The Potential Impact of Air Pollution Coverage in the Media on Respiratory Disease Admissions

Pui-Shan Tang [1], James Goulding [3], Gavin Smith [3], Dominick Shaw [1,2]

1. East Midlands Academic Health Sciences Network, 2. Nottingham Respiratory Research Unit, Nottingham City Hospital, 3. Horizon Institute, University of Nottingham.

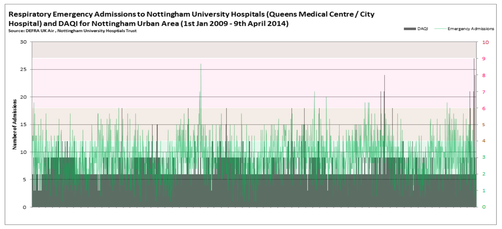
Introduction

Air pollution is well know to exacerbate respiratory disease [1, 2]. However, when air quality warnings are provided by authoritative bodies (e.g. MET office) subsequent media coverage may be disproportionate. In this study we explore whether there is an association between respiratory admissions and media pollution coverage via non-linear predictive models, and to potentially predict respiratory admissions.

Method

The relationship between media coverage of pollution and respiratory admissions was examined as follows:

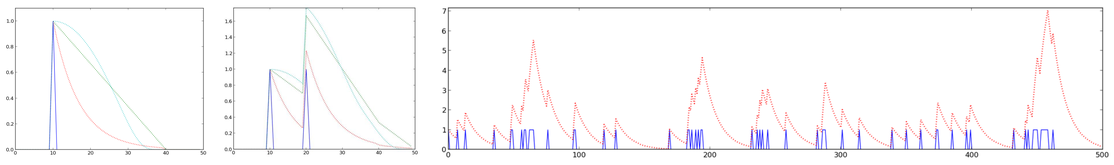
1. Baseline regression models were generated to predict daily respiratory admission episodes over the period 1st January 2009 - 9th April 2014. Predictors consisted of daily logs for PM10 particulate matter, PM2.5, Nitric oxide, Nitrogen dioxide, Ozone, Black carbon, Mean Temperature, Precipitation (obtained from *National Oceanic and Atmospheric Administration* data) and the DAQI Air Pollution Index. Models were optimized via cross-validation using daily number of admissions to Nottingham University Hospitals, which were identified by over 70 ICD10 codes.



For the study period, the air pollution level in Nottingham is low for the majority of the time, and is rarely moderate or high. The pattern of emergency admissions is not clearly linked to the level of air pollution.

2. Time series of **levels of media coverage** were generated by applying **kernel density estimation (KDE)** to daily counts of online news articles featuring pollution and air quality issues from 1st January 2013 – 9th April 2014. A range of bandwidths were examined (1, 10, 25, 50 and 100 days) in order to reflect the ongoing impact these news stories might have, using both linear and exponentially decaying kernels.

a. Single event **b. Multiple events** **c. Example yearly model of media impact using an exponential decay kernel**



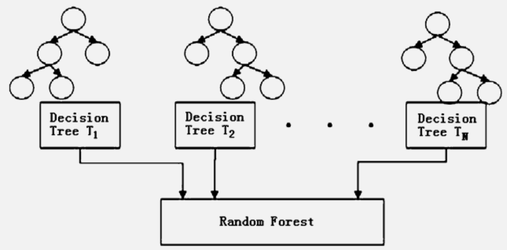
Figures above illustrate the modelling of news story events into a time series using KDE. Figure (a) illustrates a single news event (blue) being modelled using a linear (green) or different exponential (red, cyan) kernels.

3. Relationships between the input variables (**pollution levels and media coverage**) and the target variable (**respiratory admissions**) were examined using both 1. traditional linear regression techniques, and 2. non-linear predictive models drawn from the field of **machine learning**. The accuracies of these models were tested **with and without** integration of media coverage in order to ascertain the contributing impact of that variable.

.

Results

- There were a total of 16,756 emergency admissions for respiratory problems over the whole period, and the predictive model was trained on a subset of this (4,133 emergency admissions).
- Of the multiple baseline predictive models tested a **random forest** provided optimal results for air-quality predictors. When predicting daily respiratory admissions, the model's accuracy was **19.90%** better than simply predicting mean daily admissions, with an average root mean square error (RMSE) of 7.50.
- However, introduction of the media-coverage variable, this RMSE was reduced to 7.32, increasing improvement over mean prediction to 21.85%. While this did reflect an improvement in admissions forecasting, a corrected t-test indicated that these differences were not statistically significant (with a p-value of 0.06).



The random forest model's result is an amalgamation of the results of multiple decision trees, as illustrated in the figure on the left.

*An ensemble of decision trees can provide a robust exploration of **non-linear relationships**, offering insight that can easily be missed by traditional correlation-based regression techniques.*

Conclusion

Initial results indicate that consideration of media coverage may well offer improvements in predicting respiratory admissions. However, more datapoints are required as currently this effect was not found to be statistically significant. While this relationship requires further investigation, models informed by media **coverage cannot currently be considered to be accurate enough for use in a practical setting**.

Next Steps

The effect of Media on health seeking behaviour is still relatively unexplored and our attempt to model it using KDE could be refined through further qualitative research. There's also a need to define what is considered "media" - in this study only online news stories were used in the model, but media coverage of this and similar events was evident in different platform/formats (e.g. TV, Radio, Social Media, Printed Newspapers). The advantages of using machine learning techniques were restricted due to limitations in the quantity and quality of data currently available.

References

1. GWK Wong, FWS Ko et al 2001 "Temporal relationship between air pollution and hospital admissions for asthmatic children in Hong Kong" Clinical and Experimental Allergy Volume 31, p 565-569
2. FWS Ko, W Tam et al 2007 "Temporal relationship between air pollutants and hospital admissions for chronic obstructive pulmonary disease in Hong Kong" Thorax, Volume 62, p 780-785

This work was jointly supported by the **RCUK Horizon Digital Economy Research Hub grant, EP/G065802/1** and the **EPSRC Neodemographics grant, EP/L021080/1.**