005;113:823–39.

38. Hajat A, Diez-roux A V, Adar SD, Auchincloss AH, Lovasi GS, Neill MSO. Air pollution and individual and neighborhood socioeconomic status: evidence from the Multi-Ethnic Study of Atherosclerosis (MESA). *Environ Health Perspect* 2013;121:1325–33.