Module Project Proposal

Group number and group members:

Group 3
Adheesha Kumarasinghe, Elise Vong, Anne Bai, Jeremy O'Connor

Group advisor:

Jay DesLauriers

Title:

Exploring the correlation between long-term overexposure to air pollution and deaths from COVID-19 across all Greater London boroughs.

Research question:

"Does long term exposure to airborne toxins due to air pollution make one more susceptible to death from Covid-19?"

Null Hypothesis – Long term expose to air pollution in London was a considerable factor in the susceptibility of death from COVID-19
Alternate hypothesis – Long term expose to air pollution in London was not factor (or a minimal factor) in the susceptibility of death from COVID-19

Background:

A recent study conducted by NHS reveals that respiratory diseases is the third biggest cause of death in England, affecting one in every five people. In addition, hospital admissions for respiratory diseases have risen over the past seven years in a rate three times faster than the total admission [1]. For this reason, we would like to investigate in how pollution in major cities such as London has an impact on the health of the population and their ability to fight off a virus such as COVID-19 with a long-term exposure to high concentration of toxins which are abundant in air pollution.

We believe that long-term overexposure to air pollution has a severe impact on one's respiratory health, thus their ability to fight respiratory illness' such as COVID-19 was significantly less than those who were not exposed to as much air pollution in their lifetimes. We wish to make quantitative analysis on this theory to test its accuracy.

Interdisciplinary aspects of the project:

This project centres on studying environmental effects on human health, which draws together thinking from the medical and statistical fields to analyse relationships between outcomes and exposures. In addition, basic chemistry is required to understand molar

concentrations of the air pollutant molecules. A good quantitative analysis using programming is required to use extensive statistical analysis and correlation investigations. A health background is also required to draw links between the concentrations of certain pollutants, such as NO2, and their impact on the respiratory system.

Breakdown to individual steps:

1. Investigate which pollutant toxins are in the air composition in London and their respective concentrations in each of Greater London's Boroughs (GLB).
2. Investigate the health impacts of long-term overexposure to these toxins (i.e., respiratory illnesses/respiratory damage/asthma/etc).
3. Investigate the 'safe' concentration levels that it is suitable for a human to be exposed to over their lifetimes and check if these 'safe' limits are upheld across GLB.
4. Make quantitative estimations on the concentration of these toxins that an average resident in each of the GLB are exposed to on an annual basis.
5. Make quantitative estimations on the proportion of the population in each of the GLB that have a health impact, such as a respiratory illness, where one of the primary factors was long-term exposure to air pollutant toxins.
6. Make quantitative correlations on the COVID-19 death rates proportions of each GLB and our estimations for the proportion of each GLB with health defects due to over exposure to toxins as a direct result of air pollution.

Define the project core (aka the minimum viable product):

Is any correlation between deaths from COVID-19 and nitrogen dioxide levels in each of the GLBs. We should be able to confirm or reject the Null Hypothesis. This means having statistical analysis performed in at least one borough and a conclusion on the trend.

Define the project extensions:

We will look at specific instances of when members of the public encounter high levels of air pollution, such as their daily commute on the tube, and see if this has long-term health effects and/or implications to one's vulnerability to death from COVID-19.

We can look at relating the correlation between air pollution and death from COVID-19 to other Major cities worldwide such as Beijing or Los Angeles, to see if the correlation is similar for those cities.

Software tools needed for the projects:

Our Quantitative Analysis will be conducted on Python. We will import the Excel spreadsheets of data and use the Pandas module to sort, filter and organise our data. We will make extensive use of NumPy and SciPy for Statistical Analysis, such as chi squared fitting, linear regression, and uncertainty analysis from covariance matrices.

We will then use matplotlib to plot our results and trends for presentation and to include in our project report.

We are using Spyder IDE (on anaconda) for Python (version 3.8.8), the versions of modules are: Matplotlib: 3.4.3. NumPy: 1.21.2, Pandas: 1.3.4, SciPy: 1.7.1, Math:1.2.1, Operator: 2.0.1.

Datasets that will be used in the project:

NOTE: This is not the extensive list of our datasets/sources. We anticipate we will need further data, but we will include all sources in our Official Proposal.

Datasets on Covid-19:
Data on COVID-19, from the primary sources of the NHS and UK Government:
https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/deathsduetocovid19registeredinenglandandwales2020
Licence: not specified

https://data.london.gov.uk/dataset/coronavirus--covid-19--cases
Licence: not specified

https://data.london.gov.uk/dataset/covid-19-deaths-mapping-tool
Licence: not specified

https://data.london.gov.uk/dataset/coronavirus--covid-19--deaths
Licence: OGL v3

Dataset on pollution:
Data which can be chosen for specific gases and/or land-points in boroughs
https://www.londonair.org.uk/london/asp/datadownload.asp
Licence: OGL v3

In 2019 TFL took measurements on 1000s of filters across their networks, from on trains to platforms to the ticket halls.
https://tfl.gov.uk/corporate/transparency/freedom-of-information/foi-request-detail?referenceId=FOI-2525-1920
Licence: not specified

Pollution from 1000s of locations across London including schools and roads are available in raw spreadsheets from the Mayor of London's Office.
https://data.london.gov.uk/dataset
Licence: not specified

National Air Quality reports (for possible comparison) are available from The Department of environment, Food and Rural Affairs.
https://laqm.defra.gov.uk
Licence: not specified

Required hardware:

We only require standard computing processing power as of now – our laptops will be sufficient as we will be running Python using the Spyder IDE. To communicate our code and to have a backup/history log of the alternate versions, we will use the GitHub Desktop platform. It could be possible that we will need advanced desktops with a large processing power in the future due to the size of some of the datasets, however, we have access to these through Imperials computing network and remote connect.

Hardware Specifications:
Laptop - Ryzen 7 5800X (8 Core, 4.2 GHz), 2x8Gb RAM (DDR4 3200Mhz), 1Tb SSD, running Windows 11 (version 10.0.22621).

Agreed contributions from each group member:

GitHub Project Manager: Anne, Jeremy
Coding aspect: Anne, Jeremy
Research aspect: Adheesha, Elise
Presentation: ALL MEMBERS TO CONTRIBUTE
Poster: ALL MEMBERS TO CONTRIBUTE
Meeting Minutes taker: Elise
Timeline Manager: Anne

Agreed knowledge sharing:

Jeremy: I would like to learn more about the direct respiratory illnesses caused from the pollutants that the public are exposed to due to their lifestyles in London. I would like to investigate if the current 'safe' limits for pollutants is being exceeded and if so, by how much and what are the long-term health implications of this, and does it increase their likeliness to get covid-19.

Adheesha: I would like to learn how python could be used in data analysis, beyond the scope of the "introduction to python" course. This project will involve analysing the relationship between a potential risk factor and an outcome using large datasets. Such "real-world evidence" correlation studies are becoming an important part of public health medicine. As a medical student myself, I believe being well versed in this process would be invaluable.

Elise: I would like to learn how python can be applied to investigate whether pollution in London boroughs can have a significant effect on the respiratory system and if this will make a person more likely to die. I would also like to identify whether there is a "safe" level of certain pollutants which would not have a significant effect on the respiratory health of a person.

Anne: would like to learn more about respiratory diseases and see how python can be used to resolve real-life problems.

Agreed timeline:

Week 5(13/02-19/02): Submit draft proposal and begin writing the code.
Week 6(20/02-26/02): Clean and read data from files, edit the draft proposal.
27/02/2023 Final Draft (text file)
Week 7(27/02-05/03): Produce core results of data.
Week 8(06/03-12/03): Attempt on extension parts of the project and begin designing the poster and presentation.
Week 9(13/03-19/03): Produce the poster and rearrange GitHub page.
19/03/2023 Communicate project results (GitHub page, poster)
26/03/2023 Reflection on the module (reflective writing)

Agreed frequency and mode of communication:

For every week, two meetings would be held at minimum. One would be during the class session; the other would be on Wednesday 2pm. Both meetings would be done in-person. However, if some group members cannot come due to some other arrangements, they join in online for discussion. Minutes and attendance are recorded for each meeting, with summary notes found on the Teams channel.

Project repository:

All Project coding contributions will be made available to all members on the open share GitHub platform. The hyperlink is available below.
https://github.com/AnneBai0802/London-Underground-and-Respiratory-diseases

References/websites/AOB:

[1] https://www.england.nhs.uk/ourwork/clinical-policy/respiratory-disease/