Correlative Analysis of Victorian Asthma Incidence

992712, 587251, 1101512 and 1279113

Aims

We aimed to investigate potential links between the incidence of asthma, and the presence of air pollution, in particular that of PM10. PM10 refers to particulate matter found in dust or smoke, identified by the Environmental Protection Authority of Victoria as a common and dangerous air pollutant. PM10 particle levels are measured by weight, usually in micrograms per cubic metre, of particles with a diameter of one hundredth of a millimetre, or smaller.[1]

Asthma is a chronic health condition of varying severity. While there have been studies showing correlations with a number of factors, research has not been able to fully isolate or understand the causes. Asthma manifests itself with the inflammation and narrowing of small air passages in the lungs, which can cause difficulty breathing. Attacks can be triggered by common allergens, smoke, pollutants and viruses, among other things. It is estimated that asthma is responsible for almost half a million deaths every year,[2] and around one in nine Australians are affected, one of the highest prevalences in the world.[3]

Better understanding of asthma and its causes is vital to both prevention and treatment. With better data we can help improve the health in Victoria communities, inform public policy and save lives.

Data sources

We drew from three datasets for our statistical analysis:

- PM10 readings were taken from Victoria's Environmental Protection Authority's historical datasets for 2016 (found at https://discover.data.vic.gov.au/dataset/epa-air-watch-all-sites-air-quality-hourly-averages-yearly/historical).
- Information about chronic diseases (including asthma) was taken from Torrens University's Public Health Information Development Unit's modelled estimates for 2011-2013. (https://data.aurin.org.au/dataset/tua-phidu-sa2-chronicdisease-modelledestimate-sa2)
- The LGA11 Health Risk Factors - Modelled Estimate 2011-2013 dataset was used for information about weight and smoker status. (https://data.aurin.org.au/dataset/tua-phidu-lga11-healthriskfactors-modelledestimate-lga2011)

To link these datasets together we decided to operate on the local government area level, in order to get a wider view and eliminate potential noise from finer measurements. Accordingly,

---

[1] Environmental Protection Authority, Victoria. *PM10 Particles in the Air.* https://www.epa.vic.gov.au/for-community/environmental-information/air-quality/pm10-particles-in-the-air

[2] World Health Organisation. *Asthma Key Facts*. https://www.who.int/news-room/fact-sheets/detail/asthma

[3] Asthma Australia, *About Asthma*. https://asthma.org.au/about-asthma/

each of the eight PM10 measurement sites were manually mapped to an LGA code. The SA2 level chronic disease set data was aggregated and averaged for the relevant LGA. Measures we thought were out of scope or not useful for our analysis (e.g. carbon monoxide measures, musculoskeletal diseases etc.) were excised.
The PM10 measures were aggregated in three different measures. The EPA provides only 1 hour averages. We look at the maximum of these, the year-long average, and the maximum average daily measurement, as per EPA suggestions. The first one was discarded, being judged as too prone to noise, and potentially affected by unrepresentative samples from the testing site for the area (e.g. rare events like sawdust or a fire close near the reading).

Both the chronic diseases sets and the health risk factors sets recorded the percentage of population in each area self-reporting the particular behaviour or ailment. We took three measures describing overweight individuals in order to explore the nuance between these categories. ‘Pre-obese’ refers to individuals self-reporting as having a BMI of 25-30, ‘obese’ refers to individuals with BMI of 30+ and ‘overweight’ spans both of these categories; individuals with BMI of 25+. The average of the figures for each LGA examined is displayed in Table 1.

Table 1: Health Factors and Percentage of Victorians suffering

Health Factors	**Average % of Victorians afflicted**
Smoking	14.24%
Heart Disease	6.86%
Pre-obesity (25-30 BMI)	31.15%
Obesity (30+ BMI)	17.05%
Overweight (25+ BMI)	48.2%
PM10 (Maximum daily read)	50.24 µg/m3

Analysis
The Pearson correlation coefficient was calculated between data representing asthma affected individuals and data representing various health factors that may be contributing factors to asthma such as smoking, heart disease and obesity.

The resulting correlations are shown in the below Table 2. From these results it can be deduced that excluding heart disease, each factor has at least a moderate positive correlation (>0.4) with asthma with pre-obesity and overweight showing a large correlation (>0.5).

In addition to the correlations measured between the asthma and each of our health risk indicators we also calculated the mutual information shared between the datasets to better understand their potential relationships. Mutual information can help reveal the presence of non-linear relationships between data, where simple correlation cannot.

The PM10 data-set was discretized using domain knowledge provided by The EPA. The EPA categorizes PM10 readings into 5 complete standards of PM10 measured and averaged across a one hour period: less than 40µg/m3 is considered 'Good', 40-80µg/m3 is considered 'Fair', 80-120µg/m3 is considered 'Poor', 120-300µg/m3 is considered 'Very poor' and more than 300µg/m3 is considered 'Extremely poor'.[4] The data was binned into PM10 of less than 40µg/m3, 40-80µg/m3 and more than 80µg/m3; the 'Good', 'Fair' and 'Poor' categories.

The asthma percentage data was discretized into two bins; less than 10% and more than 10%.
The smoking percentage data was discretized into 3 bins; below 10%, 10-20% and above 20%.
The heart disease percentage data was discretized as below 6.9% and 7% and above, which allowed for equal frequency binning.
The pre-obese, obesity and overweight data were all discretized with equal length binning.

Table 2: Health Factors and their Entropy, and Correlation and Normalized Mutual Information with Asthma

Health Factors	Pearson Correlation	Entropy	Normalized Mutual Information
PM10 (Maximum daily average)	0.344	1.299	0.393
Smoking	0.303	1.06	0.311
Heart Disease	-0.130	1.0	0
Pre-obesity (25-30 BMI)	0.545	0.811	0.384
Obesity (30+ BMI)	0.415	1.5	0.5
Overweight (25+ BMI)	0.586	1.299	0.393

Asthma entropy was found to be 1.0.
The mutual information between the asthma data and each of the other datasets appear fairly consistent with their respective Pearson correlations. This does not indicate the presence of a non-linear relationship between any of the datasets.

[4] Environmental Protection Authority, Victoria. *PM10 Particles in the Air.* https://www.epa.vic.gov.au/for-community/environmental-information/air-quality/pm10-particles-in-the-air

In addition to the above analysis, the Pearson correlation between the average maximum 24hr PM10 levels and the mentioned health factors were calculated. This data is displayed in table 3, and was computed to better understand the potential correlations between air pollution and various health factors. The correlation between PM10 and obesity was in the higher end of the moderate range, being 0.49, slightly above the correlation between obesity and asthma at 0.41.

Table 3: Health Factors and their Correlations with PM10 readings

Health Factors	Pearson Correlation with Max 24hr PM10
Asthma	0.34
Smoking	-0.16
Heart Disease	0.34
Pre-obesity (25-30 BMI)	0.22
Obesity (30+ BMI)	0.49

These relationships perhaps indicate that populations living in areas with higher levels of air pollution are more likely to be obese and a combination of these health risks may lead an individual to develop asthma. This is a possibility based on the higher positive correlation between obesity and asthma than PM10 levels and asthma (0.34).

Results and Significance

The relationships uncovered have been fairly strong. However, it is difficult to determine the precise effect being overweight or living in an area with high PM10 levels has on the development of asthma. It is quite possibly the case that these factors are instead linked to another unknown variable, for example socio-economic circumstances or age.

It seems reasonable to conclude however that both pollution levels and being overweight play *some* part in the development of asthma, and public policy should be centred around reducing the former through better regulations and initiatives with industry and the latter through better education, especially in schools and at the GP level and policy-makers should strongly consider harsher measures than the current five star rating system, for example a sugar tax. Particular attention should be paid to child obesity and exposure to air pollutants and second-hand smoke, given the majority of cases develop in early childhood or adolescence.[5]

The following three graphs demonstrate the strong possibility that a population's overweightness and exposure to PM10 increases asthma incidence. Our findings also reflect emerging academic research.[6][7][8]

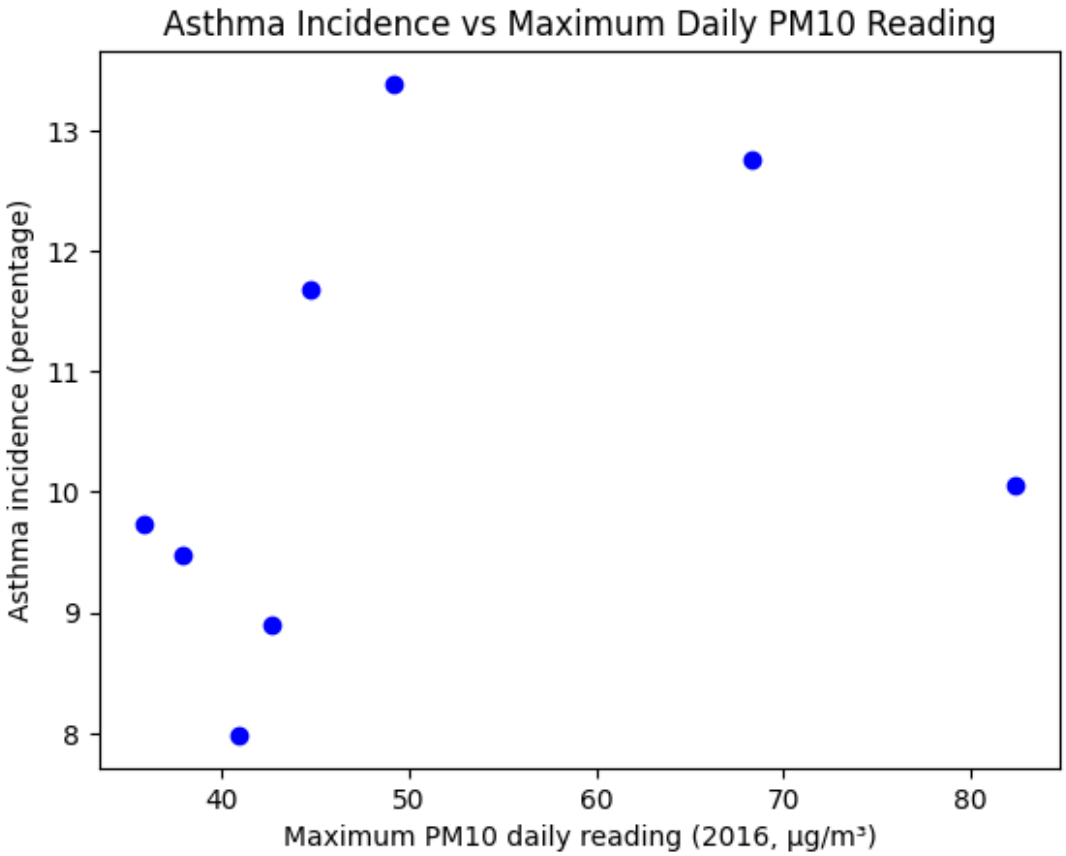
---

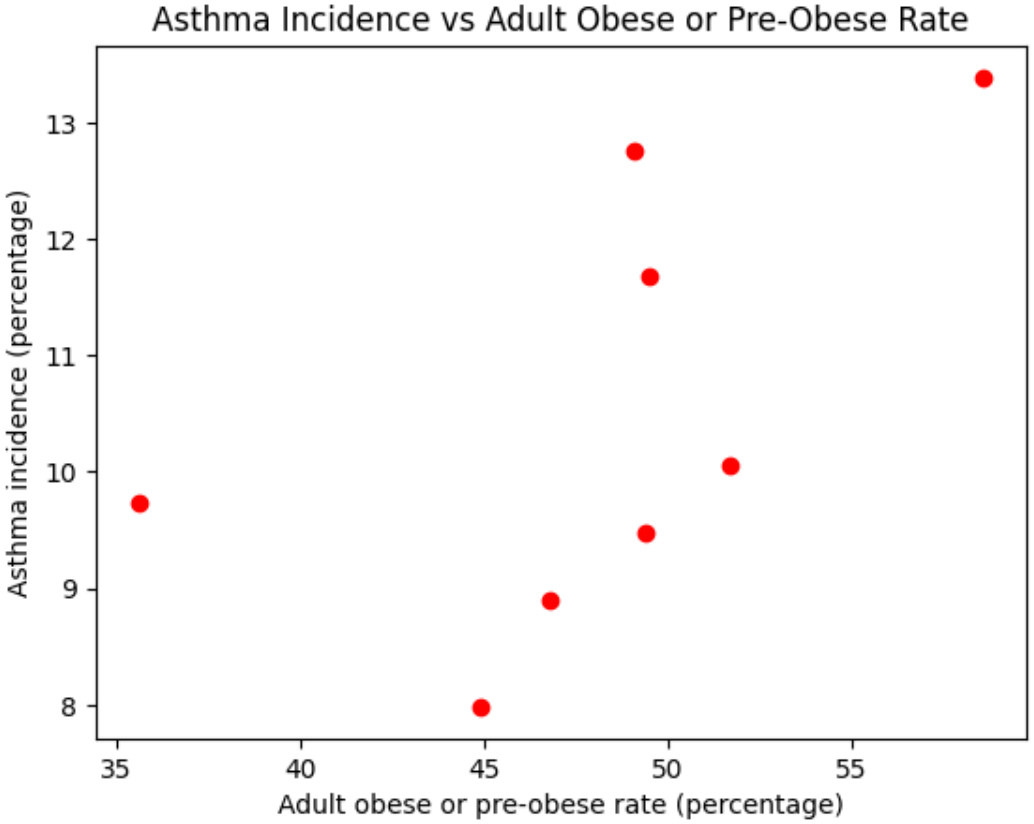
[5] Burdon, J. *Adult-onset asthma.* https://www.racgp.org.au/afp/2015/august/adult-onset-asthma

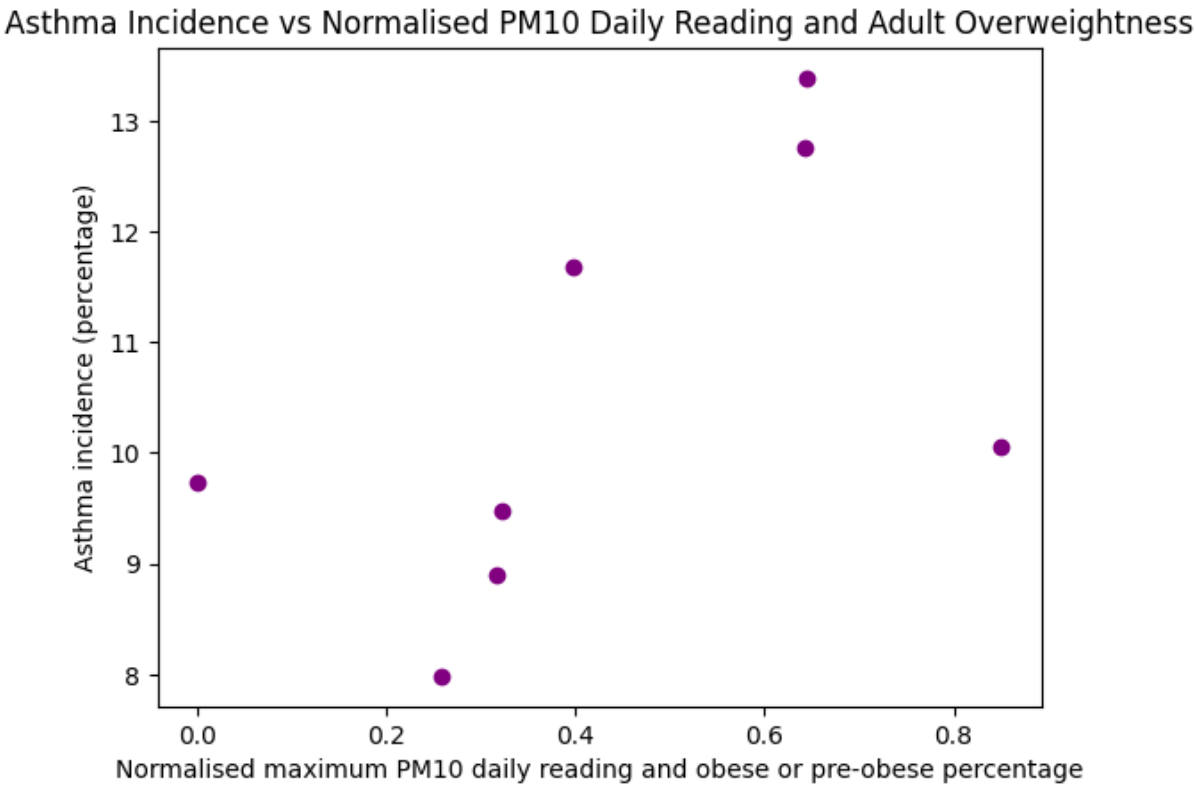
[6] L. Akinbami and C. Fryar. *Current Asthma Prevalence by Weight Status Among Adults: United States, 2001–2014*. https://www.cdc.gov/nchs/products/databriefs/db239.htm

[7] U. Peters et Al. *Obesity and Asthma.* https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5973542

[8] C. Keet et Al. *Long-Term Coarse Particulate Matter Exposure Is Associated with Asthma among Children in Medicaid*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5855070/


Asthma Incidence vs Maximum Daily PM10 Reading
Asthma incidence (percentage)
13
12
11
10
9
8
40
50
60
70
80
Maximum PM10 daily reading (2016, µg/m³)


Asthma Incidence vs Adult Obese or Pre-Obese Rate
Asthma incidence (percentage)
13
12
11
10
9
8
35
40
45
50
55
Adult obese or pre-obese rate (percentage)



Limitations and future directions

Our datasets had a number of limitations unfortunately.

The two health datasets were based on modelled estimates from self-reported data. Self-report is well-documented as an unreliable measure[9] and personal characteristics like weight are especially likely to be underreported.[10][11]

The Environmental Protection Authority unfortunately only collects PM10 data for eight locations. While these did include metro, outer suburban and rural collection points, a greater breadth would have been useful.

Future studies might look at a wider range of pollution measures, and also might like to consider the effects of gas cooking. Studies on a smaller level might be useful as well, such as examining the smoker status of parents and development of asthma as opposed to wider local government level statistical correlations. Larger time frames and information from more years would also assist in showing trends and changes, and a more longitudinal approach could give more conclusive results.

---

[9] J. Garcia and A. Gustavson, Association for Psychological Science. *The Science of Self-Report*. https://www.psychologicalscience.org/observer/the-science-of-self-report

[10] A. Bowring et al. *Measuring the accuracy of self-reported height and weight in a community-based sample of young people.* https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-12-175

[11] Australian Bureau of Statistics. *Self-reported height and weight.* https://www.abs.gov.au/articles/self-reported-height-and-weight