Impacts of Air Pollutants on Diseases

Yadu

May 23, 2016

Abstract

Air pollution is the leading cause of several diseases such as asthma, COPD, and cardiovascular disease. In this paper, the percentage of population with asthma (kids and adults), COPD, and cardiovascular disease will be correlated with the concentration of seven different air pollutants - carbon monoxide, ozone, particulate matter, sulfur dioxide, nitric oxide, nitrogen dioxide, and oxides of nitrogen in New York State counties. The air pollutant with the maximum concentration is obtained for each county. Then, each county is grouped according to maximum pollutant concentration. The atmosphere in each county is mostly polluted with ozone, carbon monoxide, or fine particulate matter. ANOVA hypothesis testing is then done to determine which air pollutant has a more statistical significance with the mentioned diseases. For COPD and cardiovascular disease, there is a negative correlation between the percentage of the population with the disease and the concentration of each air pollutant. For the percentage of adults with asthma there is a positive correlation with the concentration of each pollutant. According to the hypothesis test, there is not one pollutant that has a stronger impact on such diseases. The p-values for the percentage of kids with asthma, the percentage of adults with asthma, the percentage of the population with COPD, and the percentage of population with cardiovascular disease are all greater than the significance level of 0.05.

Through simulations, it has been anticipated that in the next 5 years, some negative correlation still exists when correlating the percentage of the population having COPD and cardiovascular disease with the concentration of the pollutants in the air. Also, when the anticipated percentage of adults with asthma is correlated with the concentration of pollutants in the air, the correlation seems to somewhat fluctuate for all 5 years. For the percentage of kids with asthma, the correlation generally appears to stagnate around 0 throughout the entire 5 years.

Keywords: Asthma; COPD; Cardiovascular Disease; Air Pollution; Kids; Adults; New York State; Carbon Monoxide; Ozone; Particulate Matter; Sulfur Dioxide; Nitric Oxide; Nitrogen Dioxide; Oxides of Nitrogen

Literature Review

Studies were conducted by the State Air Pollution Control Board only within New York City, linking asthma for children and adults with air pollution. The pollutants included in the analyses were sulfur dioxide, nitrogen dioxide, ozone, carbon monoxide, and fine particulate matter. The study was not conducted for any other respiratory disease.

Relative risks were computed for the inter-quartile range of same day 24-hour average pollutant concentration and asthma hospitalizations for children. For the cases of sulfur dioxide, nitrogen dioxide, and carbon monoxide, statistically significant associations between pollutants and asthma hospitalizations were much more common for children than for adults (Restrepo, Simonoff, Thurston, and Zimmerman 1113).

Studies were also conducted by the New York City Department of Health and Mental Hygiene, focusing on only 2 common air pollutants that contribute to lung and heart diseases, leading to death - ozone and fine particulate matter. These studies were also conducted only in New York City. Instead of examining diseases, health effects such as premature mortalities, hospitalizations, and emergency department visits for respiratory and cardiovascular diseases and conditions, were examined. The studies were conducted for different age groups.

Color maps of New York City were rendered regarding the mortality rates, respiratory hospitalization rates, cardiovascular hospitalization rates, and asthma emergency department visit rates among kids and adults attributed to both the concentration of ozone in the atmosphere and the concentration of particulate matter in the atmosphere.

The research methods used by the New York City Department of Health and Mental Hygiene were adapted from those utilized by the U.S. Environmental Protection Agency. These methods involved pollutant level characterization and a variety of calculations such as the difference between the current and the comparison levels within each neighborhood and the vicinity-wide health impact associated with the change in ambient air quality the change in air pollution concentrations. The parameters used here were the change in air pollution concentrations, concentration-response functions from the epidemiological literature, and local population and baseline health event rates. Finally, the neighborhood health impacts were combined to estimate the citywide impacts (Kheirbek, Wheeler, Walters, Pezeshki, and Kass 9).

There is a stronger link between adolescent asthma and air pollution than between adult asthma and air pollution (Restrepo, Simonoff, Thurston, and Zimmerman 1102). Weather data that was obtained from 3 different weather stations (located in Central Park, Kennedy airport, and LaGuardia airport) was also used in the study conducted by the State Air Pollution Control Board.

However, the opposite is true for ozone. Seasonal analysis of the link between asthma hospitalizations and particulate matter suggested that concentrations of this class of pollutants could be more important for asthma hospitalizations during the period between November and March, which are the colder months of the year (Restrepo, Simonoff, Thurston, and Zimmerman 1113).

In this research study conducted, there were 7 air pollutants that were examined - carbon monoxide, ozone, particulate matter, sulfur dioxide, nitric oxide, nitrogen dioxide, and oxides of nitrogen. Asthma was not the only disease that was examined. COPD and cardiovascular disease were also examined for this study. The study was conducted not only for New York City but for the entire state.

There were many drawbacks to this approach. For each county in New York State, data was not collected for all 7 air pollutants. For some of the air pollutants, the data was missing. There were many counties which had data for at most 6 of the pollutants.

Unlike the other two research studies that were conducted, a simulation was done to anticipate the results and correlations within the next 5 years. A drawback here was that ANOVA hypothesis testing could not be done on the data due to the anticipated result that all of the counties were mostly polluted with carbon monoxide.

Methodology

The data regarding the amount of the population with the diseases was scraped from the American Lung Association website. For each county the percentages of the population with COPD and cardiovascular disease were each calculated by dividing the number of people with the disease by the population of the county and then multiplying the fraction by 100. The percentages of the kids with asthma were calculated by dividing the number of kids with asthma by the number of kids living in the county and multiplying the fraction by 100. The percentages of the adults with asthma were calculated by dividing the number of adults with asthma by the number of adults living in the county and multiplying the fraction by 100.

Then, the data regarding the concentrations of different pollutants in each county of New York was obtained by downloading a csv file from the United States Environmental Protection Agency website. This dataset was then subsetted so that it would conain only data regarding the counties of New York State and the names and levels of the 7 pollutants in this study. Then, each of these levels were averaged separately according to county. Because not all of the units were the same for each quantity, the units were converted to parts per million for those quantities that had units of parts per billion. This was done by dividing each quantity by 1000.

Then, both datasets were merged by county name. A total of 28 correlation plots were rendered displaying the correlations between the percentage of the population with each disease and the concentration of each pollutant.

For each county, the name of the pollutant with the highest concentration was obtained. Then, the conditions were checked so that ANOVA hypothesis testing could be conducted relating the average percentage of each disease with the maximum average concentration of the pollutant. The box-and-whisker plots in **Figure 2** in the Appendix display constant variability across groups despite the fact that there are a few visible prominent outliers. There is also normal distribution across each group because the mean and median for each group are almost equal.

Then, ANOVA hypothesis testing was done 4 times for 3 groups of counties - 1 with the maximum concentration of carbon monoxide, 1 with the maximum concentration of ozone, and 1 with the maximum concentration of fine particulate matter.

To anticipate the possible results within the next 5 years, it was assumed that the distributions for the percentage of the population with the diseases and the distributions for the concentration of each pollutant in the air would remain the same as those of 2015. All of these distributions were normal given the fact that the mean and median were almost the same. The only distribution that was not normal was the distribution for the percentage of kids with asthma. This distribution was actually uniform due to the fact that all of the percentages were around 9 percent.

For each of the percentages and concentrations random numbers were generated. Then, charts displaying the correlations were rendered. The simulation was repeated 5 times, rendering 5 correlation charts (1 for each year). (For more information see **Figures 3 through 7** in the Appendix.)

For each of these simulations, hypothesis testing was not applicable. The reason why is because for each of the counties, the only pollutant with the maximum concentration is carbon monoxide.

Results

According to **Figure 1**, there is a negative correlation between the percentage of the population with COPD and the concentration of pollutants in the air. The same is true for the percentage of the population with cardiovascular disease.

There is a positive correlation between the percentage of adults with asthma and the concentration of each pollutant. It can be inferred that the air pollutants have a stronger impact on the health of adults with asthma as compared to the health of kids with asthma.

From the ANOVA hypothesis-testing results, it was concluded that not one of these 7 air pollutants has a stronger impact on respiratory diseases. Each air pollutant has the same impact on all of the respiratory diseases. The p-value for each of the 4 ANOVA tests were 0.79 (for asthma in kids), 0.62 (for asthma in adults), 0.56 (for COPD), and 0.54 (for cardiovascular disease).

Figures 3 through 7 show the correlations anticipated for the next 5 years. According to **Figure 3**, within the next year, there will generally be little to no correlation between the percentage of the population with respiratory diseases and the concentration of each pollutant in the air. Negative correlation still exists to some extent when correlating the percentage of the population having COPD and cardiovascular disease with the concentration of the pollutants in the air. It is shown in **Figure 4** that the correlations between the percentage of the population with cardiovascular disease and the concentrations of nitrogen dioxide, ozone, and sulfur dioxide are clearly negative.

Figure 7 shows that there is more negativity in the correlation between the percentage of the population with COPD and the concentration of carbon monoxide and in the correlations between the percentage of population with cardiovascular disease and the concentrations of ozone and particulate matter. In the same figure, there appears to be almost no correlation between the percentage of kids and adults with asthma and the concentrations of each pollutant.

Through **Figures 3 through 7** we can see that there is some fluctuation in the correlations between the percentage of adults with asthma and the concentrations of each air pollutant. In **Figure 4**, there appears to be some minute negativities in the correlations. In **Figure 5**, there appears to be one big negativity (when correlated with concentration of nitrogen dioxide) and one moderate negativity (when correlated with the concentration of the oxides of nitrogen) in the correlations. When correlated with all of the other concentrations (except carbon monoxide) there appears to be some moderate and minute positivities. In **Figure 6**, the correlations appear to be mostly positive. Correlations with the concentration of carbon monoxide and the concentration of ozone appear to be more positive than any of the other correlations. This is not surprising because ozone and carbon monoxide are 2 of the most harmful air pollutants. In **Figure 7**, the correlations appear to be mostly centered around 0 (like the correlations for the percentage of kids with asthma). There is only one moderate positivity which is the correlation between the percentage of adults with asthma and the concentration of ozone. There is also only one moderate negativity which is the correlation between the percentage of adults with asthma and the concentration of particulate matter.

Conclusion

Positive correlations were rendered when the percentage of adults with asthma was correlated with the concentration of different pollutants in the air. However, when the percentage of kids with asthma was correlated, only for 2 pollutants (ozone and sulfur dioxide), negative correlations were rendered. This implies that ozone and sulfur dioxide are air pollutants that affect adults with asthma, rather than affecting kids with asthma. However, when correlated with the concentrations of carbon monoxide, particulate matter, nitric oxide, nitrogen dioxide, and oxides of nitrogen, stronger correlations are rendered for the percentage of kids with asthma than for the percentage of adults with asthma. This suggests that asthma in kids can be strongly linked to traffic-related air pollution.

The predictions could have been more accurate if data was collected for all 7 major pollutants in each county. The group means calculated could have been significantly different.

In the future, it is possible to do correlations with the percentage of the population who smoke and also the number of vehicles used in the area. These two variables do affect the amount of pollution in the atmosphere. Therefore, the concentration of each pollutant in the atmosphere can be predicted by these two variables. In turn, the percentage of the population with respiratory diseases can also be predicted.

The percentage of kids with asthma was stagnant around 9 percent. For the other diseases, the percentages were normally distributed. We can examine why the percentages are all distributed as such.

We can also examine why the percentages of the population with COPD and cardiovascular disease would vary inversely with the concentration of air pollutants. This is a very surprising phenomenon.

Regarding the simulations, we really do not know how polluted the air will be in the years that follow. Things can change.

We can also examine how different air pollutants affect different people in different ways. Some people do not get affected much by the pollution in the air. People do have different lung capacities and immunities. These could be factors that contribute to diseases.

Appendix

Figure 1 - Percentage Diseased vs. Pollutant Concentration

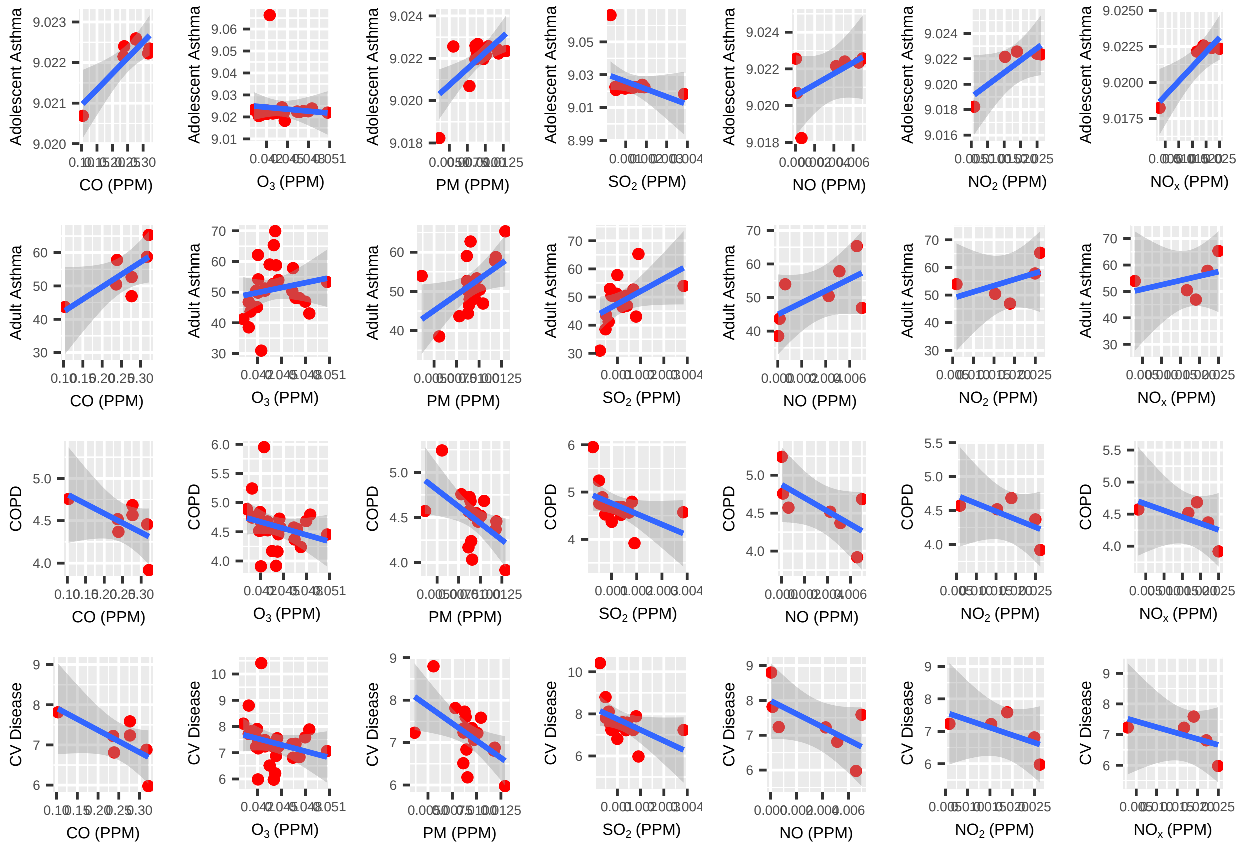


Figure 2 - Box-and-whisker Plots for Groups of Counties Highly Concentrated in Carbon Monoxide, Ozone, and Particulate Matter

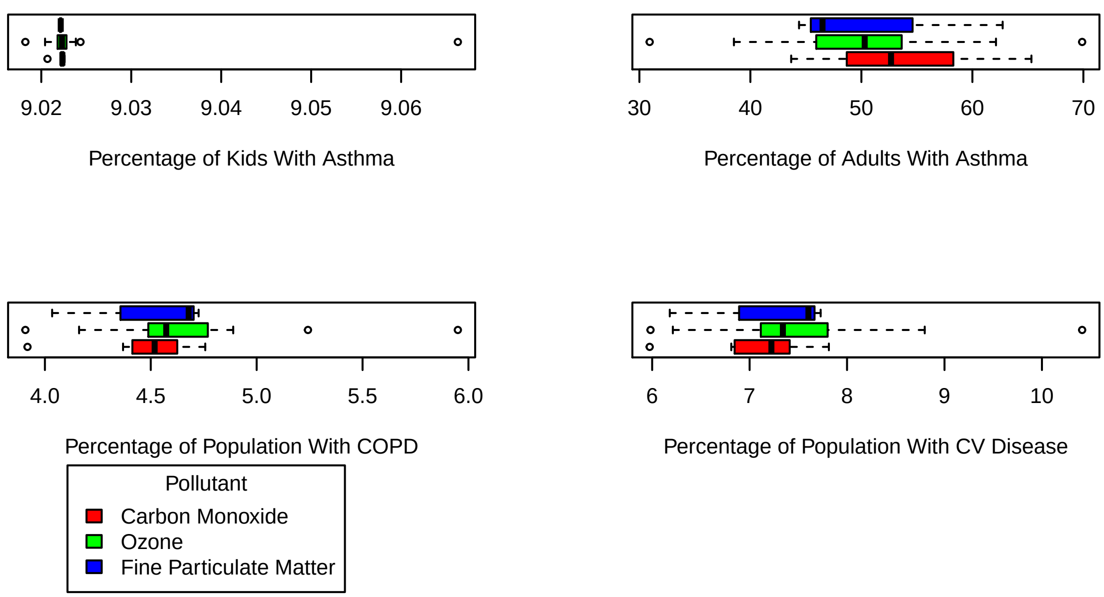


Figure 3 - Correlations Anticipated for Next Year

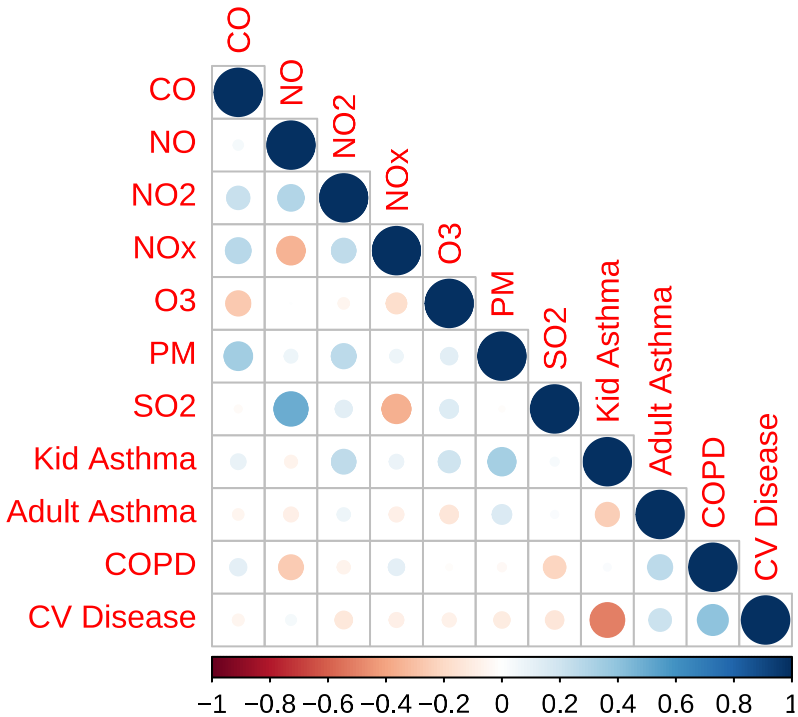


Figure 4 - Correlations Anticipated for Next 2 Years

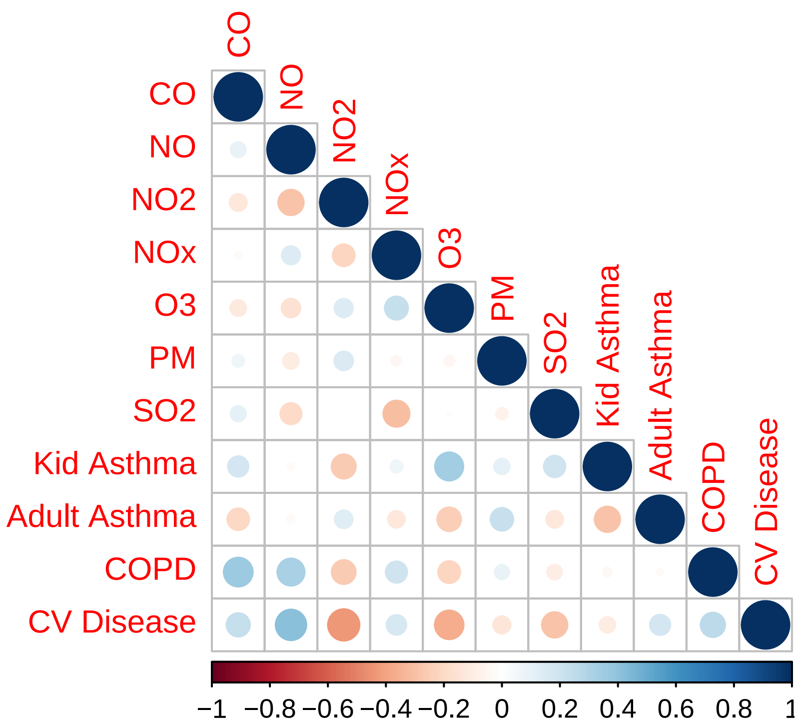


Figure 5 - Correlations Anticipated for Next 3 Years

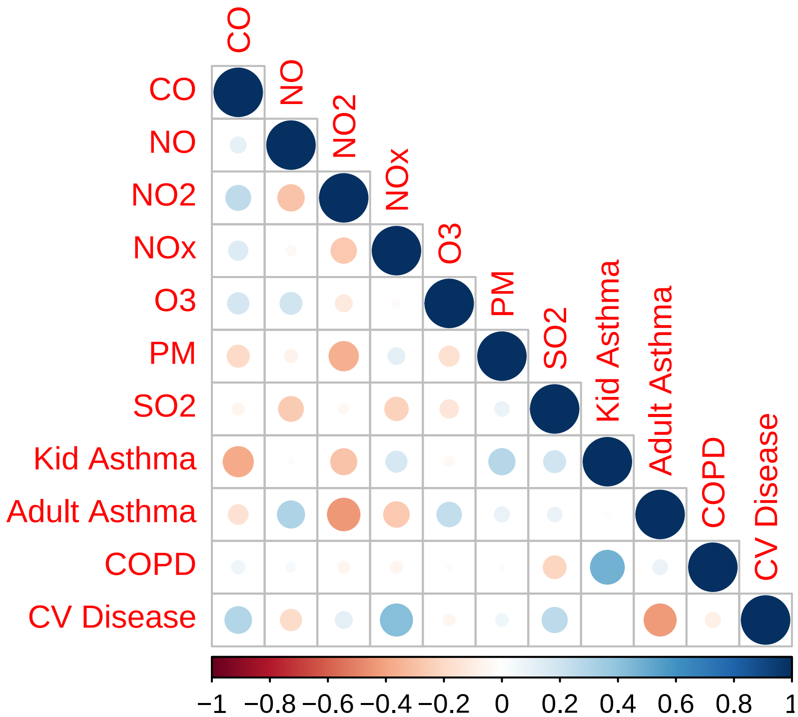


Figure 6 - Correlations Anticipated for Next 4 Years

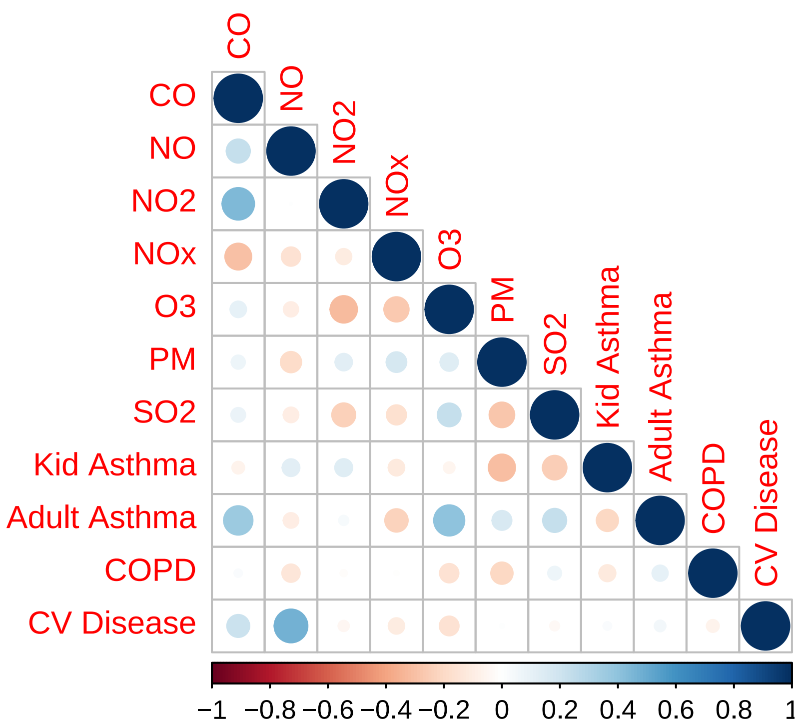
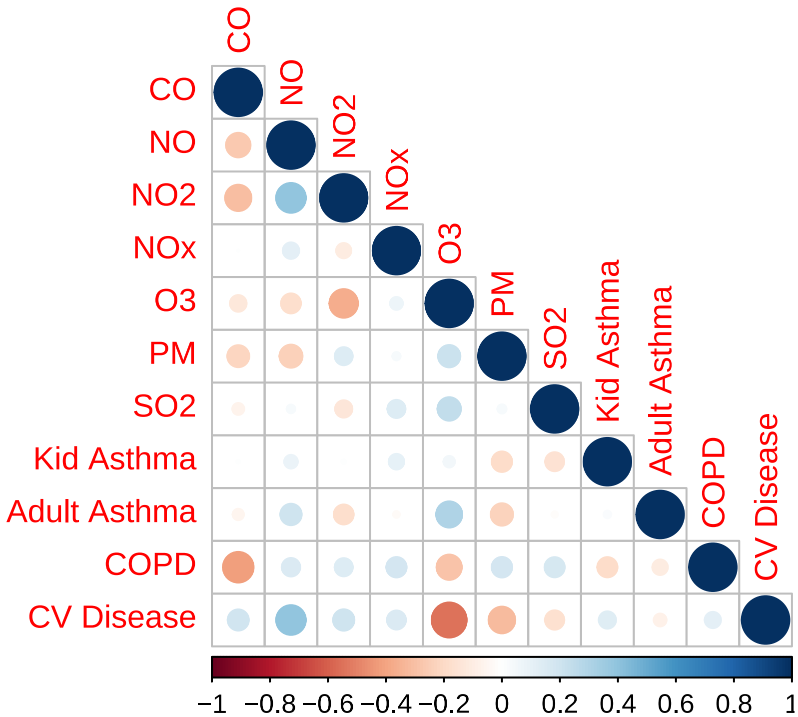


Figure 7 - Correlations Anticipated for Next 5 Years


CO
NO
NO2
NOx
O3
PM
SO2
Kid Asthma
Adult Asthma
COPD
CV Disease
−1
−0.8
−0.6
−0.4
−0.2
0
0.2
0.4
0.6
0.8
1