Spatial Cluster of Air Pollutants and its Association with Health Disparities: A County-level Ecological Study across the USA

Authors: Jing Wang[1], Rong Rong[2], Bingjie Qu[3], Xiang Shi[4], Bin Hu[5], Peng Zhao[6], Chengxiu Ling[3], Ying Chen[3]

Affiliations:

[1]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA.

[2]Department of Health Policy and Management, School of Public Health and Tropical Medicine, Tulane University, New Orleans, Louisiana, USA.

[3]Wisdom Lake Academy of Pharmacy, Xi'an Jiaotong-Liverpool University, Suzhou, China.

[4]Department of Biology, School of Science, Xi'an Jiaotong-Liverpool University, Suzhou, China.

[5]Key Lab for Biomedical Effects of Nanomaterials and Nanosafety, Chinese Academy of Sciences, National Center for Nanoscience and Technology, Beijing, China.

[6]Department of Health and Environmental Sciences, School of Science, Xi'an Jiaotong-Liverpool University, Suzhou, China.

Corresponding author: Ying Chen, Ying.Chen01@xjtlu.edu.cn, ORCID: 0000-0002-5919-743X, Wisdom Lake Academy of Pharmacy, Xi'an Jiaotong - Liverpool University, Suzhou, 215213, China

Journal for submission: GeoHealth

Key Points:

- $PM_{2.5}$ mass, PM_{10} speciation, and $NONO_xNO_y$ were identified as the three most life expectancy-associated air pollutants.
- Five spatial distribution clusters were identified across the USA based on the concentrations of the three air pollutants.
- Severe air pollution clusters were associated with higher mortality rates caused by respiratory and cardiovascular diseases, and neoplasms.

Plain Language Summary

Air pollution has been one of the major threats to public health. This ecological study first identified three most health-related air pollutants, which were $PM_{2.5}$ mass, PM_{10} speciation, and $NONO_xNO_y$, from 12 types of classical air pollutants. It also investigated five spatial patterns of the three pollutants across the USA based on their county-level concentrations. We finally assessed the associations between the five clusters and three population health metrics, including life expectancy at birth, age-specific mortality risks, and cause-specific mortality rates. We found populations lived in counties exposed to higher levels of all three pollutants were associated with lower life expectancy, increased mortality risk among people aged over 45, and increased mortality rate caused by chronic respiratory conditions, cardiovascular diseases, and neoplasms. These findings provide a novel perspective to explain the existing health disparities, and have implications on future policy making and resources allocation regarding the prevention and treatment of air pollution.

Abstract

Background: The study aimed to determine latent patterns of geographical distribution of health-related air pollutants across the USA, and to evaluate real-world cumulative effects of these patterns on public health metrics.

Methods: It was an ecological study using county-level data on the concentrations of 12 air pollutants (i.e., ozone, CO, NO_2 and SO_2, $PM_{2.5}$ mass, and speciation, PM_{10} mass and speciation, HAPs, VOCs, $NONO_xNO_y$ and lead) over 20 years, and the rigid measurements of population health including life expectancy at birth, age-specific mortality risks and cause-specific mortality rates (21 mutually exclusive disease groups). Latent class analysis (LCA) was used to identify the common clusters of life expectancy-associated air pollutants based on their concentration characteristics in the final studied counties (n=699). Multivariate linear regression analyses were then applied to assess the relationship between the LCA-derived clusters and health measurements with confounding adjustment.

Results: $PM_{2.5}$ mass, PM_{10} speciation, and $NONO_xNO_y$ were associated with life expectancy and thus were included in LCA. Five clusters were identified: the ‘all low’ cluster (n=115, 16.5%), ‘all medium’ cluster (n=285, 40.8%), ‘high particulates’ cluster (n=152, 21.8%), ‘all high’ cluster (n=136, 19.5%) and ‘mixed profile’ cluster (n=11, 1.6%). Cluster with a more severe pollutant profile was associated with a decreasing life expectancy, an increasing mortality risk among the middle-aged and elderly populations (≥45 years), and an increasing mortality rate caused by chronic respiratory conditions, cardiovascular diseases, and neoplasms.

Conclusions: Our study brings new perspectives of real-world geographical patterns of air pollution to explain health disparities across the USA.

Key words: air pollution, health inequality, geographical pattern, latent class analysis, epidemiology

Background

Air pollution is a complicated mixture composed of gaseous pollutants, particulates, and other toxic pollutants such as hazardous air pollutants (HAPs), volatile organic compounds (VOCs), and lead. It is one of the major global threats to environmental sustainability and public health. In general, air pollution exposure is associated with oxidative stress and inflammation in human cells, which may lay a foundation for many acute and chronic diseases (Enweasor et al., 2021; Hahad et al., 2020; Møller et al., 2014; X. Rao et al., 2018; Valavanidis et al., 2013; Wiegman et al., 2020; Zhang et al., 2020). Abundant research illustrates that various air pollutants could lead to adverse health consequences (Bascom et al., 1996; T.-M. Chen et al., 2007; Correia et al., 2013; Kim et al., 2018; Pope, 2007; Soni et al., 2018; Thurston, 2017).

Long-term ambient exposure to increased concentrations of NO_2 and SO_2 is a risk factor for cardiovascular and respiratory disease mortality (T.-M. Chen et al., 2007; Khaniabadi et al., 2017). Long-term exposure to NO was associated with respiratory syncytial virus infection, neurodegeneration, and diseases related to the nervous system (Bhatt & Everard, 2004; Pacher et al., 2007; Yun et al., 1997). Prolonged low-level exposure to CO was found to be associated with exacerbation of heart disease and neurological damage (Townsend, 2002). Furthermore, an increase in ozone concentration was found to be associated with excess deaths from respiratory conditions (Jerrett et al., 2009). Particulates, categorized mainly into $PM_{2.5}$ and PM_{10}, were also proved to be hazardous to health, including increased risks of mortalities from lung cancer and cardiovascular diseases and overall death (Dockery et al., 1993; Kim et al., 2018; Pope, 2007; Pope et al., 1995). Other toxic air pollutants such as VOCs and lead were found to be associated with asthma, nasopharyngeal cancer, hypertension, and they could affect the development and function of the central nervous system (Soni et al., 2018).

In addition to the effects of individual air pollutants, studies have reported the impacts of multiple air pollutants on human health. Air containing high levels of ozone, $PM_{2.5}$, and NO_2 was associated with impaired lung function and increased incidence of cardiovascular conditions (Li et al., 2022; Rice et al., 2013). Increased concentrations of $PM_{2.5}$, NO, and NO_2 were associated with a higher risk of low birth weight (Coker et al., 2016). Young populations exposed to high levels of $PM_{2.5}$, PM_{10}, SO_2, and NOx simultaneously were more likely to develop phlegm, bronchitis, and asthma (Qian et al., 2004).

The content of air pollutants differs among locations. For example, in the USA, a particularly high concentration of $PM_{2.5}$ was monitored in the eastern region (Beckerman et al., 2013; Bell et al., 2007; Y. Wang et al., 2020). The concentration of PM_{10} was relatively higher in the middle and western seaboard (Hart et al., 2009). Elevated levels of NO_2 were constantly recorded in both the eastern and southwestern regions (Hart et al., 2009; Y. Wang et al., 2020). Additionally, in research focusing on smaller regional scales, Ozone level was found to be high in the southeast of Seattle, and dense concentrations of $PM_{2.5}$, NO, and NO_2 were reported in the south of Los Angeles (Cooper & Peterson, 2000; Li et al., 2022).

Individual air pollutants, in association with particular health outcomes, have been well established in the current literature. However, a more systematic understanding of the spatial distribution of simultaneous exposure to various health-related air pollutants across a large geographical area and its cumulative influence on overall health disparities is limited. In this study, we first aimed to assess whether there were geographical associations between the concentrations of a wide variety of air pollutants and the life expectancy at birth of residents using the national data of the USA. Second, we aimed to use cluster analysis to determine the common latent patterns of the geographical distribution of these life expectancy-associated air pollutants. Finally, using these derived patterns, we aimed to evaluate the real-world collective effects of air pollutants on the measurements of overall health, including life expectancy at birth, age-specific mortality risks, and cause-specific mortality rates.

Materials and Methods

This is an ecological study with county or county equivalent as the sample unit. The USA county-level air pollutants data and rigid public health metrics, including life expectancy at birth, age-specific mortality risks and cause-specific mortality rates, were analyzed. For potential confounding adjustment, characteristics of population, socioeconomics, healthcare service, and residential environment and location data were also collected.

Database and variable

Daily records of twelve air pollutants within three categories, which are criteria gases (including Ozone, CO, NO_2, and SO_2), particulates (including $PM_{2.5}$ mass, $PM_{2.5}$ speciation, PM_{10} mass, and PM_{10} speciation), and toxics, precursors and lead (including HAPs, VOCs, $NONO_xNO_y$, and lead), at the county level between 1995 and 2014 (totally 20 years) were obtained from the USA Environmental Protection Agency (https://aqs.epa.gov/aqsweb/airdata/download_files.html#Annual) (U.S. Environmental Protection Agency, 2021). Technical information about the measured parameters, units, and classification of these air pollutants is shown in Table S1 in Supporting Information (U.S. Environmental Protection Agency, 2022). At least 1 of the 12 air pollutants was monitored in 1388 counties between 1995 and 2014. Annual average concentrations of the studied air pollutants from 1995 to 2014 were calculated based on the daily data.

County-level health metrics of the residential population in this study included life expectancy at birth, age-specific mortality risks, and 21 mutually exclusive age-standardized cause-specific mortality rates in 2014. The age-specific mortality risks were presented in the following age categories: 0-4, 5-24, 25-44, 45-65, and 65-84 years old (Institute for Health Metrics and Evaluation, 2017b). The 21 mutually exclusive causes of death were Communicable, maternal, neonatal and nutritional diseases, including 1) HIV/AIDS and tuberculosis, 2) diarrhea, lower respiratory and other common infectious diseases, 3) neglected tropical diseases and malaria, 4) maternal disorders, 5) neonatal disorders, 6) nutritional deficiencies, 7) other communicable, maternal, neonatal and nutritional diseases; Noncommunicable diseases, including 8) neoplasms, 9) cardiovascular diseases, 10) chronic respiratory diseases, 11) cirrhosis and other chronic liver diseases, 12) digestive diseases, 13) neurological disorders, 14) mental and substance use disorders, 15) diabetes, urogenital, blood and endocrine diseases, 16) musculoskeletal disorders, 17) other non-communicable diseases; and Injuries, including 18) transport injuries, 19) unintentional injuries, 20) self-harm and interpersonal violence and 21) forces of nature, war and legal intervention (Institute for Health Metrics and Evaluation, 2017a).

County-level information on population characteristics (including size, gender, age, ethnicity), socioeconomics (including educational level, annual median household income, unemployment rate, poverty rate), healthcare service (including medical insured rate, number of physicians per 1,000 population), and residential environment and location (including Rural Urban Continuum Code, latitude, longitude) were collected from the USA national official sources (Table S2 in Supporting Information) (Health Resources and Services Administration, 2021; Institute for Health Metrics and Evaluation, 2017a, 2017b; U.S. Census Bureau, 2014, 2021; U.S. Department of Agriculture, 2021).

Statistical approach and analysis

We used the mean concentrations over 20 years (from 1995 to 2014) to measure previous long-term exposure to air pollutants to evaluate its relationship with life expectancy in 2014. Univariate analyses were first carried out for the 12 individual air pollutants for potential candidates of risk factors. Multiple linear regression models with the backward selection method were then used to identify the final list of significant air pollutants associated with life expectancy, considering the possible confounding effects from each other. In this approach, a five-time repeated ten-fold cross-validation resampling scheme was carried out to assess the performance of the model via obtaining the mean squared error, without a high potential of biased estimation (James et al., 2013; Kuhn et al., 2023). This method was implemented 20 times, and the final set of air pollutants enrolled for later cluster analysis were those significantly associated with life expectancy at each time of the optimal model.

After the final list of life expectancy-associated air pollutants was determined, our study samples (USA counties) included in later cluster analysis were identified based on the following criteria: a county containing maximumly one unmonitored pollutant (missing data) from the final listed air pollutants. Where there were missing data, multiple imputation using the predictive mean matching method was implemented according to the standard approach through an avaliable package in R (the ‘mice’ package) (van Buuren & Groothuis-Oudshoorn, 2011).

Latent cluster analysis (LCA) was used to identify the common distribution patterns of the air pollutants across the USA. LCA-clustered counties were determined by the concentration features of the final listed air pollutants derived from the previous selection process. Before entry into LCA, data of air pollutant concentrations were classified into three categories (‘low’, ‘medium’, and 'high', ordinal data) according to the 33^{rd} and 67^{th} percentiles of their twenty-year mean concentrations. We reported Log-likelihood (LL) statistics with bootstrap p values, Bayes Information Criterion, and Consistent Akaike’s Information Criterion for each model containing cluster numbers from 1 to 10. LCA was carried out in the Latent GOLD (version 4.5) with Newton-Raphson algorithms and estimation-maximization being utilized for model parameters estimation (Vermunt & Magidson, 2005). One thousand different random starting values were applied, and each included 50 interactions. Bootstrap p values were determined to assess the model fit based on the LL statistics. The optimal model is the one with the largest number of clusters where the p-value remains significant at the desired significance level (5%). Each county was allocated into one cluster according to their posterior probabilities of belonging to each cluster. A mean posterior probability $\geq$ 0.7 for samples allocated to a cluster was considered a good assignment (Nagin & Kerner, 2005).

Multiple linear regression models were then developed to assess the associations between the LCA-derived clusters and life expectancy (2014 data), change in life expectancy (from 1995 to 2014), age-specific mortality risk (2014 data), and cause-specific mortality rate (2014 data), with adjustment for potential confounding factors including collected information on population characteristics, socioeconomics, healthcare service, and residential environment and location.

All statistical analyses were performed in R (version 4.0.4) except for LCA. To correct the effect of multiple testing, a p-value < 0.005 (instead of 0.05), two-tailed, was set as the threshold for statistical significance, aiming to obtain conservative results with a low level of false positive findings. The flowchart of database construction and statistical analysis is shown in Figure S1 in Supporting Information.

Results

Initial univariate analyses showed that concentrations of NO_2, $PM_{2.5}$ mass, PM_{10} mass, PM_{10} speciation, and $NONO_xNO_y$ were associated with life expectancy, whereas the other studied air pollutants (SO_2, CO, Ozone, $PM_{2.5}$ speciation, VOCs, HAPs and lead) were not. Multivariate analyses further indicated that only $PM_{2.5}$ mass, PM_{10} speciation, and $NONO_xNO_y$ were suggestive of consistent association with life expectancy (Table S3 in Supporting Information), which determined these three air pollutants to be included in the subsequent analyses.

Descriptive statistics of the final studied counties (n=699) regarding the concentrations of the three selected air pollutant, health measures, and socio-demographic variables are shown in Table 1. The geographical distribution of these counties is displayed in Figure S2 in Supporting Information, informing generally representative samples at the USA national level.

In LCA, the five-cluster model was determined as the optimal model (Table S4 in Supporting Information). Studied counties generally displayed high posterior probabilities for their assigned clusters, with mean posterior probabilities ranging from 0.66 to 0.93 across the five clusters (Table 2).

Cluster 1 (n=115, 16.5%) was featured with low concentrations of $PM_{2.5}$ mass, PM_{10} speciation and $NONO_xNO_y$ (the ‘all low’ cluster, Figure 1a). Cluster 2 (n=285, 40.8%), the most common cluster, displayed medium levels of all three air pollutants (the ‘all medium’ cluster, Figure 1b). Cluster 3 (n=152, 21.8%) was characterized by high $PM_{2.5}$ mass and PM_{10} speciation (the ‘high particulates’ cluster, Figure 1c), whereas Cluster 4 (n=136, 19.5%) had the highest levels of all the three air pollutants (the ‘all high’ cluster, Figure 1d). Cluster 5, with the smallest size (n=11, 1.6%), displayed a mixed profile: high $PM_{2.5}$ mass but low PM_{10} speciation (the ‘mixed profile’ cluster, Figure 1e).

The geographical distribution of the counties stratified by the five LCA-derived clusters is presented in Figure 1f. Counties in Cluster 1, characterized by low air pollutant concentrations, were primarily located in the middle-west and northeast regions. Most counties included in Clusters 3 and 4, which had relatively high pollution levels of the three air pollutants, were found in the middle-east and southwest coast regions. Counties classified into Cluster 2, with medium levels of all air pollutants, were distributed widely across the whole USA, whereas the distribution pattern of Cluster 5 was not to be summarized due to its small sample size.

Table S5 in Supporting Information shows the county-level descriptive statistics of the three air pollutants and health outcome measurements stratified by the five clusters. Average county-level life expectancy in Cluster 1 was the highest (79.33 years, standard deviation (SD) 1.84), while Cluster 4 was the lowest (77.40 years, SD 2.03). In multivariate analysis, compared to Cluster 1, Clusters 2, 3, and 4 were all significantly associated with a reduced county-level life expectancy at birth, with adjustment for collected potential confounding factors (left, Table 3). Similar results were obtained for life expectancy change as the outcome variable, however, only Clusters 3 and 4 reached statistical significance (right, Table 3).

Differences in mortality risks were seen between these LCA-derived clusters, specifically among middle-age and elderly populations (Figure 2a). For example, for the group of 45-64 years the mortality risks were 11.14% (SD 2.03%), 11.71% (SD 2.62%), 13.23% (SD 2.94%), 13.80% (SD 2.71%) and 12.80% (SD 3.56%) for those who lived in the counties of ‘all low’ (Cluster 1), ‘all medium’ (Cluster 2), ‘high particulates’ (Cluster 3), ‘all high’ (Cluster 4) and ‘mixed profile’ (Cluster 5) clusters, respectively (Table S5 in Supporting Information). Confounding adjusted difference in age-specific mortality risks between clusters is shown in Figure 2b (the ‘all low’ Cluster 1 as the referent group), indicating consistently increased mortality risks among middle-age and elderly populations in severely polluted regions.

For cause-specific mortality rates, health conditions associated with these LCA-derived clusters were neonatal disorders, diarrhea, lower respiratory and other common infectious diseases (in the communicable, maternal neonatal, and nutritional disease category, Figure 3a), neoplasms, cardiovascular diseases, and chronic respiratory diseases (in the noncommunicable disease category, Figure 3b), and self-harm and interpersonal violence (in the injury category, Figure 3c). In general, results suggested that clusters with more severe pollution were associated with increased mortality rates of these conditions, except for self-harm and interpersonal violence.

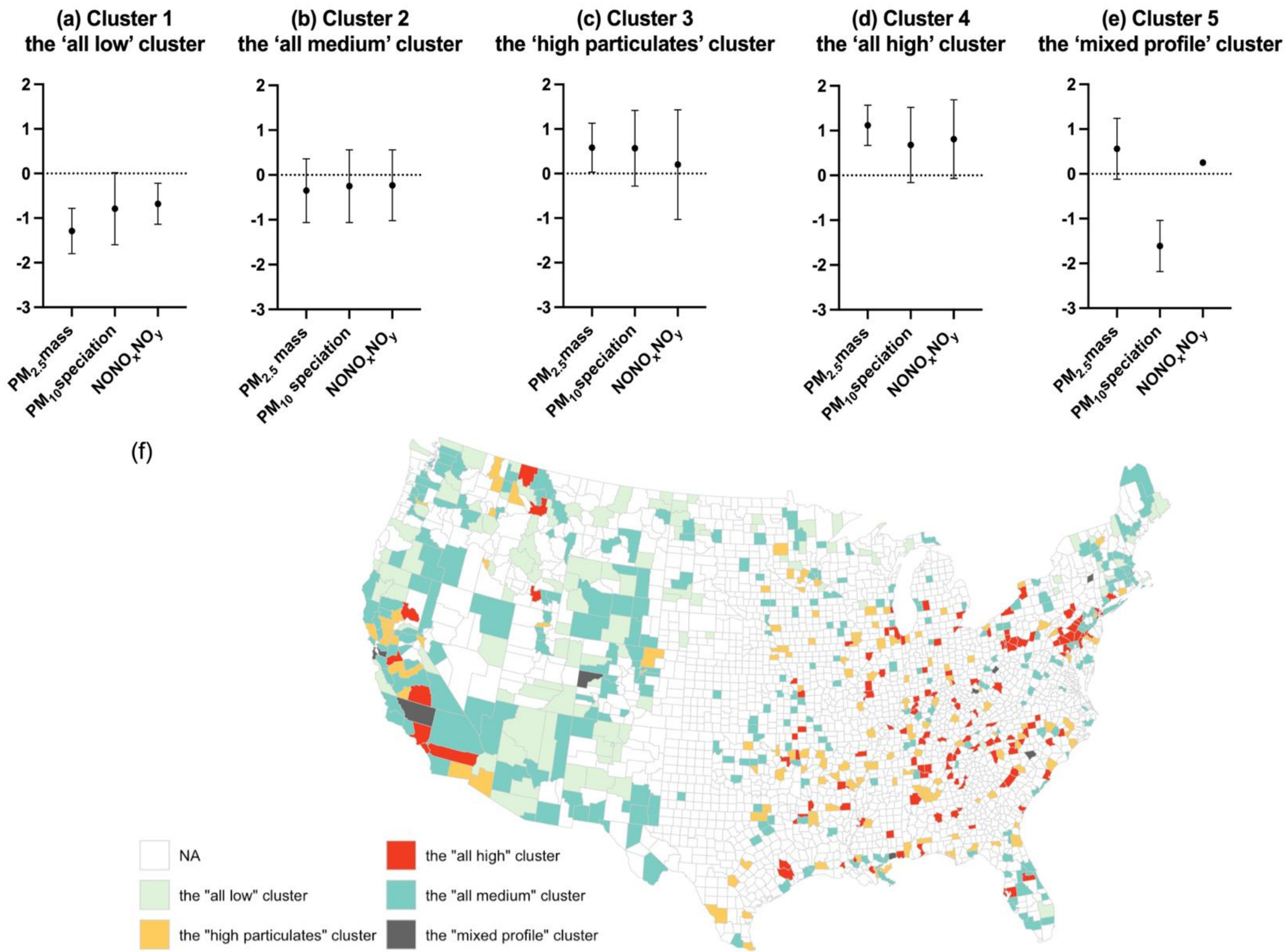

Fig 1. Concentration characteristics of $PM_{2.5}$ mass, PM_{10} speciation and $NONO_xNO_y$ by latent class analysis-derived clusters and their spatial distribution across the USA

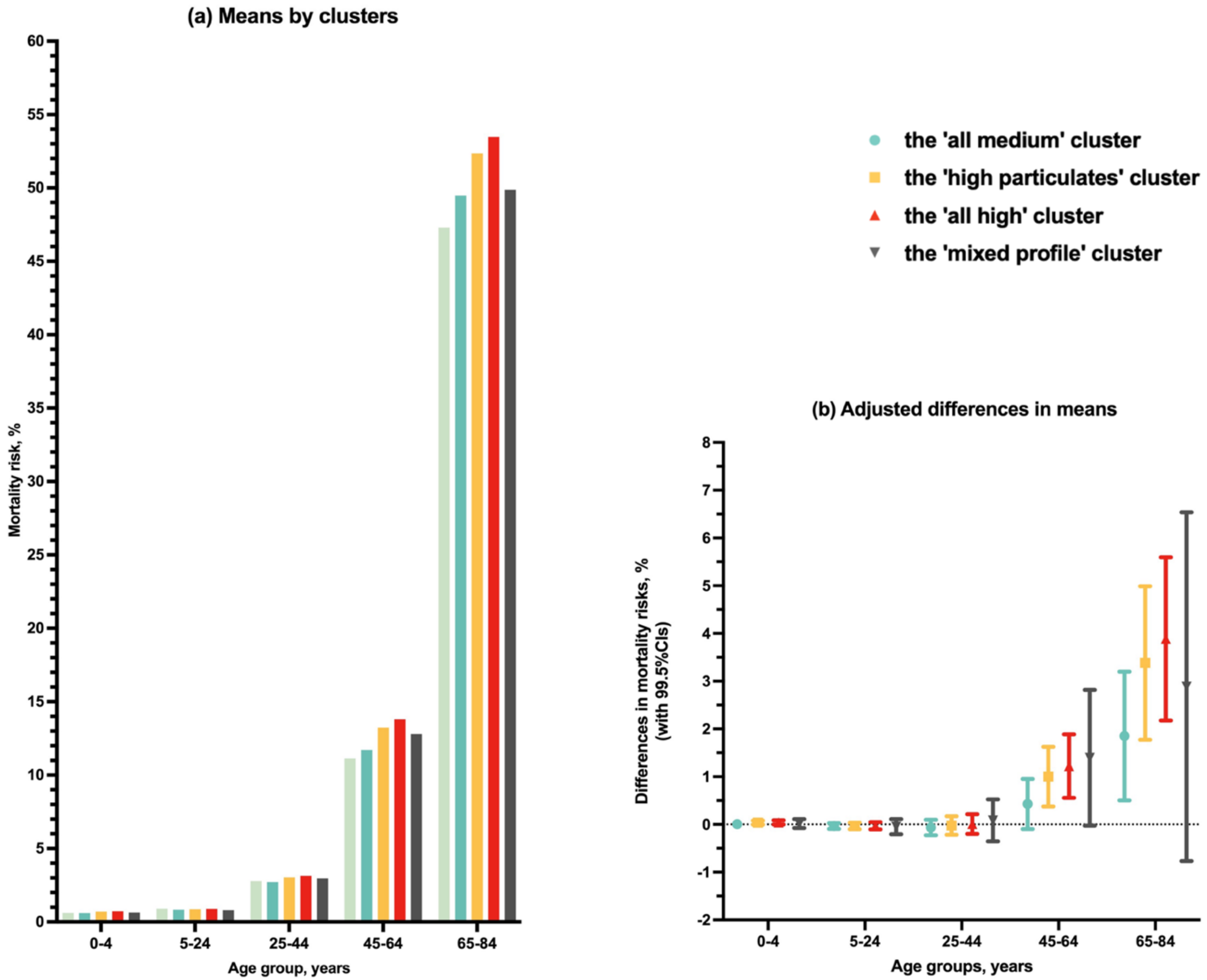

Fig 2. Age-specific mortality risk by latent class analysis-derived clusters of air pollutants

The 'all low' cluster (cluster 1) was the referent group; All models adjusted for population characteristics (including size, gender, and ethnicity), socioeconomics (including educational level, annual median household income, unemployment rate, and poverty rate), healthcare service (including medical insured rate, number of physicians per 1,000 population), and residential environment and location (including Rural-Urban Continuum Codes, latitude, and longitude).

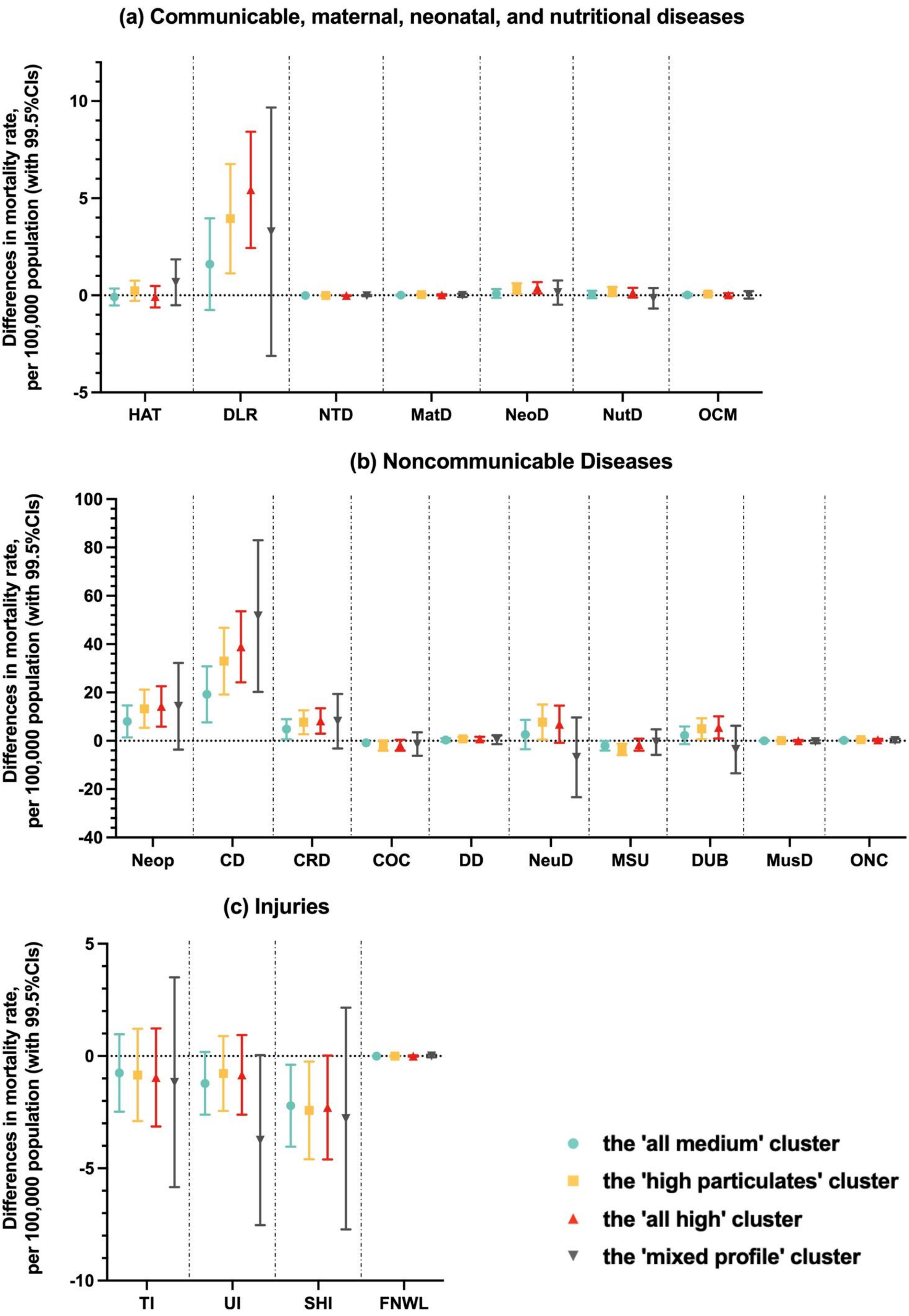

Fig 3. Cause-specific mortality rate by latent class analysis-derived clusters of air pollutants

The ‘all low’ cluster (cluster 1) was the referent group; All models adjusted for population characteristics (including size, gender, and ethnicity), socioeconomics (including educational level, annual median household income, unemployment rate, and poverty rate), healthcare service (including medical insured rate, number of physicians per 1,000 population), and residential environment and location (including Rural-Urban Continuum Codes, latitude, and longitude). Disease group: *HAT:* HIV/AIDS and tuberculosis; *DLR:* Diarrhea, lower respiratory and other common infectious diseases; *NTD:* Neglected tropical diseases and malaria; *MatD:* Maternal disorders; *NeoD:* Neonatal disorders; *NutD:* Nutritional deficiencies; *OCM:* Other communicable, maternal, neonatal, and nutritional diseases; *Neop:* Neoplasms; *CD:* Cardiovascular diseases; *CRD:* Chronic respiratory diseases; *COC:* Cirrhosis and other chronic liver diseases; *DD:* Digestive diseases; *DeuD:* Neurological disorders; *MSU:* Mental and substance use disorders; *DUB:* Diabetes, urogenital, blood and endocrine diseases; *MusD:* Musculoskeletal disorders; *ONC:* Other non-communicable diseases; *TI:* Transport injuries; *UI:* Unintentional injuries; *SHI:* Self-harm and interpersonal violence; *FNWL:* Forces of nature, war and legal intervention.

Discussion

Epidemiological research has been typically conducted only on specific air pollutants and diseases due to difficulties in design, cost, and management of a single study containing a comprehensive list of air pollutants and grand indicators of public health. Using recently available high-quality open-access databases, our ecological study at the USA national scale has made an effort to evaluate the real-world collective effects of air pollutants on human health, and could provide a novel perspective to explain the existing health disparities and a hint at future policy planning and resource allocation regarding the prevention and treatment of air pollution.

In this study, regional concentrations of air pollutants, including $PM_{2.5}$ mass, PM_{10} speciation, and $NONO_xNO_y$, were found to be associated with life expectancy. On the other hand, the regional concentrations of SO_2, CO, Ozone, $PM_{2.5}$ speciation, VOCs, HAPs, and lead were not associated. For the other two air pollutants, i.e., NO_2 and PM_{10} mass, our multivariate analysis with the resampling approach suggested that those apparent associations, which initially occurred in univariate statistics or individual resampling tests, were not robust.

Using LCA, we identified five clusters with distinct profiles of the three life expectancy-associated air pollutants, which altogether depicted the spatial distribution patterns of critical air pollutants across the USA. Previous research has described the spatial distributions of specific air pollutants separately, but no one has considered the patterns of multiple air pollutants in an integrated way, particularly concerning the most important indicator of population health, the life expectancy at birth (Beckerman et al., 2013; Bell et al., 2007; Cooper & Peterson, 2000; Hart et al., 2009; Y. Wang et al., 2020).

Further assessment of the relationship between these clusters and health measurements revealed that Clusters 3 (the ‘high particulates’ cluster) and 4 (the ‘all high’ cluster) were associated with lower life expectancy, higher age-specific mortality risks among middle-age and elderly populations, and higher rates of mortality caused by several diseases including those with the greatest burden (i.e., neoplasms, cardiovascular diseases, and chronic respiratory diseases). It was noticeable that the association of more severely polluted clusters with increased mortality risks was only seen among middle-aged and elderly populations, but not in younger people. This finding implies that the possible effects of prolonged exposure to these air pollutants on adverse health outcomes are very likely to be a slow and cumulative process. Previous studies provided varying degrees of evidence for the effects of these individual air pollutants on respiratory and cardiovascular systems and cancer (Atkinson et al., 2014; Z. Chen et al., 2022; Mohammadi et al., 2022; X. Rao et al., 2018; Sun et al., 2017; Turner et al., 2011). However, our study advances this research area by looking at the real-world patterns of multiple air pollutants within a geographical region and investigating their collective effects on diseases.

Disparities of air pollution in different regions have been a constant phenomenon across the USA. Previous studies suggested that the eastern region and California state were confronted with high concentrations of $PM_{2.5}$, while the western regions had a medium level of $PM_{2.5}$ pollution (Bell et al., 2007). Consistent with these findings, the spatial distribution pattern of studied air pollutants, as demonstrated in our study, indicated that the clusters with the highest levels of $PM_{2.5}$ mass (i.e., Clusters 3 and 4) were mainly located in the east part and southwest coast of the USA, whereas the cluster with a medium level of $PM_{2.5}$ mass (i.e., Cluster 2) showed a dense presence in the west of the country. However, regarding the distribution of PM_{10} speciation, a previous study found that the PM_{10} level in the western half of the USA was higher than in the east between 1985 and 2000 based on the geographical information system-based estimators (Hart et al., 2009), which is inconsistent with our identified patterns: most counties within Clusters 3 and 4 were gathered in the east of the USA. This variety might be due to the different measurement methods and study periods between the two studies. Furthermore, according to our study, Cluster 4, with the highest level of $NONO_xNO_y$, also appeared in California, which supports another study conducted in Los Angeles demonstrating a particularly high regional level of NO (Coker et al., 2016). Large cities with denser populations are accompanied by more transportation emissions, intenser economic activities releasing air pollutants, and thus severer air pollution (S. Wang et al., 2020). Our results demonstrated that many of these regions, such as the northeast metropolitan area, were classified as the ‘all high’ cluster (Cluster 4).

In previous literature, various research has demonstrated the substantial geographical disparities in life expectancy across the USA. For example, based on the data of 2014, the life expectancy at birth was generally low in the counties located in the middle east and southeast areas, such as the south region of Mississippi, western Virginia, and eastern Kentucky, but it was the highest in middle Colorado (Dwyer-Lindgren et al., 2017). The former areas overlapped with the counties classified into Clusters 3 and 4 analyzed in this study, whereas a considerable number of counties in Colorado were allocated to Cluster 1. Differences in life expectancy should refer to disparities in the occurrence and outcome of diseases. Previous research indicated that the southeast and middle-east regions of the USA (such as Alabama, Kentucky, Mississippi, and Tennessee) generally had high mortality rates of respiratory diseases (Vierboom et al., 2019). The west coasts, Texas and middle east regions experienced high mortality from cardiovascular diseases, whereas the mortality rates were much lower in the northeast and west regions (Dwyer-Lindgren et al., 2016; S. Rao et al., 2021). A similar pattern of geographical health disparity was also observed for the mortality rates of neurological disorders and neoplasms (Dwyer-Lindgren et al., 2016). The geographical association between health disparities and air pollutant clusters identified in this study suggests that the regional concentrations of $PM_{2.5}$ mass, PM_{10} speciation, and $NONO_xNO_y$, analyzed simultaneously way, may be an important influencing factor in the real-world setting for regional public health.

An ecological study using data reported at the county level is a feasible and effortful approach, as counties are the smallest administrative units in the USA where essential studied data can be obtained (Dwyer-Lindgren et al., 2016, pp. 1980–2014). This research took advantage of the exhaustive information for county-level demographics and socioeconomic conditions, healthcare provisions, and residual environmental statistics recorded by the US federal offices. It facilitated a thorough control for confounding factors in our statistical analyses. Although it is not available to include most of the USA counties in this study, the samples generally have a broad and representative geographical coverage of the country, ensuring the generalizability of the results. This study employs indicators of overall population health, such as life expectancy at birth and mortality rates, as opposed to more specific health measurements, such as disease status. It is important to acknowledge potential variations in the standards of disease identification and practice of disease recording among different counties. Unbiased information on specific health measurements across the USA is scarce. Moreover, some health indicators may be influenced by detection capabilities and be subject to survival bias. For instance, a higher reported incidence of lung cancer in a certain area might actually result from improved diagnostic methods or more effective treatments, subsequently extending survival times. In contrast, life expectancy and mortality statistics present fewer challenges, mainly when conducting expansive geographical investigations spanning multiple administrative regions. These metrics provide a less biased evaluation of population health.

There are some limitations in this study. Due to the issue of ecological fallacy from the nature of the ecological study, no inferences of the results could be made at the individual level (Sedgwick, 2014). It has not attempted to reveal possible mechanisms behind the observed patterns and associations. Again, as an ecological study, common issues, such as migration bias, may impact the results (Sedgwick, 2014). In this study, we summarized the prolonged exposure to air pollutants using the average level over time, however, other statistics could be used, such as those focusing on the extreme values. We only included 12 air pollutants, while other potentially important pollutants might be missed. The geographical pattern of air pollutants was determined by LCA with the mentioned selection criteria for the optimal model, however, the use of other statistical approaches may obtain slightly different results.

Conclusion

Our county-level ecological study identified the common geographical patterns of life expectancy-associated air pollutants across the USA. Five distinctive clusters were determined according to the 20-year concentration features of $PM_{2.5}$ mass, PM_{10} speciation, and $NONO_xNO_y$. In particular, the clusters, characterized by the regions having higher concentrations of these three air pollutants, were associated with a lower life expectancy, higher mortality risks among the middle-aged and elderly populations, and higher mortality rates of several specific causes including chronic respiratory diseases, cardiovascular diseases, and neoplasms. Our study brings new perspectives of real-world geographical patterns of air pollution to explain health disparities across the USA.

412

Open Research

Acknowledgements

The authors thank the other colleagues at the Wisdom Lake Academy of Pharmacy, Xi'an Jiaotong-Liverpool University for their academic and administrative support. YC would like to express his special appreciation for the arrival of his baby daughter, with whom the process of manuscript writing has become a particularly memorable and enjoyable moment.

Declarations

Competing interests: The authors have no conflicts of interest relevant to this article to disclose.

Funding: This study was supported by the Research Development Fund at Xi'an Jiaotong-Liverpool University (reference number: Research Development Fund-22-01-012).

Data Availability:

All air pollutants data were publicly accessible at https://aqs.epa.gov/aqsweb/airdata/download_files.html#Annual. County-level population health metrics, including life expectancy at birth and age-specific mortality risks, were extracted from the database provided by the Institute for Health Metrics and Evaluation at https://ghdx.healthdata.org/record/ihme-data/united-states-life-expectancy-and-age-specific-mortality-risk-county-1980-2014. County-level 21 mutually exclusive age-standardized cause-specific mortality rates, were extracted from the database provided by the Institute for Health Metrics and Evaluation at https://ghdx.healthdata.org/record/ihme-data/united-states-cancer-mortality-rates-county-1980-2014. County-level population size, gender, age, ethinic distribution, educational level, annual median household income, unemployment rate, poverty rate, insured population rate, and representative latitude and longitudes coordinates were extracted from the databases provided by the USA Bureau of the Census at https://data.census.gov/cedsci. County-level physicians' number were extracted from the database provided by the USA Health Resources and Services Administration at https://data.hrsa.gov/data/download. The Rural-Urban Continuum Code were extracted from the database provided by the USA Department of Agriculture at https://www.usda.gov/topics/data.

Ethical Approval: Not applicable.

Author contributions: Ying Chen conceived the study. Jing Wang analyzed the data. Jing Wang and Ying Chen drafted the initial manuscript. All authors contributed to the study design and interpretation of the data, and approved the final version of the manuscript submitted for publication.

References

Atkinson, R. W., Kang, S., Anderson, H. R., Mills, I. C., & Walton, H. A. (2014). Epidemiological time series studies of PM2.5 and daily mortality and hospital admissions: A systematic review and meta-analysis. *Thorax*, *69*(7), 660–665. https://doi.org/10.1136/thoraxjnl-2013-204492

Bascom, R., Bromberg, P. A., Costa, D. L., Devlin, R., Dockery, D. W., Frampton, M. W., Lambert, W., Samet, J. M., Speizer, F. E., & Utell, M. (1996). Health effects of outdoor air pollution. *American Review of Respiratory Disease*, *153*(2), 477–498. https://doi.org/10.1164/ajrccm.153.2.8564086

Beckerman, B. S., Jerrett, M., Serre, M., Martin, R. V., Lee, S.-J., van Donkelaar, A., Ross, Z., Su, J., & Burnett, R. T. (2013). A Hybrid Approach to Estimating National Scale Spatiotemporal Variability of $PM_{2.5}$ in the Contiguous United States. *Environmental Science & Technology*, *47*(13), 7233–7241. https://doi.org/10.1021/es400039u

Bell, M. L., Dominici, F., Ebisu, K., Zeger, S. L., & Samet, J. M. (2007). Spatial and Temporal Variation in $PM_{2.5}$ Chemical Composition in the United States for Health Effects Studies. *Environmental Health Perspectives*, *115*(7), 989–995. https://doi.org/10.1289/ehp.9621

Bhatt, J. M., & Everard, M. L. (2004). Do environmental pollutants influence the onset of respiratory syncytial virus epidemics or disease severity? *Paediatric Respiratory Reviews*, *5*(4), 333–338. https://doi.org/10.1016/j.prrv.2004.07.003

Chen, T.-M., Kuschner, W. G., Gokhale, J., & Shofer, S. (2007). Outdoor Air Pollution: Nitrogen Dioxide, Sulfur Dioxide, and Carbon Monoxide Health Effects. *The American Journal of the Medical Sciences*, *333*(4), 249–256. https://doi.org/10.1097/MAJ.0b013e31803b900f

Chen, Z., Liu, N., Tang, H., Gao, X., Zhang, Y., Kan, H., Deng, F., Zhao, B., Zeng, X., Sun, Y., Qian, H., Liu, W., Mo, J., Zheng, X., Huang, C., Sun, C., & Zhao, Z. (2022). Health effects of exposure to sulfur dioxide, nitrogen dioxide, ozone, and carbon monoxide between 1980 and 2019: A systematic review and meta-analysis. *Indoor Air*, *32*(11), e13170-n/a. https://doi.org/10.1111/ina.13170

Coker, E., Liverani, S., Ghosh, J. K., Jerrett, M., Beckerman, B., Li, A., Ritz, B., & Molitor, J. (2016). Multi-pollutant exposure profiles associated with term low birth weight in Los Angeles County. *Environment International, 91*, 1–13. https://doi.org/10.1016/j.envint.2016.02.011

Cooper, S. M., & Peterson, D. L. (2000). Spatial distribution of tropospheric ozone in western Washington, USA. *Environmental Pollution*, *107*(3), 339–347. https://doi.org/10.1016/S0269-7491(99)00172-4

Correia, A. W., Pope, C. A., Dockery, D. W., Wang, Y., Ezzati, M., & Dominici, F. (2013). Effect of Air Pollution Control on Life Expectancy in the United States: An Analysis of 545 U.S. Counties for the Period from 2000 to 2007. *Epidemiology*, *24*(1), 23–31. https://doi.org/10.1097/EDE.0b013e3182770237

Dockery, D. W., Pope, C. A., Xu, X., Spengler, J. D., Ware, J. H., Fay, M. E., Ferris, B. G., & Speizer, F. E. (1993). An Association between Air Pollution and Mortality in Six U.S. Cities. *New England Journal of Medicine*, *329*(24), 1753–1759. https://doi.org/10.1056/NEJM199312093292401

Dwyer-Lindgren, L., Bertozzi-Villa, A., Stubbs, R. W., Morozoff, C., Kutz, M. J., Huynh, C., Barber, R. M., Shackelford, K. A., Mackenbach, J. P., van Lenthe, F. J., Flaxman, A. D., Naghavi, M., Mokdad, A. H., & Murray, C. J. L. (2016). US County-Level Trends in Mortality Rates for Major Causes of Death, 1980-2014. *JAMA*, *316*(22), 2385.

https://doi.org/10.1001/jama.2016.13645

Dwyer-Lindgren, L., Bertozzi-Villa, A., Stubbs, R. W., Morozoff, C., Mackenbach, J. P., van Lenthe, F. J., Mokdad, A. H., & Murray, C. J. L. (2017). Inequalities in Life Expectancy Among US Counties, 1980 to 2014: Temporal Trends and Key Drivers. *JAMA Internal Medicine*, *177*(7), 1003. https://doi.org/10.1001/jamainternmed.2017.0918

Enweasor, C., Flayer, C. H., & Haczku, A. (2021). Ozone-Induced Oxidative Stress, Neutrophilic Airway Inflammation, and Glucocorticoid Resistance in Asthma. *Frontiers in Immunology*, *12*. https://doi.org/10.3389/fimmu.2021.631092

Hahad, O., Lelieveld, J., Birklein, F., Lieb, K., Daiber, A., & Münzel, T. (2020). Ambient Air Pollution Increases the Risk of Cerebrovascular and Neuropsychiatric Disorders through Induction of Inflammation and Oxidative Stress. *International Journal of Molecular Sciences*, *21*(12). https://doi.org/10.3390/ijms21124306

Hart, J. E., Yanosky, J. D., Puett, R. C., Ryan, L., Dockery, D. W., Smith, T. J., Garshick, E., & Laden, F. (2009). Spatial Modeling of PM_{10} and NO_2 in the Continental United States, 1985–2000. *Environmental Health Perspectives*, *117*(11), 1690–1696. https://doi.org/10.1289/ehp.0900840

Health Resources and Services Administration. (2021). *Data Downloads* [dataset]. https://data.hrsa.gov/data/download

Institute for Health Metrics and Evaluation. (2017a). *United States Cancer Mortality Rates by County 1980-2014*. https://ghdx.healthdata.org/record/ihme-data/united-states-cancer-mortality-rates-county-1980-2014

Institute for Health Metrics and Evaluation. (2017b). *United States Life Expectancy and Age-specific Mortality Risk by County 1980-2014*. https://ghdx.healthdata.org/record/ihme-data/united-states-life-expectancy-and-age-specific-mortality-risk-county-1980-2014

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: With Applications in R* (Vol. 103). Springer Nature. https://doi.org/10.1007/978-1-4614-7138-7

Jerrett, M., Burnett, R. T., Pope, C. A., Ito, K., Thurston, G., Krewski, D., Shi, Y., Calle, E., & Thun, M. (2009). Long-Term Ozone Exposure and Mortality. *New England Journal of Medicine*, *360*(11), 1085–1095. https://doi.org/10.1056/NEJMoa0803894

Khaniabadi, Y. O., Polosa, R., Chuturkova, R. Z., Daryanoosh, M., Goudarzi, G., Borgini, A., Tittarelli, A., Basiri, H., Armin, H., Nourmoradi, H., Babaei, A. A., & Naserian, P. (2017). Human health risk assessment due to ambient PM10 and SO2 by an air quality modeling technique. *Process Safety and Environmental Protection*, *111*, 346–354. https://doi.org/10.1016/j.psep.2017.07.018

Kim, S., Kim, H., Kim, H., Byun, G., Choi, Y., Song, H., Kim, S., & Lee, J.-T. (2018). Effects of Long-Term Exposure to Ambient Air Particles on Cause-Specific Mortality and Life Expectancy in Seoul, Korea. *ISEE Conference Abstracts*, *2018*(1), isesisee.2018.P01.0690. https://doi.org/10.1289/isesisee.2018.P01.0690

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., & Hunt, T. (2023). *caret: Classification and Regression Training*.

https://cran.r-project.org/web/packages/caret/index.html

Li, H., Deng, W., Small, R., Schwartz, J., Liu, J., & Shi, L. (2022). Health effects of air pollutant mixtures on overall mortality among the elderly population using Bayesian kernel machine regression (BKMR). *Chemosphere*, *286*, 131566. https://doi.org/10.1016/j.chemosphere.2021.131566

Mohammadi, M. J., Fouladi Dehaghi, B., Mansourimoghadam, S., Sharhani, A., Amini, P., & Ghanbari, S. (2022). Cardiovascular disease, mortality and exposure to particulate matter (PM): A systematic review and meta-analysis. *Reviews on Environmental Health*. https://doi.org/10.1515/reveh-2022-0090

Møller, P., Danielsen, P. H., Karottki, D. G., Jantzen, K., Roursgaard, M., Klingberg, H., Jensen, D. M., Christophersen, D. V., Hemmingsen, J. G., Cao, Y., & Loft, S. (2014). Oxidative stress and inflammation generated DNA damage by exposure to air pollution particles. *Mutation Research/Reviews in Mutation Research*, *762*, 133–166. https://doi.org/10.1016/j.mrrev.2014.09.001

Nagin, D. S., & Kerner, H.-J. (2005). Group-based modeling of development. *Annals of the American Academy of Political and Social Science*, *602*, 259–279. https://doi.org/10.1177/0002716205281245

Pacher, P., Beckman, J. S., & Liaudet, L. (2007). Nitric Oxide and Peroxynitrite in Health and Disease. *Physiological Reviews*, *87*(1), 315–424. https://doi.org/10.1152/physrev.00029.2006

Pope, C. A. (2007). Mortality Effects of Longer Term Exposures to Fine Particulate Air Pollution: Review of Recent Epidemiological Evidence. *Inhalation Toxicology*, *19*(sup1), 33–38. https://doi.org/10.1080/08958370701492961

Pope, C. A., Thun, M. J., Namboodiri, M. M., Dockery, D. W., Evans, J. S., Speizer, F. E., & Heath, C. W. (1995). Particulate Air Pollution as a Predictor of Mortality in a Prospective Study of U.S. Adults. *American Journal of Respiratory and Critical Care Medicine*, *151*(3_pt_1), 669–674. https://doi.org/10.1164/ajrccm/151.3_Pt_1.669

Qian, Z., Chapman, R. S., Hu, W., Wei, F., Korn, L. R., & Zhang, J. (Jim). (2004). Using air pollution based community clusters to explore air pollution health effects in children. *Environment International*, *30*(5), 611–620. https://doi.org/10.1016/j.envint.2003.11.003

Rao, S., Hughes, A., Segar, M. W., Wilson, B., Ayers, C., Das, S., Halm, E. A., & Pandey, A. (2021). Longitudinal Trajectories and Factors Associated With US County-Level Cardiovascular Mortality, 1980 to 2014. *JAMA Network Open*, *4*(11), e2136022–e2136022. https://doi.org/10.1001/jamanetworkopen.2021.36022

Rao, X., Zhong, J., Brook, R. D., & Rajagopalan, S. (2018). Effect of Particulate Matter Air Pollution on Cardiovascular Oxidative Stress Pathways. *Antioxidants & Redox Signaling*, *28*(9), 797–818. https://doi.org/10.1089/ars.2017.7394

Rice, M. B., Ljungman, P. L., Wilker, E. H., Gold, D. R., Schwartz, J. D., Koutrakis, P., Washko, G. R., O'Connor, G. T., & Mittleman, M. A. (2013). Short-Term Exposure to Air Pollution and Lung Function in the Framingham Heart Study. *American Journal of Respiratory and Critical Care Medicine*, *188*(11), 1351–1357. https://doi.org/10.1164/rccm.201308-1414OC

Sedgwick, P. (2014). Ecological studies: Advantages and disadvantages. *BMJ*, *348*. https://doi.org/10.1136/bmj.g2979

Soni, V., Singh, P., Shree, V., & Goel, V. (2018). Effects of VOCs on Human Health. In N. Sharma, A. K. Agarwal, P. Eastwood, T. Gupta, & A. P. Singh (Eds.), *Air Pollution and Control* (pp. 119–

142). Springer Singapore. https://doi.org/10.1007/978-981-10-7185-0_8

Sun, J., Barnes, A. J., He, D., Wang, M., & Wang, J. (2017). Systematic Review and Meta-Analysis of the Association between Ambient Nitrogen Dioxide and Respiratory Disease in China. *International Journal of Environmental Research and Public Health*, *14*(6), 646. https://doi.org/10.3390/ijerph14060646

Thurston, G. D. (2017). Outdoor Air Pollution: Sources, Atmospheric Transport, and Human Health Effects. In *International Encyclopedia of Public Health* (pp. 367–377). Elsevier. https://doi.org/10.1016/B978-0-12-803678-5.00320-9

Townsend, C. L. (2002). Effects on health of prolonged exposure to low concentrations of carbon monoxide. *Occupational and Environmental Medicine*, *59*(10), 708–711. https://doi.org/10.1136/oem.59.10.708

Turner, M. C., Krewski, D., Pope, C. A., Chen, Y., Gapstur, S. M., & Thun, M. J. (2011). Long-term Ambient Fine Particulate Matter Air Pollution and Lung Cancer in a Large Cohort of Never-Smokers. *American Journal of Respiratory and Critical Care Medicine*, *184*(12), 1374–1381. https://doi.org/10.1164/rccm.201106-1011OC

U.S. Census Bureau. (2014). *Gazetteer Files*. https://www.census.gov/geographies/reference-files/time-series/geo/gazetteer-files.2014.html#list-tab-264479560

U.S. Census Bureau. (2021). *Explore Census Data*. https://data.census.gov/cedsci

U.S. Department of Agriculture. (2021). *Data*. https://www.usda.gov/topics/data

U.S. Environmental Protection Agency. (2021). *Pre-Generated Data Files: Daily Summary Data*. https://aqs.epa.gov/aqsweb/airdata/download_files.html#Annual

U.S. Environmental Protection Agency. (2022). *Pollutant Standards*. https://aqs.epa.gov/aqsweb/documents/codetables/pollutant_standards.html

Valavanidis, A., Vlachogianni, T., Fiotakis, K., & Loridas, S. (2013). Pulmonary Oxidative Stress, Inflammation and Cancer: Respirable Particulate Matter, Fibrous Dusts and Ozone as Major Causes of Lung Carcinogenesis through Reactive Oxygen Species Mechanisms. *International Journal of Environmental Research and Public Health*, *10*(9), 3886–3907. https://doi.org/10.3390/ijerph10093886

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

Vermunt, J. K., & Magidson, J. (2005). *Technical Guide for Latent GOLD Choice 4.0: Basic and Advanced.* Belmont Massachusetts: Statistical Innovations Inc. https://www.statisticalinnovations.com/wp-content/uploads/LGCtechnical.pdf

Vierboom, Y. C., Preston, S. H., & Hendi, A. S. (2019). Rising geographic inequality in mortality in the United States. *SSM - Population Health*, *9*, 100478. https://doi.org/10.1016/j.ssmph.2019.100478

Wang, S., Gao, S., Li, S., & Feng, K. (2020). Strategizing the relation between urbanization and air pollution: Empirical evidence from global countries. *Journal of Cleaner Production*, *243*, 118615. https://doi.org/10.1016/j.jclepro.2019.118615

Wang, Y., Bechle, M. J., Kim, S.-Y., Adams, P. J., Pandis, S. N., Pope, C. A., Robinson, A. L., Sheppard, L., Szpiro, A. A., & Marshall, J. D. (2020). Spatial decomposition analysis of NO2 and PM2.5 air pollution in the United States. *Atmospheric Environment*, *241*, 117470.

https://doi.org/10.1016/j.atmosenv.2020.117470

Wiegman, C. H., Li, F., Ryffel, B., Togbe, D., & Chung, K. F. (2020). Oxidative Stress in Ozone-Induced Chronic Lung Inflammation and Emphysema: A Facet of Chronic Obstructive Pulmonary Disease. *Frontiers in Immunology*, *11*. https://doi.org/10.3389/fimmu.2020.01957

Yun, H. Y., Dawson, V. L., & Dawson, T. M. (1997). Nitric oxide in health and disease of the nervous system. *Molecular Psychiatry*, *2*(4), 300–310. https://doi.org/10.1038/sj.mp.4000272

Zhang, H., Wang, Q., He, S., Wu, K., Ren, M., Dong, H., Di, J., Yu, Z., & Huang, C. (2020). Ambient air pollution and gestational diabetes mellitus: A review of evidence from biological mechanisms to population epidemiology. *Science of The Total Environment*, *719*, 137349. https://doi.org/10.1016/j.scitotenv.2020.137349

Table 1. Descriptive statistics of air pollutant concentrations, health measurements, and socio-demographics at the county level of USA

Characteristics, *n*=699	**Median (interquartile range)**	**Mean (standard deviation)**	**Number (%)**
Air pollutant concentrations			
$PM_{2.5}$ Mass ($\mu g/m^3$)	9.39 (4.95)	9.03 (3.39)	-
PM_{10} Speciation ($\mu g/m^3$)	15.37 (10.11)	14.80 (7.41)	-
$NONO_xNO_y$ (ppb)	5.95 (7.71)	7.30 (6.84)	-
Health measurements			
Life expectancy, year			
Life expectancy (2014)	78.53 (3.01)	78.44 (2.23)	-
Increase in life expectancy between 1995 and 2014	3.33 (0.01)	3.41 (0.12)	-
Age-specific mortality risk, %			
0 – 4 years	0.63 (0.22)	0.65 (0.18)	-
5 – 25 years	0.83 (0.35)	0.87 (0.27)	-
25 – 44 years	2.77 (1.02)	2.88 (0.84)	-
45 – 64 years	12.05 (3.68)	12.37 (2.82)	-
65 – 84 years	50.77 (7.80)	50.52 (5.95)	-
Cause-specific mortality rate, number of deaths/100 000 Population			
Communicable, maternal, neonatal, and nutritional diseases			
HIV/AIDS and tuberculosis	1.12 (1.54)	1.82 (2.05)	-
Diarrhea, lower respiratory and other common infectious diseases	30.06 (12.10)	30.85 (9.20)	-
Neglected tropical diseases and malaria	0.05 (0.05)	0.06 (0.05)	-
Maternal disorders	0.30 (0.15)	0.33 (0.14)	-
Neonatal disorders	2.94 (1.35)	3.19 (1.17)	-
Nutritional deficiencies	1.35 (0.73)	1.44 (0.66)	-
Other communicable, maternal, neonatal, and nutritional diseases	1.29 (0.44)	1.34 (0.33)	-
Noncommunicable diseases			
Neoplasms	196.77 (37.59)	197.20 (29.60)	-
Cardiovascular diseases	252.63 (67.88)	256.96 (52.17)	-
Chronic respiratory diseases	58.15 (22.06)	59.15 (15.96)	-
Cirrhosis and other chronic liver diseases	17.71 (6.91)	18.87 (7.18)	-
Digestive diseases	15.73 (3.26)	15.63 (2.66)	-
Neurological disorders	99.01 (30.10)	97.94 (21.36)	-
Mental and substance use disorders	13.60 (7.54)	14.84 (6.71)	-
Diabetes, urogenital, blood and endocrine diseases	56.39 (19.32)	58.24 (15.87)	-
Musculoskeletal disorders	3.16 (1.04)	3.28 (0.86)	-
Other non-communicable diseases	6.22 (1.69)	6.34 (1.37)	-
Injuries			
Transport injuries	16.20 (10.11)	17.89 (8.39)	-

Unintentional injuries	22.29 (6.32)	22.52 (5.41)	-
Self-harm and interpersonal violence	21.44 (8.87)	22.49 (7.24)	-
Forces of nature, war and legal intervention	0.06 (0.06)	0.08 (0.07)	-
Socio-demographics			
Population characteristics			
Size (per 1,000 population), *n*	108.52 (282.24)	304.38 (624.33)	-
Gender, male %	49.35 (1.38)	49.63 (1.57)	-
Ethnicity, white alone %	87.71 (17.79)	82.78 (14.62)	-
Age, %			
0 – 9 years	12.35 (2.43)	12.48 (2.06)	-
10 – 19 years	12.86 (1.86)	12.89 (1.59)	-
20 – 29 years	12.74 (3.44)	13.30 (3.25)	-
30 – 39 years	12.13 (2.04)	12.23 (1.75)	-
40 – 49 years	12.33 (1.86)	12.33 (1.38)	-
50 – 59 years	14.31 (1.96)	14.25 (1.61)	-
60 – 69 years	11.36 (3.00)	11.77 (2.53)	-
70 – 79 years	6.32 (2.31)	6.71 (2.02)	-
≥ 80 years	3.93 (1.53)	4.05 (1.21)	-
Socioeconomics			
Educational level (age ≥ 25), %			
Less than a high school diploma	10.60 (6.00)	11.62 (5.28)	-
A high school diploma only	29.70 (9.15)	29.80 (7.08)	-
Completing some college or associate's degree	31.40 (6.95)	31.27 (5.16)	-
A bachelor's degree or higher	25.40 (13.55)	27.31 (10.64)	-
Median household income (annual, per 1,000 US dollars), *n*	48.57 (15.09)	51.48 (13.53)	-
Unemployment rate, %	6.00 (2.40)	6.28 (2.21)	-
Poverty rate, %	15.50 (7.05)	15.89 (5.51)	-
Healthcare service			
Medical insured population (age < 65), %	86.65 (6.20)	86.63 (4.62)	-
Physicians (per 1,000 population), *n*	1.78 (1.78)	2.15 (1.87)	-
Residential environment and location			
Rural-Urban Continuum Code			
1 (Metro areas, 1 million population or more)	-	-	171 (24.46)
2 (Metro areas, 250 thousand to 1 million population)	-	-	163 (23.32)
3 (Metro areas, population fewer than 250 thousand)	-	-	111 (15.88)
4 (Urban population of 20 thousand or more, adjacent to a metro area)	-	-	50 (7.15)
5 (Urban population of 20 thousand or more, not adjacent to a metro area)	-	-	31 (4.43)
6 (Urban population of 2,500 to 19,999, adjacent to a metro area)	-	-	62 (8.87)
7 (Urban population of 2,500 to 19,999, not adjacent to a metro area)	-	-	62 (8.87)
8 (Completely rural or less than 2,500 urban population, adjacent to a metro	-	-	14 (2.00)

area)			
9 (Completely rural or less than 2,500 urban population, not adjacent to a metro area)	-	-	35 (5.58)
Latitude	38.18 (7.06)	38.82 (5.23)	-
Longitude	-94.92 (24.69)	-94.91 (15.96)	-

Table 2. Cluster classification: posterior probability of membership of clusters

Assigned cluster, *n* =699 (%)	Mean posterior probability for each cluster (standard deviation)				
	Cluster 1 (the 'all low' cluster)	Cluster 2 (the 'all medium' cluster)	Cluster 3 (the 'high particulates' cluster)	Cluster 4 (the 'all high' cluster)	Cluster 5 (the 'mixed profile' cluster)
Cluster 1 (the 'all low' cluster), n = 115 (16.5%)	**0.66 (0.12)**	0.32 (0.08)	0.02 (0.05)	0.00 (0.00)	0.00 (0.00)
Cluster 2 (the 'all medium' cluster), n = 285 (40.8%)	0.02 (0.03)	**0.88 (0.14)**	0.06 (0.10)	0.03 (0.08)	0.02 (0.05)
Cluster 3 (the 'high particulates' cluster), n = 152 (21.7%)	0.01 (0.01)	0.13 (0.13)	**0.84 (0.13)**	0.01 (0.01)	0.01 (0.01)
Cluster 4 (the 'all high' cluster), n = 136 (19.4%)	0.00 (0.00)	0.06 (0.06)	0.01 (0.00)	**0.93 (0.06)**	0.00 (0.00)
Cluster 5 (the 'mixed profile' cluster), n = 11 (1.6%)	0.00 (0.00)	0.01 (0.01)	0.08 (0.02)	0.00 (0.00)	**0.91 (0.01)**

Table 3. Associations of air pollutant clusters with life expectancy and change in life expectancy with confounding adjustment

Variables in multivariate regression analysis	Life expectancy, 2014 Adjusted R^2=0.67		Change in life expectancy, 1995-2014 Adjusted R^2=0.45	
	Coefficient	99.5% Confidence Intervals	Coefficient	99.5% Confidence Intervals
Clusters of air pollutants				
Cluster 1 (the 'all low' cluster)	0 (referent)	-	0 (referent)	-
Cluster 2 (the 'all medium' cluster)	**-0.36**	**(-0.80, -0.08)**	-0.25	(-0.56, 0.05)
Cluster 3 (the 'high particulates' cluster)	**-0.79**	**(-1.32, -0.27)**	**-0.40**	**(-0.77, -0.04)**
Cluster 4 (the 'all high' cluster)	**-0.95**	**(-1.50, -0.39)**	**-0.46**	**(-0.84, -0.07)**
Cluster 5 (the 'mixed profile' cluster)	-0.83	(-2.02, 0.36)	0.33	(-0.50, 1.15)
Population size (per 1,000 population), n	**0.0006**	**(0.0003, 0.0008)**	**0.0004**	**(0.0002, 0.0006)**
Gender, male %	0.11	(-0.01, 0.22)	**0.12**	**(0.05, 0.20)**
Ethnicity, white alone %	**0.04**	**(0.02, 0.05)**	**-0.02**	**(-0.03, -0.01)**
Education level, bachelor's degree %	**-0.12**	**(-0.15, -0.09)**	**-0.05**	**(-0.07, -0.03)**
Median household income (per 1,000 US dollars), n	**0.04**	**(0.02, 0.06)**	0.01	(-0.002, 0.029)
Unemployment rate, %	**0.08**	**(0.004, 0.16)**	**0.16**	**(0.10, 0.21)**
Poverty rate, %	**-0.10**	**(-0.15, -0.04)**	-0.03	(-0.07, 0.01)
Insured population, %	0.02	(-0.02, 0.06)	**-0.03**	**(-0.06, -0.0008)**
Physicians (per 1,000 population), n	0.06	(-0.04, 0.16)	**0.11**	**(0.04, 0.18)**
Rural-Urban Continuum Code				
1	0 (referent)	-	0 (referent)	-
2	0.39	(-0.06, 0.84)	**-0.39**	**(-0.70, -0.08)**
3	0.49	(-0.04, 1.03)	**-0.49**	**(-0.86, -0.11)**
4	0.62	(-0.05, 1.28)	-0.17	(-0.63, 0.30)
5	0.21	(-0.58, 1.01)	-0.51	(-1.07, 0.04)
6	**0.80**	**(0.13, 1.46)**	-0.24	(-0.71, 0.22)
7	0.36	(-0.31, 1.04)	-0.37	(-0.84, 0.10)
8	0.90	(-0.20, 2.01)	0.01	(-0.76, 0.78)
9	**0.98**	**(0.15, 1.82)**	-0.14	(-0.72, 0.44)
Latitude	0.01	(-0.02, 0.05)	**0.03**	**(0.002, 0.05)**
Longitude	0.003	(-0.01, 0.01)	**0.01**	**(0.01, 0.02)**