Clustering cities with similar fine particulate matter exposure characteristics based on residential infiltration and in-vehicle commuting factors

Lisa K. Baxter, Sc.D.[1*] and Jason D. Sacks, MPH[2]

[1] National Exposure Research Laboratory, U.S. Environmental Protection Agency, RTP, NC
[2] National Center for Environmental Assessment, U.S. Environmental Protection Agency, RTP, NC

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

Abstract

Epidemiological studies have observed between city heterogeneity in $PM_{2.5}$-mortality risk estimates. These differences could potentially be due to the use of central-site monitors as a surrogate for exposure which do not account for an individual's activities or ambient pollutant infiltration to the indoor environment. Therefore, relying solely on central-site monitoring data introduces exposure error in the epidemiological analysis. The amount of exposure error produced by using the central-site monitoring data may differ by city. The objective of this analysis was to cluster cities with similar exposure distributions based on residential infiltration and in-vehicle commuting characteristics.

Factors related to residential infiltration and commuting were developed from the American Housing Survey (AHS) from 2001 – 2005 for 94 Core-Based Statistical Areas (CBSAs). We conducted two separate cluster analyses using a k-means clustering algorithm to cluster CBSAs based on these factors. The first only included residential infiltration factors (i.e. percent of homes with central air conditioning (AC) mean year home was built, and mean home size) while the second incorporated both infiltration and commuting (i.e. mean in-vehicle commuting time and mean in-vehicle commuting distance) factors.

Clustering on residential infiltration factors resulted in 5 clusters, with two having distinct exposure distributions. Cluster 1 consisted of cities with older, smaller homes with less central AC while homes in Cluster 2 cities were newer, larger, and more likely to have central AC. Including commuting factors resulted in 10 clusters. Clusters with shorter in-vehicle commuting times had shorter in-vehicle commuting distances. Cities with newer homes also tended to have longer commuting times and distances.

This is the first study to employ cluster analysis to group cities based on exposure factors. Identifying cities with similar exposure distributions may help explain city-to-city heterogeneity in $PM_{2.5}$ mortality risk estimates.

Keywords: air exchange rates, cluster analysis, exposure error, in-vehicle exposure, infiltration

1 Introduction

Multi-city population-based epidemiological studies have observed heterogeneity between community- or city-specific $PM_{2.5}$-mortality effect estimates (Dominici et al., 2006; Franklin et al., 2007). One potential reason for these differences is the use of central-site monitors as a surrogate for exposure. This may introduce bias into the observed risk estimates if the central-site monitor-exposure relationship varies by city.

Previous studies have hypothesized and reported higher air pollution risks for cities with higher overall air exchange rates (AERs) or pollutant infiltration efficiencies (Bell and Dominici, 2008; Hodas et al., 2012; Janssen et al., 2002; Levy et al., 2005; Medina-Ramon et al., 2006). A number of factors related to home characteristics can influence the infiltration of ambient air into the home. Some of the most important factorsinclude age of construction (Allen et al., 2003; Chan et al., 2005), housing type (i.e., multi- vs. single-family home) (Koenig et al., 2005; Pandian et al., 1993), and central air conditioning (AC) (Johnson and Long, 2005). In addition people may spend time away from their home (e.g. at work) or in other near-source environments (e.g. in vehicles), where the composition and toxicity of pollutants can vary. In-vehicle air pollution measurement studies have also indicated that concentrations of pollutants inside cars and buses are considerably higher than those recorded at nearby central-site monitors (Riediker et al., 2003) and exposure models suggest that even a small amount of time spent in vehicles may contribute significantly to the average daily personal PM exposure (Burke et al., 2001). Recently exposure to traffic pollution while in-vehicle has been shown to result in changes in heart rate variability (Shields et al., 2013). Estimating exposures based on community-average pollution concentrations also does not account for time spent at other locations outside the assigned community, and thus can add bias (Setton et al., 2011).

. This analysis continues our attempt to better understand the heterogeneity in $PM_{2.5}$-mortality effect estimates across cities. Our objective is to group cities with similar central-site monitor – exposure relationships by clustering them using a k-means cluster analysis based on residential infiltration and commuting characteristics. Exposure variables related to infiltration and commuting patterns were developed from the American Housing Survey (AHS) from 2001 – 2005 for 94 Core-Based Statistical Areas (CBSAs). It is anticipated that this approach will identify groups of cities with similar exposure characteristics that may explain the heterogeneity in $PM_{2.5}$ mortality risk estimates observed in multi-city epidemiologic studies.

2 Methods

2.1 Development of variables

We acquired data from the AHS, available from the Department of Housing and Urban Development's website (Department of Housing and Urban Development) on community-specific residential infiltration and commuting patterns. The AHS collects data on the Nation's housing, including number of apartments, single-family homes, mobile homes, and vacant housing units; and household characteristics including household income, housing and neighborhood quality, housing costs, heating equipment and fuels, size of housing unit, and recent moves. AHS also collects information on type of transportation (e.g., car, bus, subway) used to commute to work, commuting distance, and commuting time. National data are collected in odd numbered years, and supplemented with data for 47 selected CBSAs about every six years. The national sample covers an average 55,000 housing units while each metropolitan area sample covers 4,100 or more housing units. For this analysis we used the national surveys and any available metropolitan surveys from 2001-2005.

Using the housing units sampled in each CBSA as part of the AHS, indicators of AERs were calculated as a means to identify those cities that may have a higher fraction of ambient $PM_{2.5}$ penetrate indoors. These indicators include percent of home with central air conditioning, average home age, and average square footage of the home for each CBSA. Previous studies have shown that personal and/or indoor concentrations of sulfate (often used as a tracer for PM of ambient origin) are lower and less well correlated with outdoor concentrations for homes with AC than homes without AC (Suh et al., 1994; Suh et al., 1992). This is likely because air conditioned homes typically have lower air exchange rates (AERs) than homes that use open windows for ventilation, suggesting that the fraction of $PM_{2.5}$ from ambient origin that penetrates indoors (i.e. infiltration) is lower in homes with AC than in homes without AC. Other predictors of AER include the year a structure was built, as well as its size (Chan et al., 2003; Sherman and Matson, 2002). Newer homes are generally more tightly sealed with lower AERs due to modern methods for constructing and sealing building envelopes (Chan et al., 2005; Persily et al., 2010). Similarly, larger houses typically have higher AERs compared to smaller houses, since they contain a greater surface area for leaks to develop (Chan et al., 2003).

The mean in-vehicle commuting distance and time was also calculated for each AHS sample subject in each CBSA. Commuting was considered in-vehicle if according to the AHS the mode of transportation was car, truck, van, bus/streetcar, taxicab, or other vehicle. This in-vehicle mode of transportation was then combined with the distance traveled in miles and the time traveled in minutes.

Cluster analysis is based on the distance between points so variables need to be scaled appropriately. If variables are measured on different scales, or units variables with large values contribute more to the distance measure than variables with small values. Therefore, the variables were standardized prior to performing the cluster analysis. All variables were standardized to a mean of 0 and standard deviation of 1.

2.2 Selection of cites

The total number of CBSAs covered in the national and metropolitan surveys from 2001-2005 was 148. The population of the CBSA largely determines the daily number of clinical events, such as mortality and hospitalizations, and thus the statistical power to detect potential adverse health effects of air pollutants, as reflected in the confidence intervals around their effect estimates. Small CBSAs with relatively few daily events will have more uncertainty surrounding their city-specific effect estimates and less statistical power to detect potential adverse health effects of air pollutants. From a previous report we determined that populations of less than 500,000 would not provide enough daily deaths to perform a time-series analysis with sufficient power to detect significant associations between $PM_{2.5}$ and mortality (Baxter et al., 2012). As a result, for this analysis of the 148 CBSAs that are included in AHS, we focused on the 94 CBSAs with a population greater than 500,000 people. Population data for these 94 CBSAs was obtained from the U.S. Census Bureau's website (United States Census Bureau).

2.3 Cluster Analysis

We used a k-means clustering algorithm to cluster CBSAs based on residential infiltration factors and commuting patterns. This iterative algorithm searches for a local solution that minimizes the Euclidean distance between the observations and the cluster centers. The k-means clustering algorithm is somewhat less sensitive to outliers than hierarchical clustering methods (Punj and Stewart, 1983). In a k-means cluster analysis the number of clusters (k) must be assigned a priori based either on pre-existing knowledge of the data or observable characteristics of the data set. For our analysis there was no pre-existing knowledge of the number of unique clusters to specify. We, therefore, calculated the within groups sum of squared errors (SSE) for 14 cluster solutions with k ranging from 2 to 15 to identify an optimal number of clusters.

SSE is defined as the sum of the squared distance between each member of a cluster and its cluster centroid (Kaufman and Rousseeuw, 1990) as shown below.

$$SSE = \sum_{i=1}^{K} \sum_{x \epsilon C_i} dist(c_i, x)^2$$

where K is the number of clusters; x is a city; C_i is the i^{th}cluster; *dist* is the standard Euclidean distance between two objects of Euclidean space; and c_i is the centroid of cluster C_i. In general, as the number of clusters increases, the SSE should decrease because clusters are, by definition, smaller. A plot of the SSE against a series of sequential cluster levels can provide a useful graphical way to choose an appropriate cluster level. The most appropriate cluster solution is defined as the solution at which the reduction in SSE slows dramatically. This produces an "elbow" in the plot of SSE against cluster solutions.

We conducted two separate cluster analyses. The first only included the residential infiltration variables (i.e. percent of homes with central AC, mean year home was built, and mean home size). The second analysis incorporates both the infiltration and the commuting variables (i.e. mean in-vehicle commuting time and mean in-vehicle commuting distance). The SSE plots were generated with R 2.14.0 (R Development Core Team, 2011) and the cluster analyses were conducted using SAS 9.3 (SAS Institute Inc., 2011).

3 Results

3.1 Description of Variables

Table 1 presents the summary statistics of the residential infiltration and commuting factors across all of the 94 CBSA. To describe the range of the data we calculated the ratios between the maximum and minimum values. Across CBSAs the ratios of percent of homes with central AC and mean home size were 13.7 and 2.2, respectively. Mean year home was built ranged across the CBSAs by approximately 60 years. For the commuting variables, these ratios were approximately 2 for both in-vehicle commuting time and distance. Using the coefficient of variation (CV), we determined the variability for each exposure factor across CBSAs. The CV for percent of homes with central AC was 44%, 24% for year built, and approximately 16% for the other factors. Of note the CV for year built was calculated using the age of the home (2013 - year built). See Supplemental Material, Table 1S for exposure factor values for each CBSA. There were geographical patterns to some of the factors. Older homes and homes with lower percentage of central AC tended to be found in the Northeast. Lower percentage of homes with central AC was also observed in California CBSAs, while the CBSAs in the South consisted of more homes with central AC.

The Spearman correlations among the exposure factors are shown in Table 2. The percent of home with central AC was highly correlated ($\rho = 0.71$) with mean year home was built. Mean year home was built was weakly correlated with mean in-vehicle commuting distance at 0.26. Mean in-vehicle commuting time and distance were strongly correlated ($\rho = 0.79$). While the overall correlations of these variables were high in a few cities (e.g. Las Vegas, NV and San Francisco, CA) the correlations were 0.5-0.6 we therefore chose not to combine these two variables. All other correlations between the factors were weak ($\rho < 0.3$).

3.2 Selecting k

The first step in analyzing the data using k-means clustering was to select the correct number of clusters. The selection of the correct number of clusters is not governed by a defined approach, but encompasses analyzing the within group SEE. To identify the appropriate number of clusters we compared the within groups SSE for a number of cluster solutions ranging from 2 to 15. By plotting the number of clusters against the within groups sum of squares for the residential infiltration factors (Figure 1), and both the infiltration and commuting factors (Figure 2) it is possible to not only quantitatively, but also visually identify a representative number of clusters. Figure 1 shows that similar values of the within group sum of squares appear for values *k* greater than 5. We therefore determined that a 5 cluster solution was a reasonable choice for this dataset. Figure 2 shows that similar values of the within group sum of squares appear for values *k* greater than 10. We therefore determined that a 10 cluster solution was a reasonable choice for this dataset.

3.3 Cluster Results with Residential Infiltration Factors

As determined from the SSE analysis a 5 cluster solution was used for the residential infiltration factors. The 94 CBSAs listed by clusters are shown in Table 2S (see Supplemental Material, Table 2S). Figure 3 and Table 3 presents the characteristics of the 5 clusters. CBSAs with lower percentage of homes with central AC were observed in clusters 1 and 5. Mean home age were older in clusters 1, 4, and 5 (prior to 1970), with CBSAs in cluster 5 having slightly older homes. In contrast, the homes in the CBSAs in cluster 5 were much larger than those assigned to cluster 1. CBSAs with the highest percentage of homes with central AC were grouped into clusters 2 and 3, with similar percentages seen in each cluster. However, cluster 2 consisted of CBSAs with newer and larger homes. Cluster 4 contained CBSAs with mean home ages similar to cluster 1; larger home sizes were found for clusters 2 and 5; however their prevalence of central AC were very different. For some of the clusters there appeared to be a geographical pattern as illustrated by Figure 4. The majority of cities in cluster 1 appear to be near bodies of water and almost all of the southeastern CBSAs are grouped into cluster 3.

Cluster 4 contained mostly midwestern cities and cluster 5 consisted of all northeastern CBSAs. A complete of all cities by cluster are listed in Table 2S.

3.4 Cluster Results with Addition of Commuting Factors

The 94 CBSAs listed by clusters are shown in Table 3S (see Supplemental Material, Table 3S). As determined from the SSE analysis a 10 cluster solution was used for the analysis including both the residential infiltration and commuting factors. The characteristics of the 10 clusters are presented on Figure 5 and Table 4. CBSAs with highest percentage of homes with central AC were grouped into clusters 3, 6, and 9, and CBSAs with the lowest percentages were in clusters 2, 5, and 7. Cluster 3 had the CBSAs with the newest homes while clusters 5, 7, and 10 had CBSAs with the oldest homes. Clusters consisting of CBSAs with the largest home included clusters 1, 3, 4, and 5, and the remaining clusters characterized by smaller homes. Mean in-vehicle commuting time was similar across clusters, with CBSAs with the longest commuting times in cluster 1 and CBSA with the shortest commuting times in cluster 10. Finally, CBSAs with further in-vehicle commuting distances were grouped into clusters 1, 3, 6, 7, and 8. As with the results with the infiltration factors a few geographical patterns did emerge. The majority of cluster 4 was made up of midwestern CBSAs, cluster 7 consisted of all northeastern CBSAs, and northeastern and midwestern CBSAs grouped into cluster 10 (Figure 6).

4 Discussion

The inability to explain the regional heterogeneity, specifically city-to-city heterogeneity within a region, in $PM_{2.5}$ mortality risk estimates observed in multi-city studies remains a key uncertainty in the examination of the PM-mortality relationship. This heterogeneity in PM mortality risk estimates has often been attributed to differences in (a) PM composition (Franklin et al., 2008), (b) PM exposure (i.e., exposure error) (Baxter et al., 2013a), and/or (c) community-specific characteristic such as demographics (Bell and Ebisu, 2012).

Our previous analysis (Baxter et al., 2013b) examined the first component of this hypothesis, PM composition. Unlike previous studies that focused on identifying the most toxic components (Bell et al., 2007; Franklin et al., 2008; Ostro et al., 2007). Baxter et al. (2013) focused on trying to identify city-to-city differences in PM composition that could explain the city-to-city heterogeneity observed within regions. While we did not find clear evidence of compositional differences between cities that could explain differences in $PM_{2.5}$ mortality risk estimates, a cursory analysis found some evidence that city-specific exposure differences, such as percent of population living in apartments and the prevalence of homes with central air conditioning, between cities may help explain the difference in $PM_{2.5}$ mortality risk estimates. Using a novel approach not previously applied to exposure factors, we examined city-specific differences in population exposures for 94 CBSAs that were clustered into groups with similar exposure distributions.

In the first analysis, cities were clustered based on residential infiltration factors. Previous studies have observed that as the prevalence of central air conditioning across cities increases, PM-mortality risk estimates decrease (Bell et al., 2009; Franklin et al., 2007; Janssen et al., 2002). Building on this concept, Chen el al. (2012)found evidence that seasonal and regional differences in PM_{10} mortality coefficients reflect seasonal and regional differences in total PM_{10} exposure per unit change in outdoor exposure. This study used a number of exposure factors including ones representative of residential infiltration, specifically the leakiness of the home (determined by the age and size of the home) and fraction of residences with central AC.

The analysis using only residential infiltration factors resulted in 5 clusters. Cluster 1 consisted of cities with older and smaller homes with less central AC while cities with newer, larger homes and a large percentage of central AC comprised Cluster 2. For the remaining clusters, cluster 3 is high prevalence of AC with newer and smaller homes; 4 is moderate prevalence of AC with older and large homes; and Cluster 5 is low prevalence of AC with older and larger homes. These 5 clusters cover the most common combinations in the U.S. housing stock. To further examine if additional city-specific exposure differences are contributing to the difference in $PM_{2.5}$ mortality risks between cities, exposure factors representative of commuting patterns (both distance and time) were included in an additional cluster analysis along with the infiltration factors in an attempt to more fully understand the exposure distributions of each city. In both analyses there were some geographical patterns. For the residential infiltration factors only analysis the southeastern CBSAs tended to group together (cluster 3) while for the infiltration plus commuting factors analysis, cluster 4 consisted of mostly midwestern CBSAs.

It was hypothesized that additional information on commuting patterns could further explain differences between cities that were not evident when examining only residential infiltration factors. With the American population spending on average 6.6% of its time (95 min/day) commuting (Klepis, 1999), in-vehicle exposures represent an important consideration when examining personal exposure to air pollution. Commuting times and distances can also represent personal mobility. Ignoring daily mobility patterns can introduce exposure measurement error and therefore bias into an epidemiological study (Setton et al., 2011). The addition of the commuting factors to the analysis resulted in 10 clusters. Overall, clusters with shorter in-vehicle commuting times had shorter in-vehicle commuting distances. In addition, cities with newer homes tended to have longer commuting times and distances possibly indicating urban sprawl. For the rest of the exposure factors no clear patterns emerged; however, this was not surprising given the weak correlations between these factors.

These clusters could be used to investigate potential explanations for city-to-city heterogeneity in PM mortality risk estimates. We, therefore, examined our results in the context of PM_{10} mortality risk estimates from the National Morbidity Mortality and Air Pollution Study (NMMAPS) (Dominici et al., 2003). Of the 88 cities included in NMMAPS, 66 were included in our cluster analyses. We focused on the clusters based solely on residential infiltration factors and examined the differences in PM_{10} mortality risk estimates between cities in clusters 2 (i.e., Atlanta, GA; Kansas City, MO; Columbus, OH; and Charlotte, NC) and 5 (i.e., Newark, NJ; Syracuse, NY; and Boston, MA), representing cities with high and low residential infiltrations, respectively. Cluster 2 is representative of cities with newer homes (i.e., mean age of 1989) with a higher percentage of central AC (i.e., 72.1%); whereas homes in cluster 5 are much older (i.e., mean age of 1945) and have a lower percentage of central AC (i.e., 18.9%). We would expect higher exposures to outdoor PM in clusters with cities exhibiting higher residential infiltration resulting in larger PM_{10} mortality risk estimates. Based on the maximum likelihood estimates (MLE) using GLM from Dominici et al. (2003), the magnitude of PM_{10} risk estimates is generally smaller for the cities in cluster 2, except for Charlotte, compared to cluster 5. It is important to note that the cluster analysis may be more informative when examining $PM_{2.5}$ mortality risk estimates due to the difference in particle size and reactivity between $PM_{2.5}$ and PM_{10} that could influence PM infiltration in the home.

Although the cluster analysis based on residential infiltration factors alone provides some insight with regard to the city-to-city heterogeneity in PM_{10} mortality risk estimates, other factors are also contributing to the heterogeneity observed. As such, clusters based on infiltration and commuting factors were also examined in the context of NMMAPS. We recognize the clusters based on infiltration and commuting factors are difficult to interpret; however, when taking them into consideration to explain differences within clusters from the infiltration factor only analysis additional information is gained. For example, within cluster 2, although all of the cities have relatively similar commuting exposure distributions, when examining infiltration plus commuting clusters the infiltration distribution of Atlanta is found to be slightly different than that of the other 3 cities, which could explain the difference in PM_{10} mortality risk estimates. Overall, this exercise may indicate that grouping cities into clusters based on exposure distributions does provide additional information that can help explain the city-to-city heterogeneity in PM mortality risk estimates, but it also brings to light that additional uncertainties remain that need to be examined to more full characterize the city-to-city differences observed.

It is important to recognize that this study is subject to inherent limitations. The main limitation of the clustering analyses conducted in this study is that the temporal and seasonal patterns in exposure as well as air pollution were not considered. It is well known that AC use and other factors that influence residential infiltration, such as the opening of windows, follow strong seasonal patterns indicative of an inverted U-shape curve (Koutrakis et al., 2005) and can be affected by the number of heating and cooling degree days. By excluding this information in conducting the cluster analysis it is possible that the cities within each cluster may not have very similar exposure distributions. However, the spearman correlation between cooling degree days and percent of central AC was quite strong ($\rho = 0.7$).Our analysis also does not include all factors affecting exposure such as socioeconomic status. People living in poorer and more disadvantaged neighborhoods may have higher exposures as a result of living closer to sources(Zeka et al., 2006).

While the AHS attempts to select a representative sample of homes it still may not be representative of the population introducing some exposure misclassification bias. The AHS uses multiple counties for each metropolitan area which are not aligned with the concentration and health data that focused on the core counties. However, since the majority of the population lives in the core county we can assume that the AHS values for the larger metropolitan area are indicative of the core county. Our exposure factors are also surrogates to capture residential infiltration rather than direct measurements potentially resulting in some exposure error. In addition, measures of the mean may not adequately capture the within-community variability which may be greater than the between-community variability.

An additional limitation of the cluster analysis is the sole reliance on minimizing SSE to determine the correct number of unique clusters. However, due to the lack of prior information that could be used to determine the correct number of unique clusters it is unclear how the reliance on SSE influences the overall results of the analyses. Finally, some of the clusters consist of only a small number of cities. This could indicate that the k we used is too large or alternatively it could mean those cities are really unique and should be in their own clusters.

5 Conclusions

This is the first study to employ cluster analysis to group cities with similar exposure distributions. Identifying cities with similar exposure distributions may help explain city-to-city heterogeneity in $PM_{2.5}$ mortality risk estimates. To date, the air pollution literature has focused on PM composition and demographics in identifying heterogeneity in the observed $PM_{2.5}$ mortality risk estimates. This study builds on previous work that examined whether differences in composition within- and between-cities can explain the heterogeneity, and in combination shows that compositional as well as exposure information can provide additional insight. Additional research is warranted to examine the combination of composition, exposure differences and demographics to fully understand how each influences PM mortality risk estimates.

Acknowledgements:

The authors would like to thank Kristin Isaacs of the U.S. EPA's National Exposure Research Laboratory and Tom Long of the U.S. EPA's National Center for Environmental Assessment for their review of this paper.

Disclaimer

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

References

Allen R, Larson T, Sheppard L, Wallace L, Liu LJS. Use of real-time light scattering data to estimate the contribution of infiltrated and indoor-generated particles to indoor air. Environmental Science and Technology 2003; 37: 3484-3492.

Baxter L, Franklin M, Özkaynak H, Schultz B, Neas L. The use of improved exposure factors in the interpretation of fine particulate matter epidemiological results. Air Quality, Atmosphere & Health 2013a; 6: 195-204.

Baxter LK, Duvall RM, Sacks JD. Examining the effects of air pollution composition on within region differences in $PM_{2.5}$ mortality risk estimates. Journal of Exposure Science and Environmental Epidemiology 2013b; advance online publication, December 19, 2012.

Baxter LK, Teet S, Neas LM. Method to Select Metropolitan Areas of Epidemiologic Interest for Enhanced Air Quality Monitoring. United States Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory, 2012.

Bell ML, Dominici F. Effect modification by community characteristics on the short-term effects of ozone exposure and mortality in 98 US communities. American Journal of Epidemiology 2008; 167: 986-997.

Bell ML, Dominici F, Ebisu K, Zeger SL, Samet JM. Spatial and temporal variation in $PM_{2.5}$ chemical composition in the United States for health effects studies. Environmental Health Perspectives 2007; 115: 989-995.

Bell ML, Ebisu K. Environmental inequality in exposures to airborne particulate matter components in the United States. Environmental Health Perspectives 2012; 120: 1699-1704.

Bell ML, Ebisu K, Peng RD, Dominici F. Adverse health effects of particulate air pollution: modification by air conditioning. Epidemiology 2009; 20.

Burke JM, Zufall MJ, Ozkaynak H. A population exposure model for particulate matter: case study results for $PM_{2.5}$ in Philadelphia, PA. Journal of Exposure Analysis and Environmental Epidemiology 2001; 11: 470-489.

Chan WR, Nazaroff WW, Price PN, Sohn MD, Gadgil AJ. Analyzing a database of residential air leakage in the United States. Atmospheric Environment 2005; 39: 3444-3455.

Department of Housing and Urban Development. American Housing Survey.

Dominici F, McDermott A, Zeger SL, Samet JM. National maps of the effects of particulate matter mortality: exploring geographic variation. Environmental Health Perspectives 2003; 111: 39-43.

Dominici F, Peng RD, Bell ML, Pham L, McDermott A, Zeger SL, et al. Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. Journal of the American Medical Association 2006; 295: 1127-1134.

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

Franklin M, Zeka A, Schwartz J. Association between $PM_{2.5}$ and all-cause and specific-cause mortality in 27 US communities. Journal of Exposure Science and Environmental Epidemiology 2007; 17: 279-287.

Hodas N, Lunden MM, Meng QY, Baxter LK, Özkaynak H, Burke J, et al. Heterogeneity in the fraction of ambient $PM_{2.5}$ found indoors contributes exposure error and may contribute to spatial and temporal differences in reported $PM_{2.5}$ health effect estimates. Journal of Exposure Science and Environmental Epidemiology 2012; 22: 448-454.

Janssen NAH, Schwartz J, Zanobetti A, Suh HH. Air conditioning and source specific particles as modifiers of the effect of PM_{10} on hospital admissions for heart and lung disease. Environmental Health Perspectives 2002; 110: 43-49.
Johnson T, Long T. Determining the frequency of open windows in residences: a pilot study in Durham, North Carolina during varying temperature conditions. Journal of Exposure Analysis and Environmental Epidemiology 2005; 15: 329-349.
Kaufman L, Rousseeuw R. Finding groups in data: an introduction to cluster analysis. London: Jonhn Wiley and Sons, 1990.
Klepis NE. An introduction to the indirect exposure assessment approach: modeling human exposure using microenvironmental measurements and the recent National Human Activity Pattern Survey Environmental Health Perspectives 1999; 107: 365-374.
Koenig JQ, Mar TF, Allen RW, Jansen K, Lumley T, Sullivan JH, et al. Pulmonary effects of indoor- and outdoor-generated particles in children with asthma. Environmental Health Perspectives 2005; 113: 499-503.
Koutrakis P, Suh HH, Sarnat JA, Brown KW, Coull BA, Schwartz J. Characterization of particulate and gas exposures of sensitive subpopulations living in Baltimore and Boston. Health Effects Institute, Boston, MA, 2005.
Levy JI, Chemerynski SM, Sarnat JA. Ozone exposure and mortality: an empiric Bayes metaregression analysis. Epidemiology 2005; 16: 458-468.
Medina-Ramon M, Zanobetti A, Schwartz J. The Effect of Ozone and PM10 on Hospital Admissions for Pneumonia and Chronic Obstructive Pulmonary Disease: A National Multicity Study. American Journal of Epidemiology 2006; 163: 579-588.
Ostro B, Feng WY, Broadwin R, Green S, Lipsett M. The effects of components of fine particulate air pollution on mortaltity in California: results from CALFINE. Environmental Health Perspectives 2007; 115: 13-19.
Pandian MD, Ott WR, Behar JV. Residential air exchange rates for use in indoor air and exposure modeling studies. Journal of Exposure Analysis and Environmental Epidemiology 1993; 3: 407-416.
Punj G, Stewart DW. Cluster analysis in marketing research: review and suggestions for application. Journal of Market Research 1983; 20: 134-148.
R Development Core Team. R: A Language Environment for Statistical Computing, Vienna, Austria, 2011.
Riediker M, Willams R, Devlin RB, Griggs TR, Bromberg PA. Exposure to particulate matter, volatile organic compounds, and other air pollutants inside patrol cars. Environmental Science and Technology 2003; 37: 2084-2093.
SAS Institute Inc. Version 9.3 of the SAS System for Windows. Copyright 2002-2010, Cary, NC USA, 2011.
Setton E, Marshall JD, Brauer M, Lundquist KR, Hystad P, Keller P, et al. The impact of daily mobility on exposure to traffic-related air pollution and health effect estimates. Journal of Exposure Science and Environmental Epidemiology 2011; 21: 42-48.
Shields KN, Cavallari JM, Hunt MJ, Lazo M, Molina M, Molina L, et al. Traffic-related air pollution exposures and changes in heart rate variability in Mexico City: A panel study. Environmental Health Perspectives 2013; 12.
Suh HH, Koutrakis P, Spengler JD. The relationship between airborne acidity and ammonia in indoor environments. Journal of Exposure Analysis and Environmental Epidemiology 1994; 1: 1-23.

Suh HH, Spengler JD, Koutrakis P. Personal exposures to acid aerosols and ammonia. Environmental Science and Technology 1992; 26: 2507-2516.
United States Census Bureau. Annual Estimates of the Population of Metropolitan and Micropolitan Statistical Areas: April 1, 2000 to July 1, 2009 (CBSA-EST2009-01).
Zeka A, Zanobetti A, Schwartz J. Individual-Level Modifiers of the Effects of Particulate Matter on Daily Mortality. American Journal of Epidemiology 2006; 163: 849-859.