Molly McNamara
Springboard Capstone Project Proposal 1

US Air Pollution and Respiratory Health

High air pollution levels have been consistently documented as a major environmental risk to health. The World Health Organization has stated that the majority of the world's population is living in areas that do not meet air quality guidelines and estimated that outdoor air pollution caused 3 million premature deaths in 2012. While programs do exist in many countries to combat air pollution with the aim of reducing levels of harmful pollutants, weather patterns and geographical features can also influence the concentration of air pollution in certain urban areas. Given the health risks involved in elevated levels of air pollutants, it would be useful in a variety of ways for health care providers and hospitals to understand the acute incidence of respiratory illnesses during periods of poor air quality. Prediction of periods of increased patient illness could inform hospital staffing needs, mobilization efforts to reach high risk patients, and preventative care and messaging to reduce hospital visits. Given the challenges facing the health care system in the United States, there is a tremendous financial incentive to better manage high risk patients.

The Data

The primary dataset consists of daily levels of 4 primary air pollutants (Nitrogen Dioxide, Sulphur Dioxide, Carbon Monoxide and Ozone) and their air quality indices from major cities across the United States between 2000 and 2016. The data is sourced from the United States Environmental Protection Agency. Ideally this will be supplemented with hospital data on the intake of acute respiratory illnesses or mortality in these cities from the CDC and state health departments.

Potential Analysis Strategy:

The proposal for this project is to build a model that uses the pollutant data patterns to determine the best predictive features of acute respiratory illness, increased hospitalization or mortality. This may include identifying which of the major pollutants is most heavily linked to respiratory illness and, if the hospital data is granular enough, what types of illness may be more strongly associated with which pollutants.

Deliverables

After analysis of the data, the code used to develop the models will be shared. A formal report will be developed detailing the results of the exploratory data analysis, the statistical analysis and the results of the modeling, and this information will also be captured at a high level in a slide deck. The end goal would be to have some recommendations on the level of health risk for the level of pollutants that health care providers could utilize to better manage care of higher risk patients.