$PM_{2.5}$ and survival among older adults: Effect modification by particulate composition

Marianthi-Anna Kioumourtzoglou, Elena Austin, Petros Koutrakis, Francesca Dominici, Joel Schwartz, Antonella Zanobetti

Supplemental Material

Contents

List of Tables

List of Figures

Cluster Analyses

e-1 Spatial Clustering Based on $PM_{2.5}$ Composition

We categorized cities in clusters with similar air pollution profiles. To do so, we employed a k-means clustering analysis.[2] K-means is an iterative algorithm that minimizes the Euclidean distance between each observation and the pre-defined k cluster centers, seeking to partition M points (in our analysis M is the number of species) in N dimensions (the number of cities) into k clusters.

As observations, we used modified Z-scores for each chemical species, which were calculated using the ratio of the species average between January 2003 and December 2008 at each site, divided by the corresponding $PM_{2.5}$ concentration at that site. We then employed the k-means clustering algorithm to cluster sites that have the most similar $PM_{2.5}$ composition together, using the Z-scores. The $PM_{2.5}$ chemical species included in the clustering were the ions NO_3^-, Na^+, K^+, SO_4^{2-}, NH_4^+, elemental carbon (EC), organic carbon (OC), and the elements S, Cu, Fe, Zn, Ni, V, Ti, Mg, K, Si, Na, Cl, Ca, Br, Sr, Pb, Mn.

In our analysis, we specified the initial k values (cluster center points) to increase the stability of our solution, by identifying the k centers with hierarchical clustering.[3] The preliminary identification of the k centers also allowed for minimum impact of potential outlying observations. Given that this algorithm depends on the choice of the number of clusters k, subject specific knowledge is required to select the best number of clusters.[4] To this end, we used knowledge of air pollution sources and contributions, minimizing the variability of predefined concentration ratios of specific species,[1] as well as the number of single-city clusters. A more detailed description of the methods used for the spatial clustering is presented by Austin et al.[1]

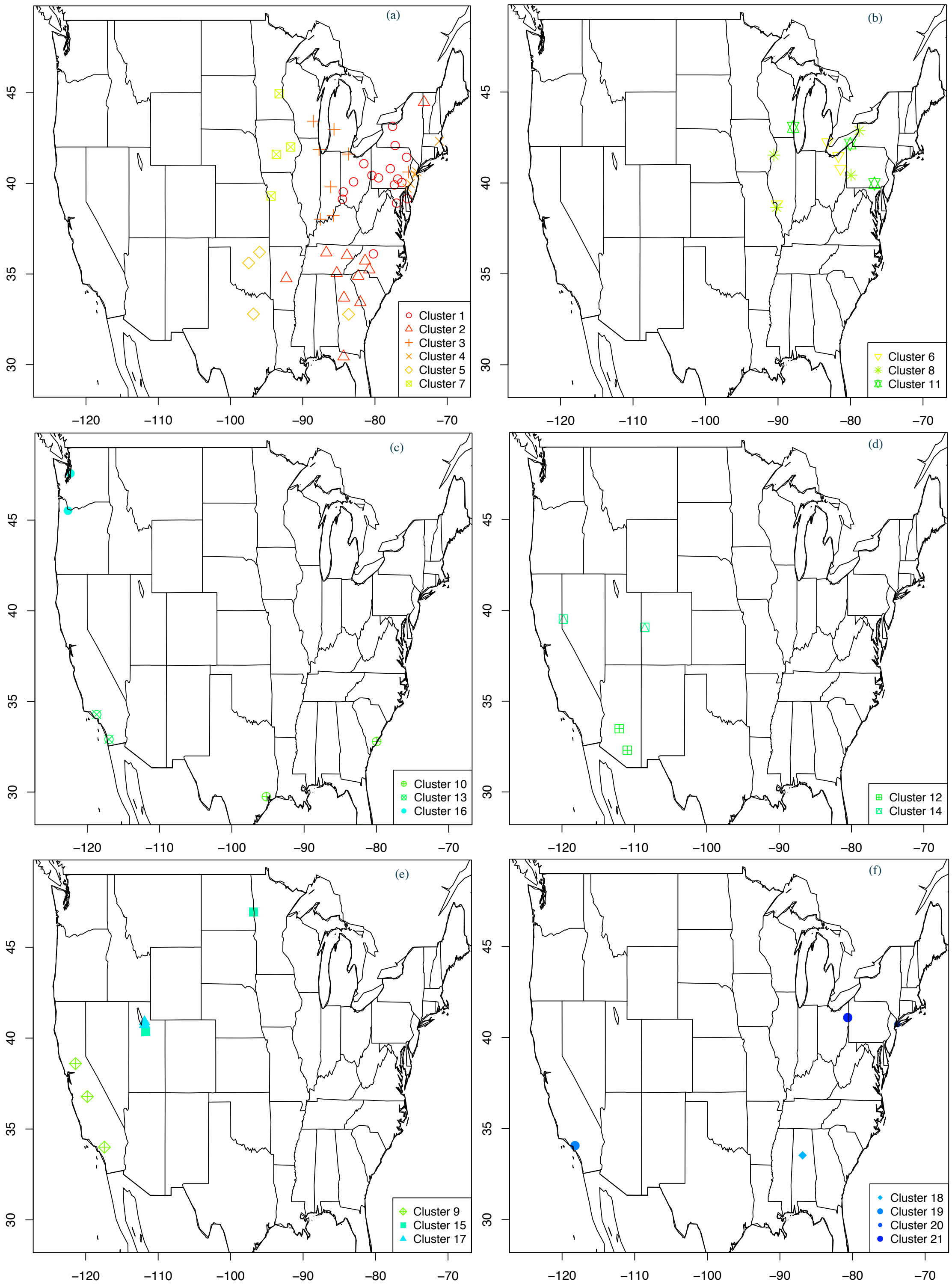
eTable 1: Cities per cluster and number of subjects and deaths.

City	Cluster	No. Subjects	No. Deaths	Mean Age
Akron, OH	1	129,158	46,899	75.56
Bath, NY	1	25,920	9,105	75.33
Cincinnati, OH	1	192,292	71,065	75.58
Columbus, OH	1	195,999	68,045	75.09
Dover, DE	1	31,421	9,375	74.18
Gettysburg, PA	1	24,151	7,449	75.04
Greensburg, PA	1	114,001	41,440	75.58
Harrisburg, PA	1	61,904	21,656	75.38
Lancaster, PA	1	118,627	40,037	75.77
Middletown, OH	1	68,712	23,393	74.73
Rochester, NY	1	166,479	58,050	75.89
Scranton, PA	1	160,546	65,063	76.06
State College, PA	1	26,552	8,540	75.10
Washington, DC	1	149,574	47,440	76.22
Washington, PA	1	62,845	23,249	75.66
Winston-Salem, NC	1	77,165	25,390	74.92
Atlanta, GA	2	439,291	130,324	74.68
Augusta, GA	2	37,936	13,153	74.66
Burlington, VT	2	29,091	9,109	75.17
Charlotte, NC	2	129,147	38,499	74.90
Chattanooga, TN	2	79,070	27,311	75.08
Greenville, SC	2	93,791	29,611	74.75
Hickory, NC	2	36,221	11,848	74.55
Knoxville, TN	2	120,649	40,972	74.92
Little Rock, AR	2	76,921	26,092	75.27
Nashville, TN	2	116,528	40,272	75.23
Tallahassee, FL	2	41,424	12,891	75.09
Allentown, PA	3	157,331	54,641	75.74
Beaver Dam, WI	3	22,737	7,774	75.76
Chicago, IL	3	1,285,802	441,220	75.50
Evansville, IN	3	44,884	16,855	75.71
Grand Rapids, MI	3	110,459	37,055	75.52
Indianapolis, IN	3	169,547	60,349	75.28
Louisville, KY	3	165,275	59,783	75.21
Toledo, OH	3	97,880	37,150	75.55
Boston, MA	4	590,615	199,689	75.88
Elizabeth, NJ	4	124,695	40,284	75.88
Middlesex, NJ	4	164,094	50,662	75.44
Philadelphia, PA	4	1,051,643	385,935	75.69
Wilmington, DE	4	105,192	34,875	75.16

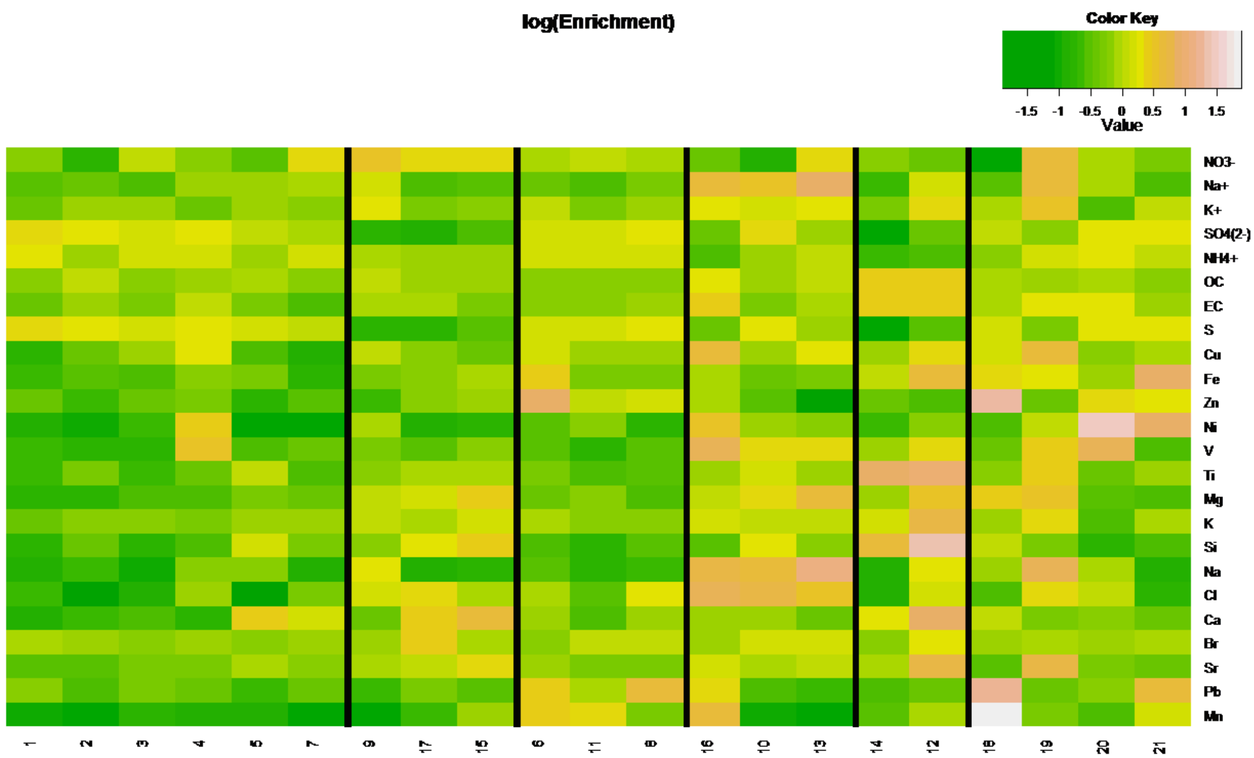
Continued on next page

eTable 1 – *Continued from previous page*

City	Cluster	No. Subjects	No. Deaths	Mean Age
Dallas, TX	5	345,887	107,824	74.80
Macon, GA	5	38,533	13,994	75.19
Oklahoma city, OK	5	150,558	52,091	75.07
Tulsa, OK	5	131,556	45,488	75.07
Canton, OH	6	106,223	38,076	75.43
Cleveland, OH	6	460,334	171,576	75.74
Detroit, MI	6	833,786	303,496	75.47
Madison, IL	6	65,611	23,933	75.35
Cedar Rapids, IA	7	43,765	14,513	75.45
Des Moines, IA	7	78,372	26,044	75.38
Kansas City, KS	7	303,035	104,278	75.34
Minneapolis, MN	7	316,088	103,998	75.92
Buffalo, NY	8	239,595	90,500	75.74
Davenport, IA	8	73,588	25,380	75.45
Pittsburgh, PA	8	351,259	135,693	76.00
St. Louis, MO	8	351,018	130,270	75.63
Fresno, CA	9	146,228	46,892	75.34
Riverside, CA	9	695,165	206,934	74.79
Sacramento, CA	9	253,698	78,074	75.23
Charleston, SC	10	70,448	21,607	74.95
Houston, TX	10	510,950	154,406	74.44
Tampa, FL	10	240,184	73,363	75.00
Erie, PA	11	67,508	25,019	75.69
Milwaukee, WI	11	276,230	101,527	75.77
York, PA	11	95,883	31,220	75.12
Phoenix, AZ	12	682,866	200,388	75.01
Tucson, AZ	12	231,725	66604	74.99
San Diego, CA	13	571,128	17,4001	75.44
Ventura, CA	13	157,695	44,466	75.23
Grand Junction, CO	14	360,15	10,907	75.21
Reno, NV	14	84,379	23,909	74.35
Fargo, ND	15	23,093	7,253	75.71
Provo, UT	15	49,561	14,338	74.83
Portland, OR	16	292,293	97,052	75.68
Seattle, WA	16	340,433	103,316	75.70
Layton, UT	17	37,272	10,189	74.43
Salt Lake City, UT	17	138,049	42,689	75.06
Birmingham, AL	18	211,866	79,621	75.14
Los Angeles, CA	19	1,629,036	485,732	75.34
New York, NY	20	1,517,097	474,471	75.80
Youngstown, OH	21	130,955	49,676	75.65



eFigure 1: Cities within the 21 clusters, by groupings.



eFigure 2: Heatmap of the log of the species enrichment factors by cluster. The enrichment factors are defined as $\frac{\overline{S_{ij}}}{PM_{2.5i}} \div \frac{\overline{S_j}}{PM_{2.5}}$, for j species at site i,[1] and they represent the enrichment of a specific species of $PM_{2.5}$ within a cluster, as compared to the entire sample, i.e. whether a cluster has higher relative contributions of that specific species vs. the other clusters.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
PM2.5	12.92	13.33	13.41	13.09	12.43	14.54	10.30	13.25	16.68	10.94
NO3	0.13	0.07	0.17	0.13	0.09	0.16	0.22	0.16	0.28	0.06
NA+	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02
K+	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.01
SO4	0.33	0.29	0.27	0.29	0.25	0.27	0.23	0.28	0.10	0.31
NH4	0.14	0.10	0.14	0.13	0.10	0.13	0.14	0.14	0.11	0.10
OC	0.27	0.35	0.27	0.30	0.33	0.26	0.28	0.28	0.38	0.31
EC	0.04	0.05	0.04	0.07	0.04	0.05	0.03	0.05	0.06	0.04
S	108.09	95.75	89.99	94.14	84.64	87.77	79.86	95.99	33.79	101.11
Cu	0.20	0.29	0.38	0.57	0.23	0.50	0.19	0.40	0.46	0.39
Fe	4.88	5.76	5.34	7.65	6.91	15.71	4.53	7.29	7.05	6.58
Zn	0.99	0.74	0.96	1.09	0.64	3.90	0.87	1.68	0.70	0.85
Ni	0.07	0.06	0.08	0.27	0.05	0.10	0.06	0.08	0.17	0.15
V	0.09	0.09	0.08	0.34	0.11	0.11	0.13	0.11	0.14	0.28
Ti	0.19	0.29	0.19	0.28	0.45	0.30	0.23	0.25	0.34	0.47
Mg	0.41	0.41	0.52	0.51	0.66	0.62	0.64	0.52	0.97	1.31
K	4.45	5.58	5.66	4.88	5.89	6.54	5.91	5.60	7.52	7.15
Si	4.62	7.15	4.72	5.58	13.27	6.05	7.87	6.15	8.78	13.83
Na	2.67	3.22	2.40	5.15	5.10	3.88	2.62	2.99	8.17	12.58
Cl	1.55	0.95	1.33	3.02	0.96	3.13	2.46	4.26	3.94	7.01
Ca	2.30	2.98	3.27	2.72	9.41	4.97	7.14	5.38	4.06	5.10
Br	0.28	0.24	0.22	0.25	0.22	0.24	0.25	0.31	0.24	0.32
Sr	0.06	0.06	0.07	0.07	0.09	0.08	0.07	0.07	0.09	0.09
Pb	0.29	0.20	0.28	0.25	0.19	0.58	0.24	0.68	0.18	0.19
Mn	0.14	0.12	0.16	0.16	0.15	0.63	0.13	0.29	0.13	0.14

	Cluster 11	Cluster 12	Cluster 13	Cluster 14	Cluster 15	Cluster 16	Cluster 17	Cluster 18	Cluster 19	Cluster 20	Cluster 21
PM2.5	13.44	7.80	12.26	9.02	9.43	8.36	11.07	16.78	17.88	12.82	14.14
NO3	0.16	0.10	0.24	0.13	0.23	0.11	0.23	0.05	0.30	0.16	0.12
NA+	0.01	0.01	0.03	0.01	0.01	0.02	0.01	0.01	0.02	0.01	0.01
K+	0.00	0.01	0.01	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.01
SO4	0.27	0.15	0.20	0.08	0.13	0.15	0.09	0.25	0.17	0.29	0.29
NH4	0.13	0.06	0.12	0.05	0.10	0.06	0.10	0.09	0.13	0.14	0.12
OC	0.28	0.56	0.35	0.53	0.31	0.44	0.30	0.32	0.29	0.31	0.28
EC	0.05	0.09	0.05	0.10	0.04	0.09	0.06	0.06	0.08	0.08	0.05
S	90.11	46.88	64.23	25.55	43.29	49.98	31.83	85.21	56.94	97.78	95.70
Cu	0.38	0.61	0.57	0.39	0.28	0.82	0.36	0.52	0.86	0.36	0.44
Fe	7.51	18.83	7.27	11.10	9.38	9.43	8.03	14.37	13.17	8.97	25.64
Zn	1.57	0.84	0.46	0.99	1.26	1.39	1.14	5.85	0.99	2.11	1.94
Ni	0.14	0.13	0.13	0.08	0.08	0.29	0.07	0.09	0.18	0.82	0.45
V	0.08	0.26	0.27	0.16	0.16	0.45	0.11	0.13	0.30	0.45	0.10
Ti	0.21	1.13	0.36	1.06	0.40	0.38	0.41	0.32	0.63	0.26	0.36
Mg	0.77	1.72	1.83	0.79	1.51	0.96	1.09	1.50	1.58	0.57	0.51
K	5.25	14.56	7.23	7.87	8.38	8.02	6.71	6.03	9.52	3.77	6.67
Si	4.96	43.93	8.46	21.33	17.59	6.31	14.17	11.26	7.84	4.70	5.85
Na	2.78	8.12	20.85	2.53	2.93	14.12	2.56	5.40	14.34	6.31	2.66
Cl	1.95	3.88	5.77	1.37	3.13	7.81	4.78	1.81	4.56	3.65	1.46
Ca	3.08	15.74	3.89	8.06	11.55	5.01	9.26	6.11	4.37	4.80	3.75
Br	0.29	0.38	0.33	0.22	0.27	0.25	0.47	0.24	0.29	0.25	0.28
Sr	0.07	0.20	0.10	0.10	0.14	0.11	0.10	0.06	0.20	0.07	0.06
Pb	0.36	0.25	0.19	0.20	0.23	0.53	0.26	1.24	0.24	0.28	0.74
Mn	0.54	0.36	0.13	0.22	0.34	0.71	0.19	2.53	0.27	0.21	0.45

eFigure 3: Ratios of species over $PM_{2.5}$ mean concentrations within clusters (dimensionless). For $PM_{2.5}$, the average cluster concentrations are presented ($\mu g/m^3$).

eTable 2: Within cluster (CL) estimated HR per 10 μg/m^3 of $PM_{2.5}$ and their 95% CIs, by group of clusters.

Cluster	HR	95% CI
Group 1		
CL-1	0.94	(0.76, 1.16)
CL-2	**1.46**	**(1.14, 1.86)**
CL-3	1.00	(0.76, 1.33)
CL-4	1.17	(0.82, 1.66)
CL-5	1.25	(0.84, 1.85)
CL-7	1.24	(0.82, 1.88)
Group 2		
CL-9	0.87	(0.56, 1.36)
CL-15	0.65	(0.34, 1.25)
CL-17	0.84	(0.48, 1.48)
Group 3		
CL-6	1.12	(0.75, 1.65)
CL-8	1.35	(0.91, 2.00)
CL-11	1.49	(0.94, 2.35)
Group 4		
CL-10	**1.71**	**(1.08, 2.72)**
CL-13	**0.44**	**(0.25, 0.77)**
CL-16	**1.91**	**(1.10, 3.32)**
Group 5		
CL-12	0.93	(0.52, 1.65)
CL-14	0.77	(0.43, 1.38)
Group 6		
CL-18	1.66	(0.76, 3.60)
CL-19	1.42	(0.66, 3.08)
CL-20	1.17	(0.54, 2.53)
CL-21	0.97	(0.44, 2.15)

e-2 Related Code

SAS code: City-specific Cox Models

```
proc phreg data=city_i;
model count*dead(0)= pm25 year coronarycare intenscare diabetes
copd mi chf / ties=exact;
strata followup age sex race;
run;
```

R code: Random Effects Meta-analysis

```
library(meta)
library(rmeta)

model.pm <- mvmeta(I(dta$est*10)~1, S=(dta$se*10)^2, method="reml",
na.action="na.omit")
model.clust <- mvmeta(I(dta$est*10) ~ as.factor(dta$cluster),
S=(dta$se*10)^2, method="reml", na.action="na.omit")
```

References

1. E. Austin, B.A. Coull, A. Zanobetti, P. Koutrakis. A framework to spatially cluster air pollution monitoring sites in US based on the $PM_{2.5}$ composition. *Environment International*, 59:244–254, 2013.
2. JA Hartigan, MA Wong. Algorithm as 136: A k-means clustering algorithm. *Applied statistics*, pages 100–108, 1979.
3. R Maitra, AD Peterson, AP Ghosh. A systematic evaluation of different methods for initializing the k-means clustering algorithm. *IEEE Transactions on Knowledge and Data Engineering*, pages 522–537, 2010.
4. AK Jain, MN Murty, PJ Flynn. Data clustering: a review. *ACM computing surveys (CSUR)*, 31 (3):264–323, 1999.