Group 21: 21 and Me

Linfeng Hu, Cynthia Ma, Pluto Zhang

The Association between Asthma-related Emergency Department Visits and Exposure to Air Pollution

Abstract

Exposure to ambient air pollutants such as ozone and particulate matter has been linked to higher risks of respiratory and cardiovascular diseases. This study aims to investigate the relationship between ambient air pollutant concentration and the age-adjusted rate of Asthma-related Emergency Department (ED) visits in the state of California, while adjusting for socioeconomic indicators and other environmental factors. Census-tract level cross-sectional data were analyzed using a negative binomial model to address overdispersion in the dataset. We found a quadratic association between age adjusted asthma-related ED visits and daily maximum 8-hour ozone concentration, and a positive linear relationship between annual mean PM2.5 and ED visits.

Introduction

According to the World Health Organization (WHO), ambient air pollution is responsible for 4.2 million premature deaths every year[1]. Driven by rapid urbanization, globalization of industrial production, and growing use of automobile vehicles, ambient air pollution is currently one of the most crucial threats for global health. Within the United States, over 137 million Americans are still living in places with unhealthy levels of air quality despite the introduction of the Clean Air Act in 1970[2]. Components of air pollution include particulate matter (PM), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), volatile organic compounds (VOC), and ozone (O_3). Both acute and chronic exposure to these air pollutants have been linked with several adverse health outcomes including cardiovascular events, respiratory disease (ex. Chronic Obstructive Pulmonary Disease (COPD), asthma, bronchiolitis), lung cancer, low birthweight, neurologic disease, and many more[3].

Literature Review

Previous studies have assessed the relationship between air pollution and emergency department visits yet findings are mixed and inconsistent most likely due to different study designs. A multi-city case-crossover study in Canada found that a short-term increase in air pollution increased emergency department (ED) visits related to upper and low respiratory illness even in areas with relatively low air pollutant concentration[4]. However, another ecological study in New York did not identify a change in ED visits for asthma associated with air pollution, which the authors thought might be masked by other influential triggers[6]. A cohort study in Atlanta, Georgia identified CO as the strongest predictor for cardiovascular-related ED visits and ozone as the strongest predictor for respiratory-related ED visits[5]. While most studies utilized hospital records data to identify ED visits, we were not able to gain access to dataset at a similar level of resolution that could achieve satisfactory statistical power. Therefore, we decided to use census-tract level data to investigate the relationship between the age-adjusted rate of ED visits for asthma based on ambient air pollution level while adjusting for socioeconomic indicators in the state of California, where data were available.

Methods

Data Source

Census tract-level data were extracted from CalEnvironScreen 3.0 compiled by the California Environmental Protection Agency. The outcome of interest is ED visits related to asthma and exposure variables are ozone, PM2.5 and Diesel PM. Variable definitions can be found in Appendix

A. Covariates with missing data include PM2.5, DrinkingWater, Traffic, LowBirthWeight, Education, LinguisticIsolation, Poverty, Unemployment, and HousingBurden. The proportion of observations with missing data = 4.966%, which is less than the empirical rule of 5%. Most census tracts with missing data have and only have missingness in LowBirthWeight. We believe that data were missing at random (MAR) because missingness was mostly identified in socioeconomic indicators that should not be dependent on our outcome of interest, and the missingness can be fully accounted for by variables with thorough information. The MAR assumption is restrictive and impossible to assess, and may be dealt with using many different statistical approaches. For now, we proceed with complete case analysis because we do not believe the complete cases differ systematically from the incomplete cases in this data set.

Preliminary Variable Selection

Based on the results of the stepwise selection procedure, 18 out of 55 available variables in the dataset were selected as possible covariates in the final model. Paired scatter plots in Appendix B were then visually examined to evaluate collinearity among the variables as well as each variable's crude association with age-adjusted rate of asthma-related ED visits. Pollution burden index and age-adjusted cardiovascular-related ED visits were excluded as potential model covariates due to high correlation with other environmental exposure indices and percentage of low birth weight respectively.

Model Selection

The Poisson regression model was initially considered since the outcome variable of asthma-related ED visits is a count variable. However, because the variable of interest exhibited overdispersion (dispersion quotient = 5), the negative binomial model was chosen instead for a more robust variance estimation. The number of covariates were further shrunk based on the p-values and variable importance (VIP) scores from the regression results from the negative binomial model with 16 potential covariates. Possible higher order terms and effect measure modification by pollutant concentrations were also evaluated to improve the model fit. The final model was determined based on AIC values and goodness-of-fit tests.

Model Specification

Variable Selection

We first ran a Poisson model that included all the linear terms of the available variables selected from the stepwise selection as covariates. Since Pesticides, Traffic, Haz.waste, Imp.WaterBodies, and HousingBurden have statistically insignificant p-values, they were excluded from the next round of analysis. We then ran an elastic net and variable importance in projection analysis to further exclude the covariates that demonstrated multicollinearity or were not as influential to asthma rate. Elastic net modeling was used rather than LASSO or ridge regression due to its relative flexibility that allows for preservation of covariates information while penalizing for multicollinearity.

Based on the VIP scores, the daily maximum 8-hour ozone concentration was the most influential on asthma rate and was thus considered as the primary covariate in the modeling process. DrinkingWater, CleanupSites and GroundwaterThreats were subsequently excluded due to low importance on asthma as shown in Appendix D. The final set of covariates were PM2.5, Ozone, DieselPM, Unemployment, Poverty, LinguisticIsolation, and LowBirthWeight.

Model Development

A multiple Poisson model with all seven variables was examined first. While all covariates had statistically significant p-values, the dispersion quotient was 10.8. To accommodate this strong overdispersion effect, the negative binomial model was then adopted, which had a substantially lower AIC value (AIC = 66901.9) compared to the original Poisson model (AIC = 116849.8). Higher order terms and interaction terms were then considered to further improve the model fit.

All possible interactions between ozone and each of the other six covariates were examined. Except for diesel PM, the rest of the covariates demonstrated effect measure modification on the multiplicative scale for the association between ozone and asthma. We decided to focus on the interaction term between ozone and PM2.5 due to its low p-value and known physiological mechanisms. In addition, the AIC value of the negative binomial model with Ozone * PM2.5 interaction term is the lowest among the 6 models, at the value of 66723.9 (Table 1).

We also attempted to incorporate a higher order term of ozone into the model as it had a significant and strong influence on asthma. The model with the additional quadratic term of ozone was compared with the multiple linear model using the chi-sq test. The null hypothesis was that the negative binomial model with the quadratic term was better in explaining asthma rate as compared to the negative binomial model with only linear forms of the covariates. While the alternative hypothesis was that the model with the quadratic term was as good as the model with only linear covariates. The resulting chi-sq test statistics had a p-value less than 0.001, which rejected the null hypothesis and allowed us to infer that the quadratic term of ozone should be included in the final model. The choice was further confirmed by the shrunken overdispersion quotient in the Poisson model and lower AIC value in the negative binomial model. We therefore conclude that the model with the quadratic germ of ozone improved the goodness of fit.

Table 1. Comparison of AIC values and dispersion quotient of Poisson and negative binomial models

Model covariates	**Model form**	**AIC of NB version**	**Dispersion of Poisson version**	**AIC of Poisson version**
Ozone + PM2.5 + DieselPM + LinguisticIsolation + Poverty + Unemployment + LowBirthWeight	All linear terms	66901.919	9.821	116849.8
Ozone + PM2.5 + DieselPM + LinguisticIsolation + Poverty + Unemployment + LowBirthWeight + Ozone * PM2.5	All linear terms + Interaction term between Ozone and PM2.5	66723.971	9.621	115331.7
Ozone + PM2.5 + DieselPM + LinguisticIsolation + Poverty + Unemployment + LowBirthWeight + Ozone * DieselPM	All linear terms + Interaction term between Ozone and Diesel PM	66902.876	9.821	116846.6
Ozone + PM2.5 + DieselPM + LinguisticIsolation + Poverty + Unemployment + LowBirthWeight + Ozone * LinguisticIsolation	All linear terms + Interaction term between Ozone and Linguistic Isolation	66881.814	9.797	116658.3
Ozone + PM2.5 + DieselPM + LinguisticIsolation + Poverty + Unemployment + LowBirthWeight + Ozone * Poverty	All linear terms + interaction term between Ozone and Poverty	66889.174	9.814	116792.0
Ozone + PM2.5 + DieselPM + LinguisticIsolation + Poverty + Unemployment + LowBirthWeight + Ozone * Unemployment	All linear terms + interaction term between Ozone and Unemployment	66837.192	9.754	116339.5
Ozone + PM2.5 + DieselPM + LinguisticIsolation + Poverty + Unemployment + LowBirthWeight + Ozone * LowBirthWeight	All linear terms + interaction term between Ozone and Low Birth Weight	66879.599	9.801	116694.6
Ozone + PM2.5 + DieselPM + LinguisticIsolation + Poverty + Unemployment + LowBirthWeight + I(Ozone^2)	All linear terms + Quadratic Ozone	66710.836	9.555	114836.5
Ozone + PM2.5 + DieselPM + LinguisticIsolation + Poverty + Unemployment + LowBirthWeight+ I(Ozone^2) + PM2.5*Ozone	Add the quadratic term of Ozone and interaction term of Ozone and PM 2.5	66665.100	9.520	114563.5

Lastly, we compared the negative binomial model that included both the quadratic term of ozone and the interaction term of ozone and PM2.5 with the model that included only the quadratic term. The p-value of the F-statistics was less than 0.001 while the AIC value was also lower for the former. This motivated our decision to include both the quadratic term and interaction term in addition to the linear model. Thus, our final model is:

$$log(\text{asthma}) = \beta_0 + \beta_1 \cdot \text{Ozone} + \beta_2 \cdot \text{PM2.5} + \beta_3 \cdot \text{DieselPM} + \beta_4 \cdot \text{LinguisticIsolation} + \beta_5 \cdot \text{Poverty} + \beta_6 \cdot \text{Unemployment} + \beta_7 \cdot \text{LowBirthWeight} + \beta_8 \cdot \text{Ozone}^2 + \beta_9 \cdot \text{Ozone} * \text{PM2.5} + \varepsilon$$

Assumptions

We presumed that there was no violation of the assumptions of linearity, independence and stationarity. Despite potential correlation in air pollutant concentrations due to spatial proximity among the census tracts, we believe that asthma-related ED visits were only driven by the characteristics of the corresponding census tracts and independent from the other census tracts. We validated the linearity assumption by comparing the residuals and predicted values of the outcome variable using the negative binomial model with all covariates included in our final model. The residuals did not demonstrate a specific pattern associated with the predicted asthma rate as desired (Appendix C). The residuals also did not demonstrate a specific pattern associated with the log of the predicted asthma rate (Appendix D), which indicates that the relationship between the outcome variable and the predictors is linear. Stationarity is also satisfied as the data is cross-sectional and the relationship between asthma-related ED visits and the covariates are consistent across census tracts.

Secondary Analysis

Sensitivity analysis was performed to determine if missing data had substantial impacts on the model output. Final model performances using complete case dataset versus imputed dataset were compared to examine the potential discrepancy in the association between asthma-related ED visits and air pollutant concentrations.

Discussion

Census Tract Characteristics

Table 2. Census tract characteristics (N = 8035)

	Mean (SD) or Median [Min, Max]
Environmental Exposure	
Ozone (ppm)	0.0474 (0.0102)
$PM_{2.5}$ ($\mu g/m^3$)	10.4 (2.60)
Diesel PM (kg/day)	19.2 (17.0)
Pesticides (lb/mi^2)	314 (2780)
Impaired Water Bodies (# of pollutants)	3.28 (4.52)
Drinking Water Index	479 [6.92, 1250]
Toxic Release Index	474 [0, 843000]
Traffic Density Index	700 [22.4, 45700]
Cleanup Sites Index	2.00 [0, 324]
Groundwater Threats Score	5.60 [0, 1610]
Hazardous Waste Score	0.0500 [0, 28.7]
Socioeconomics Indicators	
Linguistic Isolation (% of population)	10.4 (9.98)
Poverty (% of population)	36.4 (20.3)
Unemployment (% of population)	10.2 (5.14)
Housing Burden (% of households)	19.3 (8.73)
Health Indicators	
Low Birth Weight (%)	4.98 (1.55)
Cardiovascular Disease (aged-adjusted rate of related ED visits per 10,000)	8.27 (2.97)

Table 2 summarized the census tract characteristics on environmental exposure, socioeconomics indicators, and health indicators. All census tracts met the National Ambient Air Quality Standards (NAAQS) of 0.07 ppm for daily maximum 8-hour ozone concentration. However, 30% of the census tracts were still above the NAAQS requirements of 12.0 $\mu g/m^3$ for annual mean $PM_{2.5}$ concentration. Methodology and interpretations for the other weight-adjusted indicators on environmental exposure and socioeconomic indices can be found at *CalEnvironScreen 3.0 Report* online[7].

Modeling Results

Table 3 provides the summary of our final model output. The dispersion parameter was around 1.02, which is smaller than the empirical standard of 1.10. Therefore, overdispersion was considered to be corrected. All coefficient estimates were statistically significant as a result of our model selection process.

Table 3. Summary of estimated negative binomial model on the association between age-adjusted asthma-related ED visits and ozone and PM2.5 concentration while adjusting for socioeconomics indicators at the census tract level

Variable	**Estimate**	**SD**	**z-value**	**p-value**
Ozone	-57.3	4.38	-13.07	<0.001
PM2.5	-0.0801	0.012	-6.92	<0.001
DieselPM	0.00328	0.00035	9.43	<0.001
LinguisticIsolation	-0.00961	0.00069	-14.02	<0.001
Poverty	0.0128	0.00040	31.81	<0.001
Unemployment	0.0208	0.0013	15.75	<0.001
LowBirthWeight	0.0649	0.0034	18.86	<0.001
I(Ozone^2)	410.03	52.27	7.84	<0.001
Ozone:PM2.5	1.44	0.206	6.70	<0.001
(Intercept)	4.80	0.108	44.34	<0.001

The directions of association for the selected confounding variables in the model were consistent with existing epidemiological studies. The net effect of Linguistic Isolation, Poverty, and Unemployment—all of which are indicators of low socioeconomic status (SES)—was positively associated with the number of asthma-related ED visits. This is consistent with epidemiological studies that had identified neighborhood-level SES as a confounder and modifier of air pollution-asthma association, especially among children[9, 10]. The prevalence of low birth weight was also associated with a higher rate of asthma-related ED visits. This is consistent with the research findings that maternal exposure to air pollution increases the risk of low birth weight while asthmatic mothers have even more elevated risks[11, 12].

While the positive linear association between diesel PM and asthma as expected due to ample physiological and epidemiological evidence[13], the relationship between ozone and $PM_{2.5}$ with asthma was more complicated to interpret. As the results from the ANOVA test (described in the Methods section) indicated, there is a statistically significant and meaningful quadratic relationship between ozone exposure and asthma-related ED visits. The negative coefficient for the linear $PM_{2.5}$ term seems to suggest a protective effect, yet it is very likely to be masked by the interaction between the air pollutants. The interaction term of $PM_{2.5}$ and ozone was statistically significantly and positively associated with asthma-related ED visits, which reflect the synergistic effects of $PM_{2.5}$ and ozone level on respiratory health that had been investigated and confirmed in many other studies[14].

Due to the presence of effect modification and quadratic relationship, we were not able to produce a specific dose-response slope. Therefore, we isolated the relationship between ozone, PM2.5 and the predicted asthma-related ED visits by adjusting for all other covariates at their respective mean values (Figure 1). The association was dominated and driven by the quadratic relationship between ozone and asthma, which suggests that ozone should be prioritized for achieving desired health outcomes.

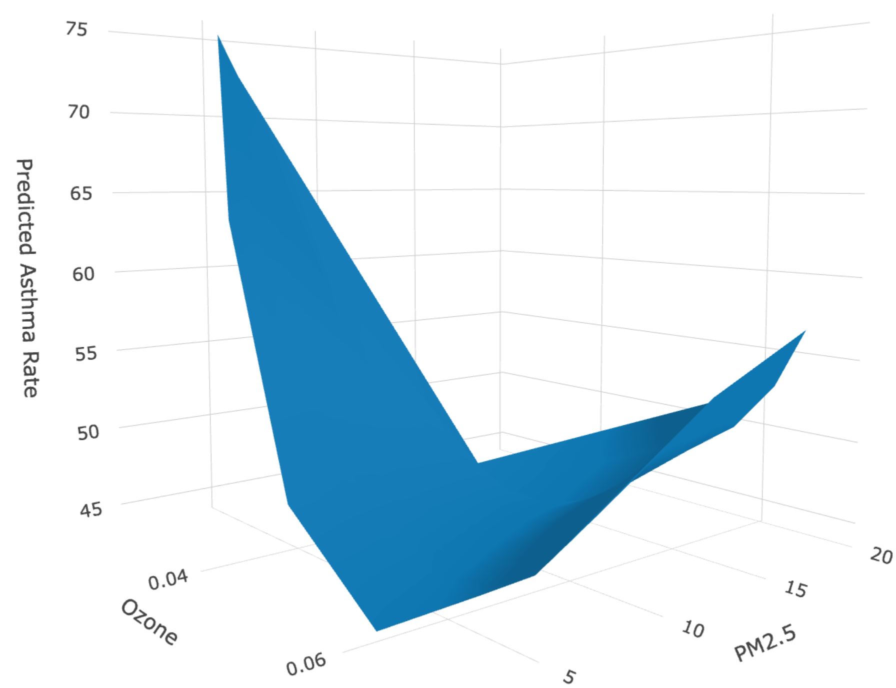


Figure 1. Three-dimensional surface plot of daily maximum 8-hour ozone concentration, annual mean $PM_{2.5}$ concentration, and the predicted age-adjusted asthma-related ED visits. DieselPM, LinguisticIsolation, Poverty, Unemployment, and LowBirthWeight were taken at their respective mean values. The interactive graph can be accessed at https://imaginative-lebkuchen-84fb15.netlify.app/

Sensitivity Analysis

Previously, we used a naive approach, complete case analysis, where we fully remove all observations with missing data since the amount of not available data (NAs) is less than 5%. However, complete case analysis in this case might be biased here in theory, since we have missing at random instead of missing completely at random. Therefore, we also attempt a less naive approach, multivariate imputation by chained equations using fully conditional specification, to fill in for existing NAs and see how that changed our model performance.

We used the mice package and set the number of imputation to be 25, meaning that our analysis now has fitted 25 regression models based on the observed data, conditional on all the covariates with complete information; then stochastically imputed for and generated 25 data sets with complete values for every covariate. We chose this number because there is slightly less than 5% of missingness so the 25 imputations should be a sufficient amount of synthetic datasets that are generated to replace for each missing value. We also set the number of iterations for each imputation to be 10. This refers to the maximum number of draws of the imputed value to replace the missing data. Thus, our final guess of the missing data would the calculated average from each of these 25 synthetic datasets. The key we always kept in mind when performing multiple imputation is that the imputed data is only as good as the model that we assumed for the distribution, based on the observed data. This is because we assume the distribution of missing variables is the same among those with and without complete information when we perform the imputation.

From the summaries of the two models obtained from complete case analysis and multiple imputation, we observed that the AIC score for imputed data's model slightly increased, indicating that the model using complete case analysis is a better fit. The model from the imputed dataset also shows a slightly higher residual deviance / square root of unit deviances, indicating higher sum of squares and therefore higher variability and worse fit of the data. However, since we have a larger size for the imputed dataset, it gives the model higher degrees of freedom, which can be translated to higher statistical power to reject a null hypothesis. The small magnitude of differences between two models using two datasets are understandable considering that although there is missing data, the magnitude is acceptable to be ignored, and there is no sufficient evidence of systematic missingness with respect to asthma-related ED visits.

Limitations

This study has underlying limitations. Firstly, since all of our data are from only one year (2018), and the same region, there is the problem of generalizability of our study results to the broader population. Unfortunately, we are unable to obtain a more diverse dataset on the census tract level with the similar indicator covariates that we are interested in. We are also aware the ED visit may not be a fully representative indicator of the number of people showing symptoms of asthma. The only other alternative we have found was asthma-related death. We chose to use ED visit because it may be a more comprehensive census-tract level indicator for asthma prevalence and has the potential to demonstrate a larger proportion of the disease burden compared using data on asthma-related death only.

Additionally, the dataset is entirely from the state of California while each observation is on a census-tract level. Under the same state government and federal regulations, there may present potential violations to the independence assumption for modeling the association. We assume independence in our dataset mainly because in the U.S, census tracts are "designed to be relatively homogeneous units with respect to population characteristics, economic status, and living conditions" and "average about 4,000 inhabitants"[8]. But the overall sampling method in examining asthma-related ED visits may be improved to have higher degrees of independence in order to examine the potential underlying associations.

Future Scope

In general, we have examined how environmental factors and socioeconomic indicators may be associated with the number of asthma-related ED visits on a census tract level in California with the data collected in 2018. However, as we have mentioned in the limitations section, there are many aspects that may be improved to further examine the potentially impacting factors. In the future, we hope to examine causal relationships between these covariates and the outcome of interest incorporating longitudinal data using mixed effect models. By doing so, we want to gain an increased statistical power and identify the trends over time. Prediction models may also be an area of interest to benefit the communities that are within the range of our analysis and guide potential improvements in regions where this type of chronic lung disease is more severe. We are also interested in exploring more approaches to correct the overdispersion problem that exist in our original dataset. Through adjusting the unexpected greater variability, asthma-related conditions may be better characterized, understood and therefore managed.

References

[1] *Ambient and household air pollution and Health: Frequently asked questions*. PAHO/WHO | Pan American Health Organization. (n.d.). Retrieved November 5, 2022, from https://www.paho.org/en/topics/air-quality-and-health/ambient-and-household-air-pollution-and-health-frequently-asked#:~:text=Exposure%20to%20high%20levels%20of,been%20associated%20with%20health%20impacts

[2] *Key findings: State of the Air*. State of the Air | American Lung Association. (n.d.). Retrieved November 5, 2022, from https://www.lung.org/research/sota/key-findings

[3] Bazyar, J., Pourvakhshoori, N., Khankeh, H. *et al.* A comprehensive evaluation of the association between ambient air pollution and adverse health outcomes of major organ systems: a systematic review with a worldwide approach. *Environ Sci Pollut Res* 26, 12648–12661 (2019). https://doi.org/10.1007/s11356-019-04874-z

[4] Szyszkowicz, M., Kousha, T., Castner, J., & Dales, R. (2018). Air pollution and emergency department visits for respiratory diseases: A multi-city case crossover study. *Environmental Research*, *163*, 263–269. https://doi.org/10.1016/j.envres.2018.01.043

[5] Tolbert, P., Klein, M., Peel, J. *et al.* Multipollutant modeling issues in a study of ambient air quality and emergency department visits in Atlanta. *J Expo Sci Environ Epidemiol* 17 (Suppl 2), S29–S35 (2007). https://doi.org/10.1038/sj.jes.7500625

[6] Castner, J., Guo, L. & Yin, Y. Ambient air pollution and emergency department visits for asthma in Erie County, New York 2007–2012. *Int Arch Occup Environ Health* 91, 205–214 (2018). https://doi.org/10.1007/s00420-017-1270-7

[7] Faust, J., August, L., Bangia, K., Galaviz, V., Leichty, J., Prasad, S., Schmitz, R., Slocombe, A., Welling, R., Wieland, W., & Zeise, L. (2017, January). *CalEnvironScreen 3.0*. California Office of Environmental Health Hazard Assessment. Retrieved November 5, 2022, from https://oehha.ca.gov/media/downloads/calenviroscreen/report/ces3report.pdf

[8] Archived on 2017-05-13. Retrieved 2017-10-19. https://web.archive.org/web/20170513191843/https://factfinder.census.gov/help/en/census_tract.htm

[9] O'Lenick, C. R., Winquist, A., Mulholland, J. A., Friberg, M. D., Chang, H. H., Kramer, M. R., Darrow, L. A., & Sarnat, S. E. (2016). Assessment of neighbourhood-level socioeconomic status as a modifier of air pollution–asthma associations among children in Atlanta. *Journal of Epidemiology and Community Health*, 71(2), 129–136. https://doi.org/10.1136/jech-2015-206530

[10] Neidell, M. J. (2004). Air Pollution, health, and socio-economic status: The effect of outdoor air quality on childhood asthma. *Journal of Health Economics*, *23*(6), 1209–1236. https://doi.org/10.1016/j.jhealeco.2004.05.002

[11] Westergaard, N., Gehring, U., Slama, R., & Pedersen, M. (2017). Ambient air pollution and low birth weight - are some women more vulnerable than others? *Environment International*, 104, 146–154. https://doi.org/10.1016/j.envint.2017.03.026

[12] Bekkar, B., Pacheco, S., Basu, R., & DeNicola, N. (2020). Association of air pollution and heat exposure with preterm birth, low birth weight, and stillbirth in the US. *JAMA Network Open*, 3(6). https://doi.org/10.1001/jamanetworkopen.2020.8243

[13] Pandya, R. J., Solomon, G., Kinner, A., & Balmes, J. R. (2002). Diesel exhaust and asthma: Hypotheses and molecular mechanisms of action. *Environmental Health Perspectives*, 110(suppl 1), 103–112. https://doi.org/10.1289/ehp.02110s1103

[14] Rhee, J., Dominici, F., Zanobetti, A., Schwartz, J., Wang, Y., Di, Q., Balmes, J., & Christiani, D. C. (2019). Impact of long-term exposures to ambient PM2.5 and ozone on ards risk for older adults in the United States. *Chest*, 156(1), 71–79. https://doi.org/10.1016/j.chest.2019.03.017

Appendix A. Variable Definitions

Asthma - spatially modeled age-adjusted rate of emergency department (ED) visits for asthma per 10,000 averaged over 2011-2013, rounded to the nearest integer

Ozone - Mean of summer months (May-October) of the daily maximum 8-hour ozone concentration (ppm), averaged over three years (2012 to 2014)

PM2.5 - Annual mean concentration of PM2.5 (average of quarterly means, µg/m3), over three years (2012 to 2014)

DieselPM - Spatial distribution of gridded diesel PM emissions from on-road and non-road sources for a 2012 summer day in July (kg/day)

LinguisticIsolation - Percent limited English-speaking household

Poverty - Percent of the population living below 2 times the federal poverty level

Unemployment - Percent of the population over 16-year-old that's unemployed and eligible for the labor force. Excluding retirees, students, homemaker, institutionalized persons except prisoners, those not looking for work, and military personnel on active duty

LowBirthWeight - Percent low birth weight averaged over 2006-2012.

Appendix B. Paired Scatter Plots

Figure B1. Paired scatter plot between asthma and exposure variables of interests (PM2.5, ozone, diesel PM)

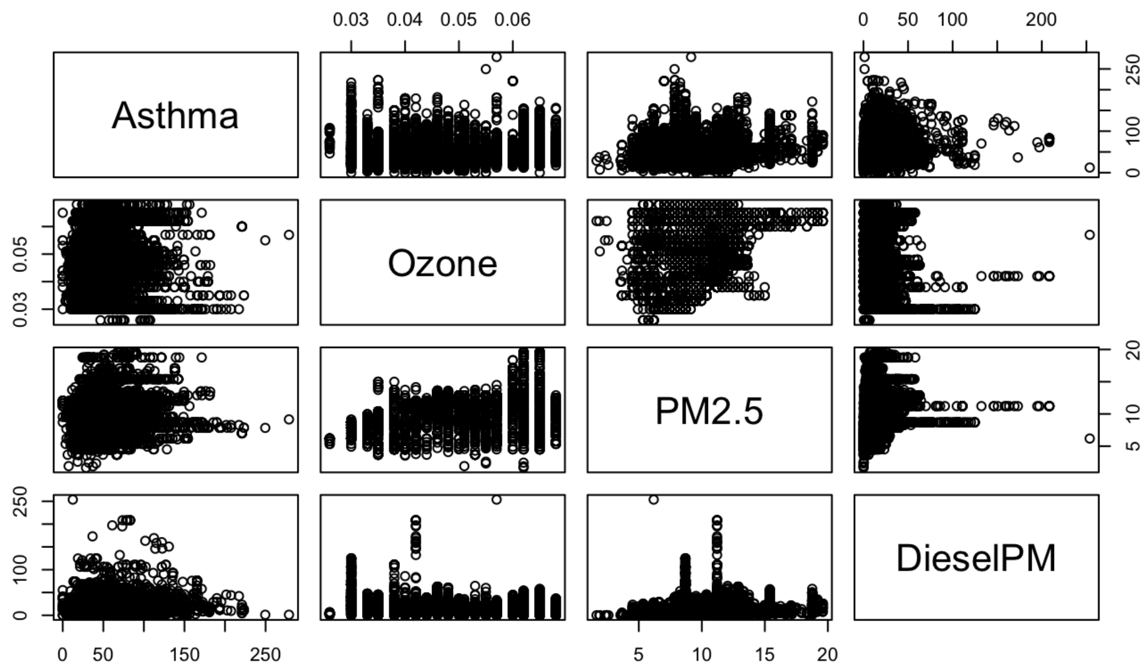


Figure B2. Paired scatter plot between asthma and possible environmental exposure covariates (drinking water, pesticides, toxic release, traffic density, cleanup sites, groundwater threats, hazardous waste, impaired water bodies, and pollution burden)

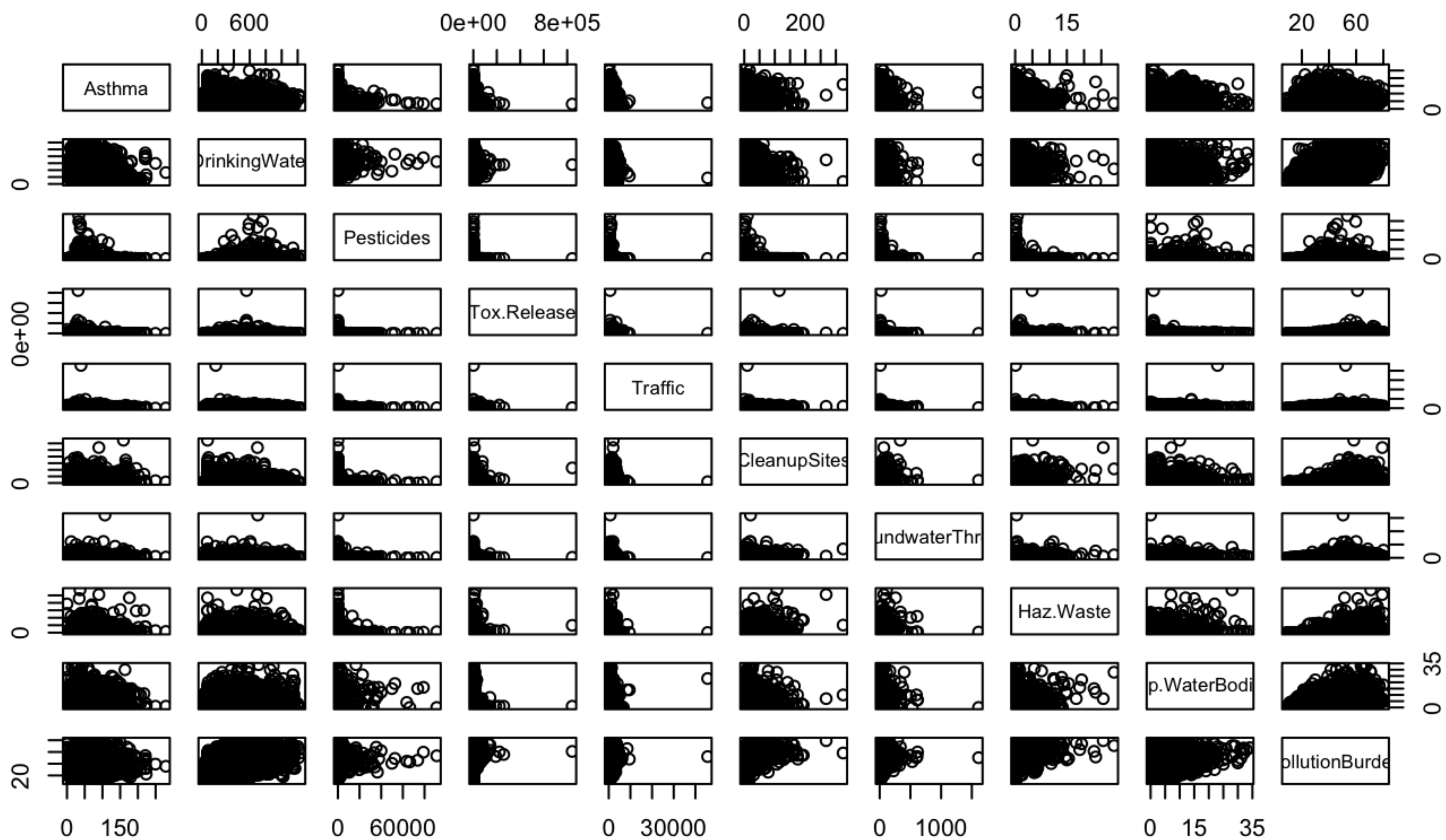
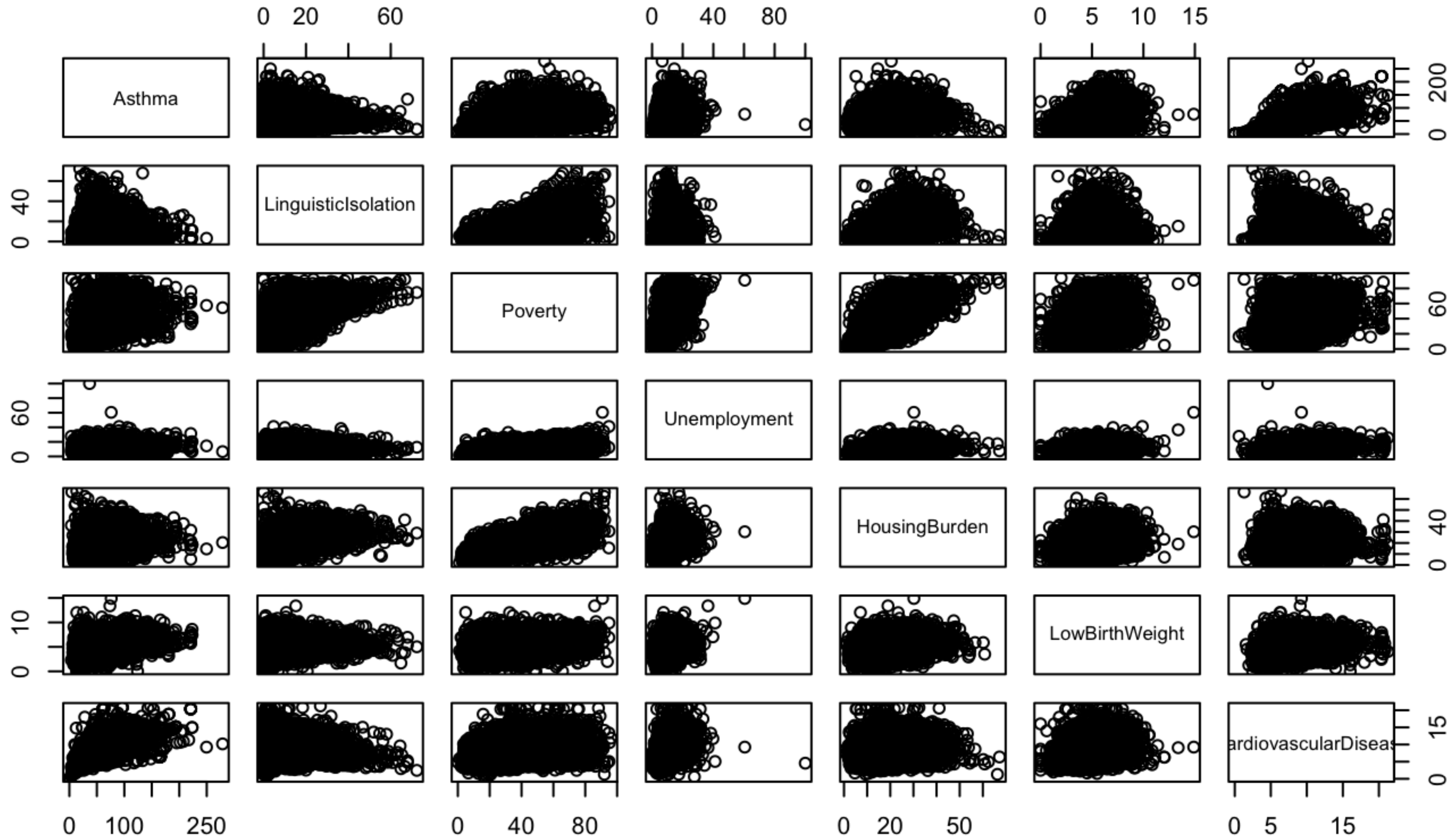
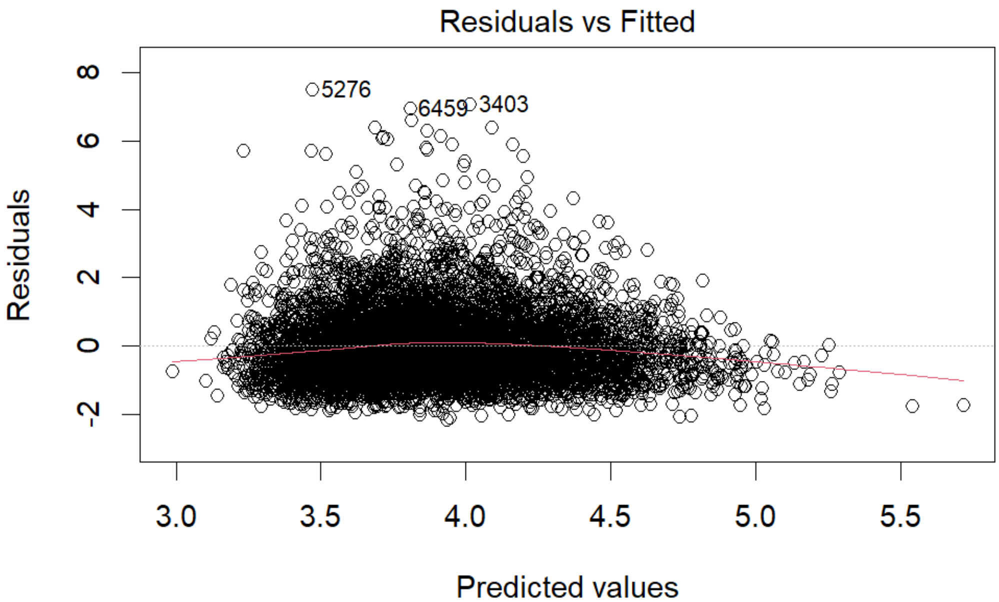


Figure B3. Paired scatter plot between asthma and possible socioeconomics and health indicators (linguistic isolation, poverty, unemployment, housing burden, low birth weight, and cardiovascular disease-related ED visits)



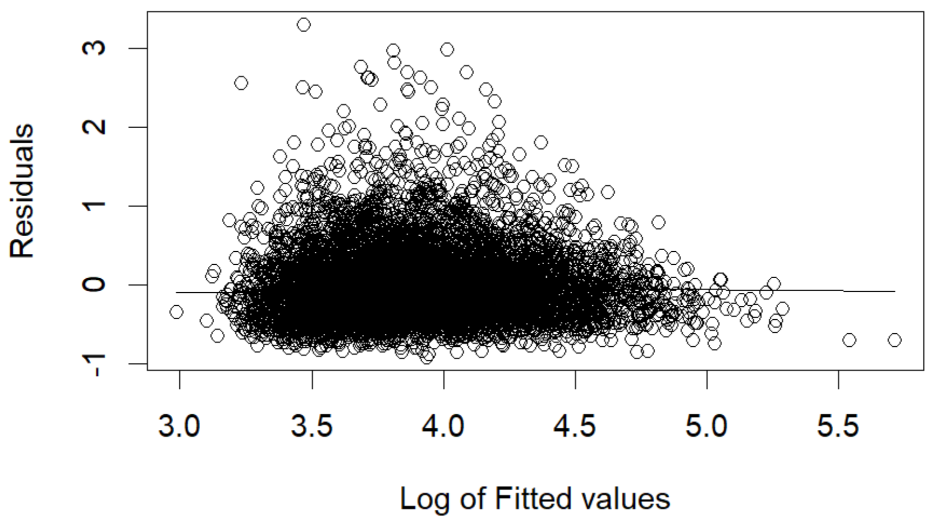
Appendix C. Residuals vs. Fitted Plot

The residuals vs predicted values plot for Asthma rate, using the final negative binomial model. As is shown on the graph, the residuals do not have a trend of increase or decrease as the predicted values increase, indicating that the linearity assumptions hold for the final model.



Appendix D. Residuals vs. Log Fitted Plot

The residuals vs log of predicted values plot for Asthma rate, using the final negative binomial model. As is shown on the graph, the residuals do not have a trend of increase or decrease as the log of predicted values increase, indicating that the linearity assumptions hold for the final model.



Appendix E. VIP Scores from Elastic Net Regularization

Figure E. the vip scores of the variables with respect to the log of Asthma-related visits. As can be seen on the graph, the Ozone has the highest power of association with log of age adjusted Asthma rate, while Groundwater Threats, Cleanup Sites and Drinking Water have the lowest importance score.