Do You Think Air? Public interest in Air Pollution

Barry Peter Kirby[1], Sam Paul Lewis[2] & Paul David Lewis[3]

[1] K Sharp Ltd, [2]School of Management, Swansea University and Vindico ICS, [3]Medical School, Swansea University and Vindico ICS

ABSTRACT

Air pollution is an issue that concerns everyone impacting both health and the environment, but generally it takes a back seat when there are more pressing issues. This paper kicks off a research programme looking at air quality with an exploration of the initial question of just how aware and interested are the general public in air pollution in the UK, with a long term aim of developing behavioural change methods underpinned by meaningful data to benefit long term health.

KEYWORDS

Air pollution, Google Trends, public interest, media headlines, behaviour change.

Introduction

Air pollution is the leading environmental cause of early death contributing to the equivalent of 5% of all deaths globally (World Health Organization, 2017). The UK has statutory obligations to ensure that concentrations of specified air pollutants are kept below certain limits. In order to address this issue, we have to understand how to influence people in order to change their behaviour. In terms of health impacts, two of the most damaging pollutants are fine particulate matter ($PM_{2.5}$) and nitrogen dioxide (NO_2) and epidemiological studies have shown that long-term exposure to air pollution can reduce life expectancy, mainly due to cardiovascular and respiratory causes (COMEAP, 2009). The UK annual mortality burden of air pollution is greater considered than 20,000 and annual costs to society estimated to be more than 20 billion pounds (Royal College of Physicians, 2016). Whereas air pollution has reduced significantly over the last fifty years, the UK still has illegal levels of NO_2.

Road traffic emissions account for around two-thirds of air pollution in urban areas and, along with new legislation and new cleaner technologies, the UK Government recognises that it is behavioural change by citizens that will help tackle air pollution (UK Government, 2019). The choices an individual makes in the decisions they make (e.g. type of heating to install, type of car to buy) and the way they behave (e.g. transport use, engine idling, use of wood burners) can significantly impact our air quality. Understanding public knowledge of air pollution is also fundamental in helping governments develop policy and communications around air pollution and reduce exposure. Some key findings of a recent study on people’s perceptions of air pollution were that level of environmental concern is strongly associated with knowledge, attitudes and behaviour on air quality issues but also people with health conditions are more likely to be aware of and concerned about it (Turner and Struthers, 2018).

Realising that a key solution to reducing air pollution is the influencing of behavioural change in citizens, the aim of this study was to gain an initial understanding of public interest and awareness of air pollution in the UK. Using Google Trends, we collated monthly air pollution-related search data between January 2010 and September 2020 and aligned with monitored air pollution levels and BBC News UK web-site headlines as a proxy to media coverage over time. We initially visualised and statistically analysed the temporal patterns of Google searches for air pollution terms and actual $PM_{2.5}$ and NO_2 pollutant levels at that time. We then used multivariate statistical analysis to assess the interplay and covariation between patterns of Google searches, media headlines and pollutant levels to provide insight into the influence of media on public awareness. Google search data aggregates billions of instances each day. Google Trends, as a web-based tool, can be used for comparative keyword research and to discover event-triggered spikes in keyword search volume. Google Trends has been widely used across research disciplines including predicting economic trends (Choi and Varian, 2012), measuring concerns of the COVID-19 pandemic (Knipe *et al*., 2020) and predicting impacts of Brexit (Simionescu *et al*., 2020). Google Trends has also been used to assess the public understanding of air pollution but research has focused generally on Far East countries (Dong *et al*.,2019; Cori *et al*., 2020; Misra and Takeuchi, 2020) with no analysis carried out on UK data to our knowledge.

Methods

Google Trends data are available from 2004 to the present day (Google Trends, 2015). Querying Google Trends provides a representative sample of Google searches within each time-point over a specified time period (sampling is considered sufficient by Google as they handle billions of searches per day). Google Trends uses a normalization procedure to allow for comparisons between search terms. Briefly, search results are normalized to the time and location of a query whereby each data point is divided by the total searches at a location and time range to compare relative popularity. This normalisation process prevents places with higher search volumes (e.g. London) as always be ranked highest by popularity. The resulting numbers across the chosen timeline are scaled on a range between 0 to 100 based on a topic's proportion to all searches on all topics. Thus, a score of 80 for a search term at a timepoint shows that the term made up a large majority of Google searches carried out in that region on that day, week or month. For time-periods spanning greater than five years, Google Trends returns search interest data reported for each month in the timeline. For timelines between nine months and five years, the data are reported as weekly trends.

In this study, we used the R Environment for statistical computing to retrieve, visualise and analyse all data (R Core Team, 2017). We used the gtrendsR R package (Massicotte and Dirk Eddelbuettel, 2020) to retrieve Google Trends search interests for several air pollution related search terms between January 2010 and September 2020 via. Search terms used were 'air pollution', 'air quality', 'air quality index', 'PM2.5' and 'nitrogen dioxide'. We did not use the term 'NO2' as it is ambiguous and led to retrieval of search interests unrelated to nitrogen dioxide. The search term 'particulates' retrieved too few search interests to allow a valid comparison with other variables. UK rural background monthly mean data for PM2.5 was retrieved for six Department for Environment Food & Rural Affairs (Defra) managed Automatic Urban and Rural Network (AURN) monitors at Auchencorth Moss, Chilbolton Observatory, Lough Navar, Harwell, Narberth and Rochester Stoke using the openair R package (Carslaw and Ropkins, 2012) over the same study period. PM2.5 data were further averaged per month across each site to provide a single UK mean time-series. Monthly mean NO_2 data were retrieved from three distributed rural background AURN monitors at Ladybower, Narberth and Rochester Stoke.

To retrieve data on UK media coverage for air pollution over the 10-year period, the BBC News UK website was searched as a proxy using Google and the search parameters 'site:www.BBC.co.uk/news/uk', 'allintitle:' entered in the search bar followed by either 'air pollution' or 'air quality'. The search range was set between January 2010 and September 2020 using the 'Tools' option on the Google search page. The search returned a total list of 194 headlines that were categorised according to whether the news story was national or local. Headlines that were relevant to a UK air pollution problem (i.e. not local within a small urban area and reaching a small audience) were then retained yielding a final list of 74 stories. A single multivariate time-series dataset was then created by aligning all variables by matching time-points.

Pearson's correlation was used to determine similarity between monthly distributions of Google searches of 'air pollution' and mean monthly levels of $PM_{2.5}$ and NO_2. Principal component analysis (PCA), a multivariate statistical method and data reduction technique, was used to examine how Google search terms covary, over time, with pollutant levels and media coverage. PCA has been widely used in human factors research to aid interpretation of how large datasets vary by reducing the number of variables to a few, interpretable linear combinations of the data (Papadimitriou *et al*., 2017). These linear combinations correspond to uncorrelated variables called principal components that maximize variance. PCA was initially used to determine any long-term patterns (January 2010 to February 2020) that might exist, pre-COVID-19 lockdown, between people's use of Google in the UK to search for information on air pollution, levels of measured air pollution and media headlines for air pollution. PCA was then applied to an extended dataset (January 2010 to September 2020) to determine if and how air pollution levels and media coverage during COVID-19 lockdown impacted on patterns of Google searches. PCA was carried out using the FactoMineR (Le *et al*., 2008) package in R.

Results

Visualisation of time series patterns

Timeseries plots between January 2010 and September 2020 for Google search terms 'air pollution', 'air quality', 'air quality index', '$PM_{2.5}$' and 'NO_2' are shown in Figure 1. Monthly mean background levels for $PM_{2.5}$ and NO_2 are also shown. Searches for 'air pollution' and 'air quality' were much more common than for 'air quality index' or either pollutant name. The levels of searches for 'air pollution' and 'air quality' began to increase from 2014. This increase coincided with an increase in BBC UK news headlines for 'air pollution' over the same period. Notable large increases in Google searches for 'air pollution' and 'air quality' were observed at key months when high air pollution levels were reported widely in the media.

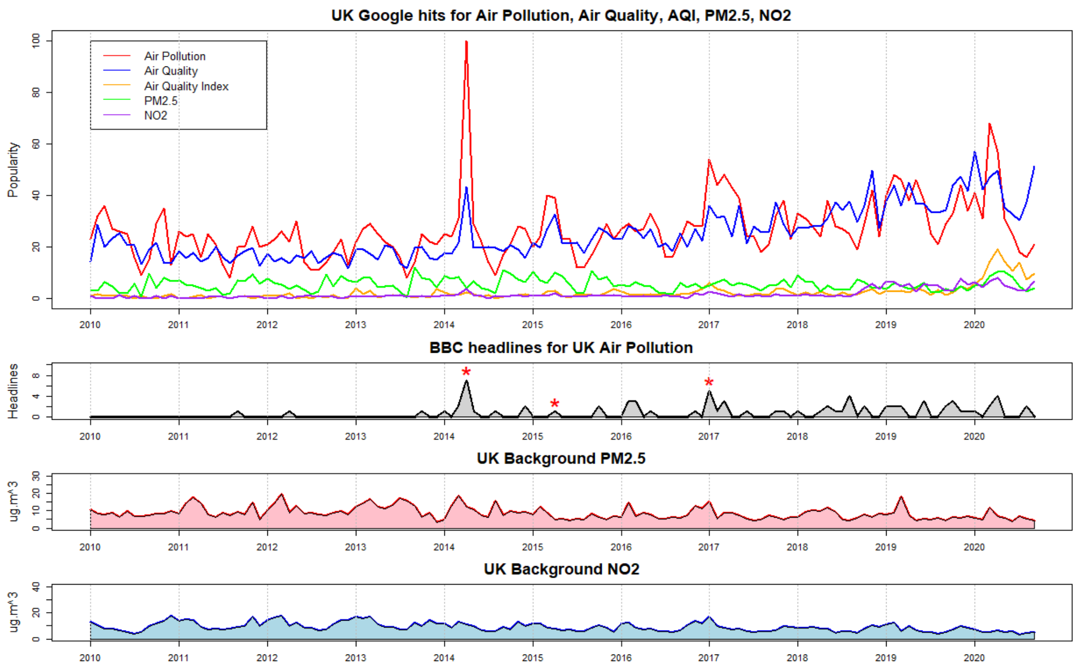

Figure 1: Timeseries plots between January 2010 and September 2020 for air pollution related Google search terms ('air pollution', 'air quality', 'air quality index', '$PM_{2.5}$', 'NO_2'), BBC

headlines for UK air pollution, monthly mean background $PM_{2.5}$ and monthly mean background NO_2. Red asterixis denote significant pollution events.

Analysis of monthly Google search patterns and pollutant levels

$PM_{2.5}$ and NO_2 levels are heavily influenced by meteorological factors. As such, pollutant levels are well known to show seasonal patterns of change where they increase over winter months, through Spring, and then reduce over summer months. Visual analysis of the timeseries patterns of Google searches suggested that search frequencies also change over the year and this pattern is repeated. We therefore grouped each pollutant and 'air pollution' search data by month over the date range January 2010 to February 2020 to determine if the distributions were correlated. Both $PM_{2.5}$ (r = 0.68, P= 0.016) and NO_2 monthly median levels showed a strong and significant (r = 0.63, P = 0.030) correlation with the monthly median for 'air pollution' search interest (Figure 2). A similar result was observed when comparing pollutant levels with the 'air quality' search term.

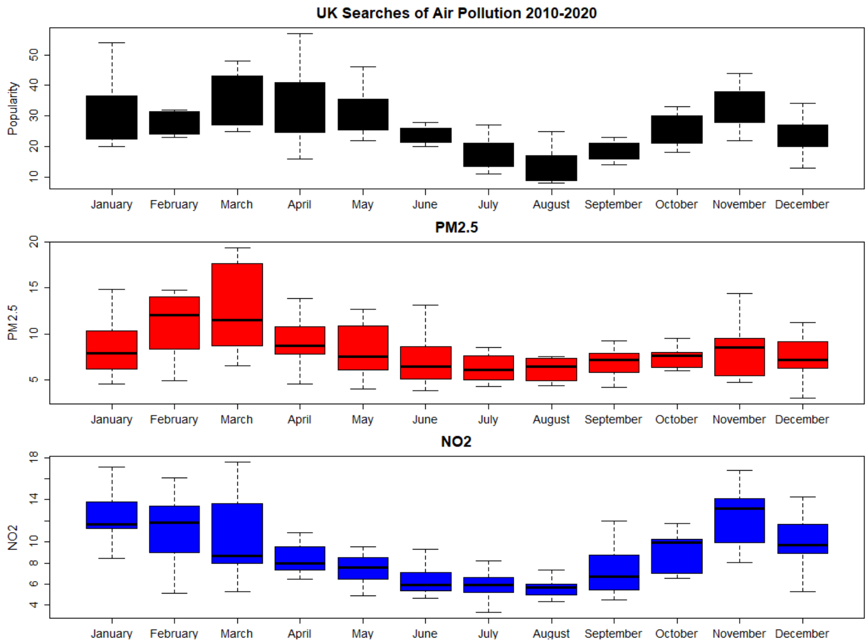

Figure 2: Boxplots showing the monthly distributions between January 2010 and February 2020 of Google searches for 'air pollution', mean $PM_{2.5}$ levels and mean NO_2 levels.

Multivariate analysis of Google search patterns, media headlines and pollutant levels

Using PCA, we explored the underlying structures of how Google search patterns for 'air pollution', 'air quality', 'air quality index', 'PM2.5' and 'nitrogen dioxide' covaried over the time with media headlines and pollutant levels over the decade prior to COVID-19 lockdown. With known reductions in NO2 levels across UK towns and cities during lockdown, with high media interest we also assessed whether public interest increased during this period. For the PCA analysis of January 2010 to February 2020 data, we retrieved 3 principal components that each explained greater than 10% of the overall variance. The percent contributions of each variable to each component are shown in Figure 3 (A-C) and the percent contributions of the 15 most significant months are shown in Figure 4 (A-C).

The first component (PC1, Figure 3A), which explained 38.6% of the total variation, showed a strong significant contribution from BBC headlines and a lesser but still significant contribution from Google searches for 'air quality', 'air pollution' and 'air quality index'. Figure 4A shows that high polluting months (including March 2014, April 2015, January 2017, March 2017 and January

2020) also associate strongly with PC1 with significant dates tending to be post-2016. This suggests that media coverage of major air pollution events leads to high levels of public interest in air pollution and that interest has increased since 2016. Interestingly, Google searches for $PM_{2.5}$ contributed around 9% to the component which was much higher for that of NO_2 suggesting that the public are having increasing awareness of that pollutant despite greater coverage of illegal NO_2 levels since 2015.PC2 (Figure 3B), which explained 27.4% of the variation showed a strong contribution from the two pollutant levels but not other variables. Significant contributing dates were generally pre-2016 confirming that there was little association between public interest during that period and pollution levels *per se*. PC 3 (Figure 3C) explained 12.2% of the variance with BBC headlines, Google searches for ‘$PM_{2.5}$’ and ‘air quality index’ as well as NO_2 level being significant contributors. PC3 was difficult to interpret but it’s possible it generally accounted for media reports occurring at times of increased air pollution such as August 2018 when the BBC report on research studies into the effects of air pollution on mental health (BBC News, 2020a) and cardiovascular disease (BBC News, 2020b).

The PCA performed on data including the lockdown period showed that public interest increased as shown by PC1 (Figure 3D) where Google searches for ‘air quality index’ increased in contribution. Google searches during early lockdown in March and April were the greatest contributors to PC1 during this period (Figure 4D). Indeed, the large increase in search for this term during that time can be seen in Figure 1. PC 2 (Figure 3E) also shows the increase in BBC coverage during this time relative to pollutant levels. PC3 (Figure 3F) perhaps shows the greatest interpretable change in public interest where this time Google searches for ‘NO2’ contribute most strongly along with BBC headlines and Google searches for ‘PM2.5’. The variances explained by each of these principal components remained the same as when lockdown data were not included.

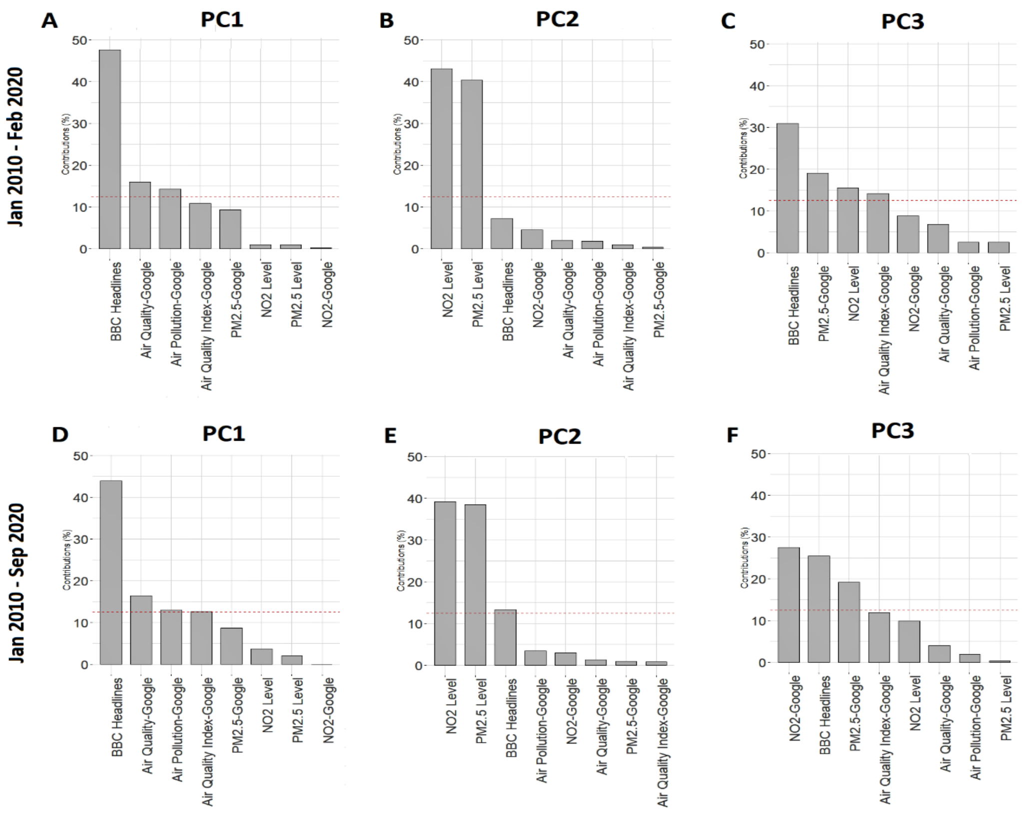

Figure 3: Percent contributions of Google search, pollutant level and BBC headlines to the first three principal components. A-C: principal components retrieved for the PCA on January 2010 to

February 2020 data. D-F: principal components retrieved for the PCA on January 2010 to September 2020 data. If a variable had a percent contribution greater than the level shown by a red dotted line, then that contribution was deemed significant.

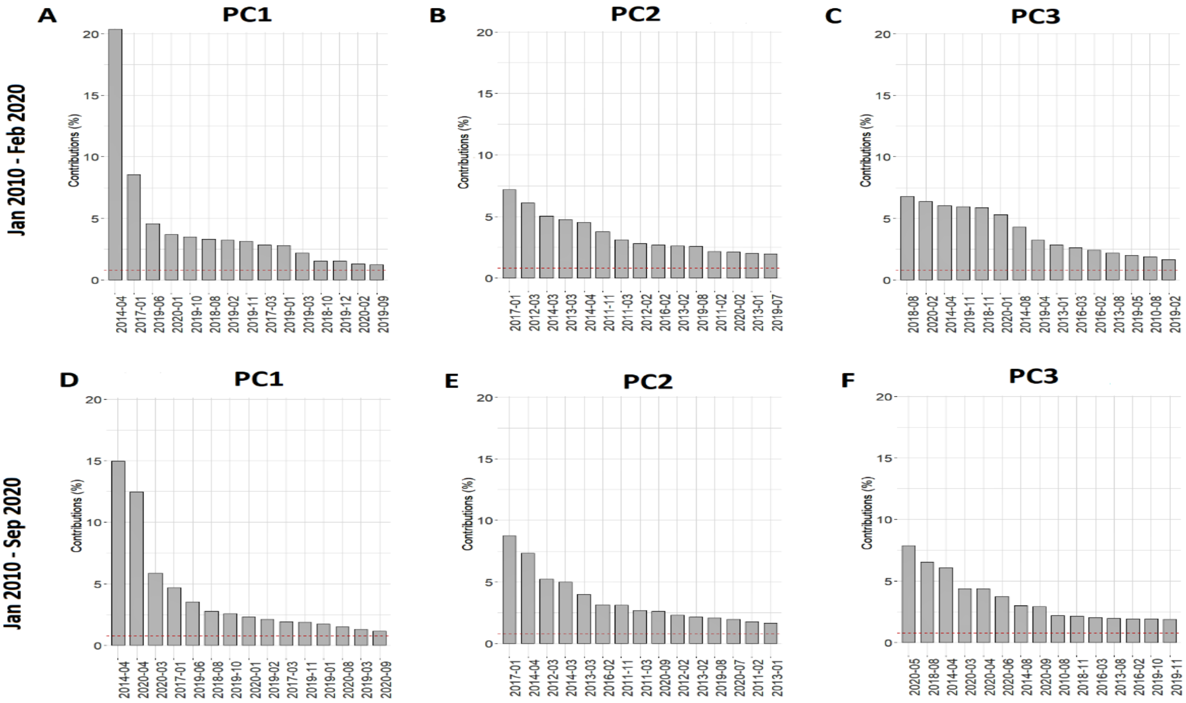

Figure 4: Percent contributions of the 15 most significant dates by month and year to the first three principal components. A-C: principal components retrieved for the PCA on January 2010 to February 2020 data. D-F: principal components retrieved for the PCA on January 2010 to September 2020 data. If a variable had a percent contribution greater than the level shown by a red dotted line, then that contribution was deemed significant.

Discussion

Our analysis of Google Trends search data for air pollution terms aligned with BBC headline data as a proxy for media coverage reveals that an increase in public interest in air pollution correlates with an increase in the frequency of news headlines. The increase in the volume of Google searches from around 2014 likely reflects a combination of contributing factors that led to greater media coverage. Firstly, in March 2014, record pollution levels hit the UK (BBC News, 2014) with much media coverage coinciding with the huge spike in Google searches for ‘air pollution’ that month. This was attributed in the media to a combination of particulates from the continent, the fine weather conditions and Saharan dust. Similar conditions and reporting happened in April 2015 (BBC News, 2015) as evidenced in Figure 1. Secondly, media coverage of illegal levels of NO_2 has increased since 2015 following continued successful action against the UK and Welsh Governments by ClientEarth and debate over Clean Air Zones in towns and cities across the UK. PCA suggested though that, despite media coverage of illegal levels of NO_2 over the years there was greater interest in $PM_{2.5}$. This could be explained by public interest increasing strongly at times of air pollution episodes which generally reflect increased levels of $PM_{2.5}$ as shown by PCA. PCA also revealed that the higher media coverage of NO_2 reductions during COVID-19 lockdown increased public interest.

When comparing Google searches for ‘air pollution’ with monthly distributions of pollutants we did observe a correlation in how the variables vary. This suggests that the public are aware of general seasonal increases and decreases of air pollution and could reflect interest by people with medical conditions impacted by $PM_{2.5}$ and NO_2. PCA did show though that the pattern of association between BBC headlines and Google searches is far greater suggesting that public interest is much increased when the media reports events. This result also highlights that the public are less aware of actual air pollution levels or the air quality index (AQI) provided by Defra. This is supported by a lower frequency of Google searches for 'air quality index' relative to 'air pollution' or 'air quality' It is unsurprising that higher air pollution episodes due to transboundary events such as Saharan dust attract media attention but another major contributor to $PM_{2.5}$ concentrations is agricultural ammonia (UK Government, 2019), which does not get reported as much and therefore the public are largely unaware.

Conclusion

Our results show that public interest in air pollution has increased gradually over the last decade which correlates with increased media coverage. Public interest does correlate with changing levels of air pollution over the year but, using PCA, we observed that interest has a greater association with air pollution episodes covered by the media. This shows that to change human behaviour in this topic, there needs to be greater information provided on a regular basis.

Impact

This research shows that there is value in ensuring a constant presence of the topic of Air Quality within the media. This not only educates the general public in the first instance, put acts as a catalyst for people to search for further information to make themselves more aware. This lends us to believe that the next step is to be able to provide high quality data and information for decision makers and the public at large to utilise in their life choices.

References

AQEG Report. (2020). Estimation of changes in air pollution emissions, concentrations and exposure during the COVID-19 outbreak in the UK. Retrieved from https://uk-air.defra.gov.uk/library/reports.php?report_id=1005

BBC News. (2014). UK air pollution: How bad is it? Retrieved from https://www.bbc.co.uk/news/uk-26851399

BBC News. (2015). High pollution hits southern England. Retrieved from https://www.bbc.co.uk/news/uk-32233922

BBC News. (2017). How bad is air pollution in the UK? Retrieved from https://www.bbc.co.uk/news/science-environment-38979754

BBC News. (2020a). Air pollution may harm cognitive intelligence, study says. Retrieved from https://www.bbc.co.uk/news/health-45326598

BBC News. (2020b). Low levels of air pollution linked to changes in the heart. Retrieved from https://www.bbc.co.uk/news/health-45034972

Carslaw, D. C. and K. Ropkins, (2012) openair - an R package for air quality data analysis. Environmental Modelling & Software. Volume 27-28, 52-61.

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. Economic record, 88, 2-9.

Committee on the Medical Effects of Air Pollutants (COMEAP). (2009). Long-term exposure to air pollution: effect on mortality. Available from:

www.gov.uk/government/uploads/system/uploads/attachment_data/file/304667/COMEAP_long_term_exposure_to_air_pollution.pdf

Cori, L., Donzelli, G., Gorini, F., Bianchi, F., & Curzio, O. (2020). Risk Perception of Air Pollution: A Systematic Review Focused on Particulate Matter Exposure. International Journal of Environmental Research and Public Health, 17(17), 6424.

Dong, D., Xu, X., Xu, W., & Xie, J. (2019). The Relationship Between the Actual Level of Air Pollution and Residents' Concern about Air Pollution: Evidence from Shanghai, China. International journal of environmental research and public health, 16(23), 4784.

Google Trends. 2015. Google Trends. Available at: https://www.google.com/trends/

Knipe, D., Evans, H., Sinyor, M., Niederkrotenthaler, T., Gunnell, D., & John, A. (2020). Tracking online searches for emotional wellbeing concerns and coping strategies in the UK during the COVID-19 pandemic: a Google Trends analysis. Wellcome Open Research, 5(220), 220.

Misra, P., & Takeuchi, W. (2020). Assessing Population Sensitivity to Urban Air Pollution Using Google Trends and Remote Sensing Datasets. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 42, 93-100.

Papadimitriou E., Lassarre S., Yannis G. Human factors of pedestrian walking and crossing behaviour. Transportation Research Procedia, 25 (2017), pp. 2002-2015

Philippe Massicotte and Dirk Eddelbuettel (2020). gtrendsR: Perform and Display Google Trends Queries. R package version 1.4.7. https://CRAN.R-project.org/package=gtrendsR

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Royal College of Physicians (RCP). (2016). Every breath we take: the lifelong impact of air pollution. Available from: www.rcplondon.ac.uk/file/2914/download?token=qjVXtDGo

Sebastien Le, Julie Josse, Francois Husson (2008). FactoMineR: An R Package for Multivariate Analysis. Journal of Statistical Software, 25(1), 1-18. 10.18637/jss.v025.i01

Simionescu, M., Streimikiene, D., & Strielkowski, W. (2020). What Does Google Trends Tell Us about the Impact of Brexit on the Unemployment Rate in the UK?. Sustainability, 12(3), 1011.

The Guardian. (2018). Air pollution: UK government loses third court case as plans ruled 'unlawful'. Retrieved from https://www.theguardian.com/environment/2018/feb/21/high-court-rules-uk-air-pollution-plans-unlawful

Turner M. and Struthers R. (2018). Public attitudes to Air Quality: Report for Defra. Retrieved at http://randd.defra.gov.uk/Document.aspx?Document=14242_DefraBMGResearch-PublicAttitudestoAirQuality-SummaryReport-May2018FINAL.pdf.

UK Government. (2019). Clean Air Strategy 2019. Retrieved from https://www.gov.uk/government/publications/clean-air-strategy-2019

World Health Organization. (2017). Global Health Observatory data. Retrieved from http://www.who.int/gho/phe/outdoor_air_pollution/burden/en/

World Health Organization. (2018). Ambient (outdoor) air pollution. Retrieved from https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health