Office for National Statistics

Coronavirus (COVID-19) related mortality rates and the effects of air pollution in England

Research report examining the potential relationship between long-term air pollution exposure and coronavirus (COVID-19) mortality rates.

Table of contents

1 . Main points

- The effects of long-term exposure to air pollution as a factor that increases coronavirus (COVID-19) mortality appear smaller than those reported in previous studies -- though our upper-bounded estimates are similar in magnitude to some studies.

- The estimated correlation (in models without further controls "raw") between air pollution and age-adjusted COVID-19 mortality rates when calculated using deaths earlier in the pandemic was higher than that found later on (including later deaths) as the disease spread more widely.

- Re-analysing the raw correlation with each new week of mortality data showed that the correlation fell rapidly and then stabilised but at a similar rate to the death rate change; it is therefore not clear whether the remaining air pollution effect shows an independent causal connection or reflects other factors such as where the infection reached before lockdown took effect.

- Further modelling was carried out including (where they improved the model fit) controls for sex, ethnicity, Indices of Multiple Deprivation (IMDs), smoking rates, cardiovascular co-morbidities for COVID-19, "other" co-morbidities for COVID-19, and population density.

- There is significant collinearity between ethnicity and air pollution, making it impossible to entirely separate the effects of these covariates with the confounding variables for which data are available; if there is a causal link between air pollution and COVID-19-related mortality, it would partially explain the disparities in COVID-19 outcomes for minority ethnic groups.

- For long-term exposure to fine particulate matter ($PM_{2.5}$), we estimated odds ratios for a 1 µg m^{-3} change in long-term average exposure of between 1.01 (statistically insignificant) and 1.07 (when ethnicity is removed from the model entirely).

- For NO_2, we estimated odds ratios for a 1 µg m^{-3} change in long-term average exposure of between 1.006 (statistically insignificant) and 1.02 (when ethnicity is removed from the model entirely).

- This analysis indicated that air pollution was unlikely to be the sole driver of disparities in mortality statistics for minority ethnic groups and as such, the scale of correlation found when ethnicity is not controlled for is likely to be an overestimate of the air pollution effect.

- A similar trend but with a negative correlation with COVID-19 mortality was found for ozone exposure: in the absence of any known reason for why ozone would provide a protective effect, a more likely explanation is that exposure to higher ozone is acting as proxy for living in the rural environment; this provides some further evidence that at least some of the correlation we can see is driven by infection rates rather than an underlying causal relationship between air pollution exposure and COVID-19-related mortality.

2 . Executive summary

The Office for National Statistics (ONS) was asked by the Scientific Advisory Group for Emergencies (SAGE) to take the lead in investigating UK data for any correlations between common air pollutants that are known to impact respiratory and cardiovascular health and rates of coronavirus (COVID-19) related mortality. This was in response to some initial studies from the US and Italy that suggested a significant positive correlation between $PM_{2.5}$ and NO_2 exposure and COVID-19 mortality rates. The work was to be carried out at a population scale, rather than individual, and based on existing data sources.

Literature

Previous studies have proposed a relationship between air pollution and COVID-19-related mortality. Wu et al. (2020) reported that based on particulate matter concentrations derived from satellite aerosol optical depth measurements, "long-term exposure to $PM_{2.5}$ is positively associated with increased COVID-19 mortality". Wu et al. specifically reported that a 1 µg m^{-3} increase in average $PM_{2.5}$ exposure would lead to an 8% increase in the baseline death rate. Conticini et al. (2020) used ambient ground-level air pollution data from air quality monitoring sites in Italy to provide "evidence that people living in an area with high levels of pollutant are more prone to develop chronic respiratory conditions and suitable to any infective agent." Travaglio et al. (2020) worked at regional and individual scales in England in April 2020, finding "an association between a 1 µg m^{-3} increase in sulphur dioxide and nitrogen oxide levels with a 17% and approximately 2% increase in COVID-19 mortality, respectively." Cole et al. (2020), the most recent paper, based on data from the Netherlands, found that a 1 µgm^{-3} increase in $PM_{2.5}$ exposure would increase the baseline death rate by between 13% and 21.4%.

Air pollution and public health

This study looks at the relationship between COVID-19 mortality and air quality using English datasets. Three major air pollutants that form part of the EC Ambient Air Quality Directive (2008/50/EC) are included as variables in this study. These are: $PM_{2.5}$ (an operationally defined metric for fine particulate matter with an aerodynamic diameter smaller than 2.5 microns), nitrogen dioxide (NO_2) and ozone (O_3). While there are a very wide range of different air pollutants that are known to be harmful to health, $PM_{2.5}$, NO_2 and O_3 are the most abundant and relevant in the context of COVID-19. These have well-established negative effects on respiratory and cardiovascular health. They are also linked to adverse outcomes in neurodevelopment, cognitive function and other chronic diseases such as diabetes. The effects of exposure to each air pollutant were reviewed in detail by the World Health Organization (WHO) in 2013. Air pollution can negatively affect human health through short-term (days to weeks) transitory exposure and long-term accumulated exposure (over years to decades), with the latter considered to cause the greater harm, according to a study by Pope (2008).

The geographic distribution of the three pollutants across the UK is different in each case, reflecting their emissions sources and atmospheric lifetimes. NO_2 is predominantly an urban air pollutant with highest concentrations found in city centres and at the roadside and with a dominant source from vehicle exhaust. It has a 1/*e*-folding atmospheric lifetime of around one hour, and so lower concentrations are found in suburban areas and the rural environment.

Ozone is a secondary pollutant that is formed from photochemical reactions. Ozone reacts rapidly with nitric oxide, a component of combustion exhaust, and this leads to its suppression in urban centres and near roads. The highest ambient concentrations, and hence possible exposure to ozone, occurs in the rural environment in the UK.

$PM_{2.5}$ has a complex range of sources. It is emitted directly from processes such as combustion and friction and is also formed as a secondary pollutant [4]. It has an atmospheric lifetime of around one to two days and concentrations in the UK, and particularly Southern England, can be influenced by transboundary transport of pollution from mainland Europe. While the highest concentrations of $PM_{2.5}$ are found typically in city centres, concentrations reduce more gradually moving from urban to rural environments, leading to a relatively narrow range of annual average concentrations and exposures, when compared to NO_2 and O_3.

The challenge

All of the analyses face the basic challenge of not having a reliable figure for levels of infection in the population across the full period of infection and at a sufficiently granular spatial level. Early in the pandemic, we would expect infection rates to be highest in cities with global travel connections and with high population densities that may lead to greater contagion rates. These are also geographic locations that have higher concentrations of $PM_{2.5}$ and NO_2 air pollution. It has also become clear over the course of the pandemic that [socioeconomic and demographic factors](#) are strongly associated with COVID-19 mortality rates, and these are also associated with higher long-term exposure to $PM_{2.5}$ and NO_2. It is therefore challenging to tease apart a correlation between air pollution and COVID-19 and geographically other co-located factors that can also influence mortality.

Confounding variables such as deprivation, existing illnesses and ethnic minority groups are all also correlated with one another and with air pollution concentrations. These "collinearities" further reduce the capacity of standard statistical analyses to find clear correlations at the scale these studies work at. If the correlation is very strong, it may be possible to find it, and if it does not exist, it may be possible to present some evidence of a null outcome. However, it is very difficult to produce clear definitive conclusions using the data currently available and this type of analysis, and it must be accepted that the true picture will likely only emerge once data are available for highly detailed individual-based modelling. While we wait for the opportunity to undertake more granular level examination of the data, there is potential to undertake further sensitivity analyses on the same basis as this study to test its robustness.

Our approach

The analysis in this article takes a particular statistic approach to examining the relationship between air pollution and COVID-19 in an attempt to overcome some of the issues of collinearity and varying rates of infection. This is achieved by breaking the country up into sample areas based on the variables of interest rather than census or governance-based geographies. In this way, a portion of London may be in the same sample population as another part of Newcastle if it shares the same salient characteristics. An assumption in this approach is that an increment or decrement in a pollutant such as $PM_{2.5}$ will have the same health effect wherever it occurs in the country. As such, differing rates of spread of the infection are then also distributed more widely.

This approach enables us to examine how our conclusions would have looked if we had used death rates at earlier stages of the pandemic and see how any correlation with air pollution has changed over time. This reflective approach -- rather than directly controlling for infection rate -- enabled us to see the direction of travel of any correlation as the infection spread more widely across the country. If the correlation was increasing or very stable with infection, that might be a clear indication we would see a very strong correlation were the infection to spread uniformly. A declining correlation as deaths increased may indicate the "real" correlation is smaller than measured or perhaps non-existent and that an early association between air pollution and COVID-19 mortality was linked to an initial outbreak of disease in large urban centres.

The results

The results suggest that $PM_{2.5}$ and NO_2 may correlate with increased mortality rates from COVID-19 infection but that the scale of impact may be smaller than that reported in earlier papers. Most importantly, once controlling for ethnicity as a confounding variable, this reduces the significance of correlation between $PM_{2.5}$ and NO_2 and COVID-19 mortality. This suggests that either $PM_{2.5}$ and NO_2 are drivers of disproportionate outcomes for minority ethnic groups or that $PM_{2.5}$ and NO_2 only show up as correlates because of the strong relationship between populations of minority ethnic groups and areas of high exposure to $PM_{2.5}$ and NO_2.

In addition, the calculated correlation between deaths and air pollution (for $PM_{2.5}$ and NO_x) was in fact falling rapidly before the lockdown and continued to fall as deaths rose before levelling out around Week 19 (week ending 8 May) of 2020. Conversely, we find that ozone exposure mirrored the correlations for $PM_{2.5}$ and NO_2 with a strongly negative correlation to COVID-19 that fell over time. We believe there is no good reason to believe that long-term exposure to higher ozone would provide a substantial protective effect. Instead, this negative correlation is more likely indicate that higher ozone is acting as a proxy for living in the rural environment -- with a potentially lower infection rate. (We note that higher ambient ozone concentrations are plausibly a factor that could reduce the viable airborne lifetime of the SARS-Cov-2 virus, but this analysis examines only the effects of cumulative exposure over 3, 5 and 10 years.)

It is therefore possible that the relationships we can see represent a snapshot of where the infection reached in the country. If that is the case, then air pollution correlations with COVID-19 may have continued to fall further if the infection had moved more uniformly across the nation.

3 . Method

Sampling and data linkage

We took a novel approach to sampling to mitigate complexities with respect to:

- varying rates of infection

- geographic collinearity of explained variables

- multi-collinearity of explanatory variables

Instead of using census or governance-based geographic boundaries as the basic unit (the approach used elsewhere and more traditionally in air pollution literature), we grouped geographic areas into treatment groups. These treatment groups were chosen based on: Indices of Multiple Deprivation (IMDs), population density and average $PM_{2.5}$ exposure over five years. The basic geographic unit built from was individual residential postcodes. The mortality and health care data could be linked at postcode level, but all other data had to be linked to from larger geographic levels.

National annual average air pollution concentration data are produced at the Ordnance Survey 1 km grid square level while the majority of confounding data are produced at the Lower-layer Super Output Area (LSOA) level. Whatever geography was chosen as the basic unit, there would be some degree of imprecision in the linking and it would depend upon assumptions. Air pollution concentrations were averaged across the 1 km grid square already, something that introduces considerable smoothing to the distribution of urban NO_2. This can average within the same 1 km grid square high roadside concentrations with lower concentrations away from major roads.

This sub-grid smoothing effect is, however, less pronounced for $PM_{2.5}$. 1 km air pollution data could be linked directly to the postcode. This remains a significant but unavoidable assumption since actual exposure of individuals will vary significantly even within a grid square. All other variables were linked to the postcode level and aggregated into sample groups based on a weighting of the total number of residential postcodes in the LSOA (or, in the case of smoking rates, local authority). This linkage is imprecise at the level of the postcode but once those postcodes are aggregated into sample groups, we consider that this level of imprecision is unlikely to significantly affect the analysis.

To create the sample groups concentrations of $PM_{2.5}$, five-year averages were ranked and broken into seven groups -- equally split by concentration -- of 1 km grid squares in England. The $PM_{2.5}$ and NO_2 are so strongly correlated (and O_3 chemically anti-correlated with NO_2) that it was not considered necessary to rebuild the frame for each pollutant individually. Those seven groups were each split up into quintiles by IMD score (less the environmental aspect of the IMD scale, to prevent double counting of air pollution effect). Those 35 groups were then each split into quintiles by population density to create 175 sample areas.

This approach combines areas across the country to mitigate to some degree the varying spread and rates of infection. We also avoid geographic collinearities in the explained variable, removing the need for weighted geographic approaches. A final analytical feature of this approach is that a much smaller proportion of the sample would lack deaths related to COVID-19 and have to be excluded at any time period through the pandemic. This enabled us to examine how any analysis of correlations between air pollution and mortality might have evolved through the pandemic.

The sampling approach also avoids large numbers of poorer, more polluted parts of the country being represented in the raw data, which could squeeze out contrast with polluted but wealthier or less densely populated neighbourhoods (notably in South-East England). It is easiest to imagine that we have filed different areas into different treatment groups for air pollution, deprivation and population density for which the statistical unit would be the treatment group.

More about coronavirus
- Find the latest on coronavirus (COVID-19) in the UK.
- All ONS analysis, summarised in our coronavirus roundup.
- View all coronavirus data.
- Find out how we are working safely in our studies and surveys.

Deaths data

Deaths were defined using the International Classification of Diseases, 10th edition (ICD-10). Deaths involving the coronavirus (COVID-19) include those with an underlying cause, or any mention, of ICD-10 codes U07.1 (COVID-19 virus identified) or U07.2 (COVID-19, virus not identified). The spatial linkage was based on the deceased's place of residence.

The analysis includes a total of 46,471 deaths involving COVID-19 among usual residents of England where the date of death was between 7 March 2020 and 12 June 2020, registered by 22 June 2020. The first death involving COVID-19 occurred on 2 March 2020, though analysis in the first week of deaths would not provide enough variance to be informative.

Age-adjusted death rates per 100,000 were calculated for each of the 175 sample areas using the standard approach. Age could be adjusted for as a covariate in the model, but the number of covariates required makes this costly in terms of explanatory power. Sex as a single variable is less costly in explanatory power and so was included later in the model development.

This sampling strategy will produce a fragmented geography but with the maximum possible variation in air pollution and the main socioeconomic variables. This work will need to be done for both NO_2 and $PM_{2.5}$ and so one, both or a combination could be used to create the sample points. To begin with, we use $PM_{2.5}$.

Air pollution

$PM_{2.5}$, NO_2, NO_x (the combination of NO_2 and NO, pollutants that are co-emitted) and O_3 exposure were all included. The number of days on which the daily max 8-hour concentration is greater than 120 μg m-3 at a 1 km grid resolution since 2003 is used for O_3; annual average air pollution exposure data are available at 1 km grid square resolution since 2002 for $PM_{2.5}$ and 2001 for NO_2 and NO_x from the [UK Air Information Resource (AIR) website](). Please see the UK AIR website for more details on the methods involved in producing these data. The value used in the sample area was the average across all postcodes included expressed as a concentration in units of μg m^{-3}.

Adjusting for levels of infection

There are a range of possible metrics that could be used as a proxy for infection rate, but all are either potentially misleading or lack sufficient granularity. We have taken two steps to mitigate the impact of varying infection rates. First, the fractured geography used in our sampling technique will combine different parts of the country, meaning that regional variations in infections will, to some degree, be smoothed out. Secondly, we repeated the analysis taking multiple snapshots of the infection for each week from Week 11 (week ending 13 March) 2020 to as close to publication of this report as possible (Week 24, week ending 12 June 2020).

If any correlation with air pollution begins to fall quickly or even disappear as the infection moves out of urban centres, it would be a sign that it could be an ultimately relatively weak or non-existent correlation. This approach does not suffer the biases and uncertainties of taking date or first infection or raw test data; however, some care must be taken in conclusions based on this approach since it is indicative only and no rigid conclusion should be drawn.

Population density

Early discussions of this work identified population density as likely to be related to rate of infection. It is therefore included in the sampling approach and confounding variables set as a weak form of infection rate control. The data are taken directly from the latest Office for National Statistics (ONS) [population projections from mid-year 2018 at Lower-layer Super Output Area (LSOA) level]().

Socio-economic characteristics

We used [English Index of Multiple Deprivation (IMD)]() scores without the human environmental domain at output area level since this includes air pollution indices. This precluded the inclusion of Scotland or Wales since they use different measures of area deprivation. However, it is a more rounded measure of deprivation than, for example, income alone.

Public health

We do not have data on smoking rates at high spatial resolution. However, we do have [local authority-level smoking prevalence (CSV, 22KB)](). This represents the poorest linkage to the postcode-based sampling areas in this analysis with all other linkages at output area or postcode level.

Co-morbidities

Hospital visit rates for known co-morbidities were calculated from [NHS data]() in 2017 to 2018. These were split into cardiovascular and "other" co-morbidities to also examine known relationships of air pollutants on cardiovascular diseases. The conditions included were:

- Alzheimer disease

- asthma

- influenza and pneumonia

- other acute respiratory infections

- bronchiectasis

- cancer

- cardiovascular conditions (all: current or recent) -- ischaemic heart disease, angina, myocardial infarction; heart failure; stroke; and Atrial fibrillation

- chronic kidney disease including renal failure

- chronic liver disease including liver failure

- chronic obstructive pulmonary disease including respiratory failure

- dementia

- diabetes

- epilepsy

- hypertension

- inflammatory bowel disease

- neurological conditions motor neurone disease, Parkinson's disease and multiple sclerosis

- osteoarthritis

- osteoporosis

- rheumatoid arthritis

- serious mental illness

To align co-morbidities with the deaths data, these were adjusted for age in the same manner as the deaths data to create a hospital visit rate per 100,000 people. We intended not to reweight them by the age functions related to the specific diseases but simply to put them back into alignment with the death data. There was a very clear relationship between the unadjusted death and comorbidity data, which was lost when examining the age-adjusted deaths data.

Ethnicity

Ethnicity data from the 2011 Census were used to estimate percentages of each population in broad ethnic groups of:

- White

- Asian or Asian British

- Black, African, Caribbean or Black British

- Other ethnic group

- Mixed or multiple ethnic group

4 . Statistical approach

Statistical model choice

The explained variable chosen is a rate and not count data. Given that the rates of mortality from the coronavirus (COVID-19) are relatively low if the age-standardised mortality rates (ASMRs) appear normally distributed, it is reasonable to apply a standard Poisson-based linear regression. We instead took a standard approach to analysing a rate-based outcome in producing a logit transform [1] and carrying out a standard linear regression of the form:

$$\ln\left(\frac{P_g}{1 - P_g}\right) = F(air\ pollution + density + socioeconomics + health + healthcare)$$

where P_g is the individual level probability of having died from COVID-19 between Weeks 11 (week ending 13 March 2020) and 14 (week ending 3 April 2020) of the pandemic in England.

Model building

Exposure periods are, unsurprisingly, strongly colinear and so could not be included in models simultaneously. We instead began by choosing the exposure period -- from those available -- for each air pollutant with the strongest correlation with age-adjusted death rates. All exposure periods were put alone into a linear model with the logit-transformed cumulative deaths data by Week 24 (week ending 12 June) of 2020. The exposure period with the strongest p-value was chosen. The ethnicity percentages are also colinear, so the same approach was taken choosing a single ethnicity to include.

The first analysis carried out was to run regressions with each air pollutant against the cumulative deaths for each week from Week 11 of 2020 to Week 24 of 2020. This was used to examine how the raw correlation (if any) changed over time.

A model controlling for confounding effects was built (without air pollution data) based on the cumulative death data for Week 25 (week ending 19 June) of 2020. The best model was found by carrying out forward and backward stepwise regressions, and the model with the strongest actual individual consumption (AIC) from the two methods was chosen. Once the control model was chosen, we added the 10-year average exposure data for $PM_{2.5}$ and NO_2 individually to examine their effect. We then carried out sensitivity testing by removing each control variable in turn to examine the impact on air pollution correlations and vice versa.

Governance

This work was commissioned from the Office for National Statistics (ONS) by the Scientific Advisory Group for Emergencies (SAGE). The Chairs of both the Air Quality Expert Group (AQEQ) and the Committee on the Medical Effects of Air Pollutants (COMEAP) were both on a steering group alongside representatives from Public Health England (PHE) hosted by the Department for Environment, Food and Rural Affairs (Defra), which helped guide the direction of the analysis. COMEAP, as a committee, had reservations about the sampling approach and how it might differ from more traditional approaches using standard governance- or census-based geography. COMEAP suggested, at a minimum, that sensitivity analysis in which the numbers of groups in each metric used in sampling are changed and results are compared is carried out in future work. The ONS is releasing this work as a first indication of findings and will take guidance on the demand for future work from SAGE. Other work looking at the drivers of COVID-19-related mortality will continue.

Notes for: Statistical approach

1. For example, Warton, D.I. and Hui, F.K.C. (2011), The arcsine is asinine: the analysis of proportions in ecology. Ecology, 92: 3 to 10. doi: 10.1890/10-0340.1

5 . Results

Exposure period choice and initial exploration of the raw correlations

NO_x, NO_2 and $PM_{2.5}$ all had their strongest correlations with logit deaths for 10-year exposures. Ozone, on the other hand, showed a stronger correlation with a five-year exposure. Ozone was the only pollutant to show a negative correlation with logit deaths.

Table 1: Logit(age-standardised mortality rates (ASMR) for the coronavirus (COVID-19) in Week 21 2020) regressed on average air pollutant exposures over different periods of time

Pollutant	Exposure Period	Coefficient	P_value
NO	10 years	0.019	7.73E-07
NO	5 years	0.02	1.38E-06
NO	1 years	0.022	1.71E-06
Ozone	10 years	-0.061	7.11E-03
Ozone	5 years	-0.317	3.67E-06
Ozone	3 years	-0.22	3.79E-06
PM .	10 years	0.088	2.42E-04
PM .	5 years	0.091	4.28E-04
PM .	3 years	0.082	7.84E-04
PM .	1 years	0.087	3.48E-04
NO	10 years	0.034	5.80E-07
NO	5 years	0.036	7.92E-07
NO	3 years	0.036	1.13E-06
NO	1 years	0.038	8.07E-07

Source: Office for National Statistics – Coronavirus and the effect of air pollution on mortality in England

However, scatterplots of each pollutant at the chosen exposure period against death rates suggest no visible correlation for $PM_{2.5}$, a weak positive correlation for NO_2 and NO_x and a weak negative correlation for ozone.

Figure 1: Weak visual positive relationship between NO_2 and age-adjusted death rate

Scatterplot of age-adjusted death rate against average NO_2 concentration, England, Week 24 2020

Download this chart

.XLSX

Figure 2: Weak visual positive relationship between NO_x and age-adjusted death rate

Scatterplot of age-adjusted death rate against average NO_x concentrations, England, Week 24 2020

Download this chart

.XLSX

Figure 3: Weak visual negative relationship between ozone and age-adjusted death rate

Scatterplot of age-adjusted death rate against average ozone exposure, England, Week 24 2020

Download this chart

[.XLSX](#)

Figure 4: Weak visual positive relationship between PM$_{2.5}$ and age-adjusted death rate

Scatterplot of age-adjusted death rate against average PM$_{2.5}$ exposure, England, Week 24 2020

Download this chart

[.XLSX](#)

Figure 5: Age-adjusted death rate for the highest-level pollution is statistically different from all other groups, but the picture is more uncertain below that

Average age-adjusted death rate per 100,000 against PM2.5 exposure group

Figure 5: Age-adjusted death rate for the highest-level pollution is statistically different from all other groups, but the picture is more uncertain below that

Average age-adjusted death rate per 100,000 against PM2.5 exposure group
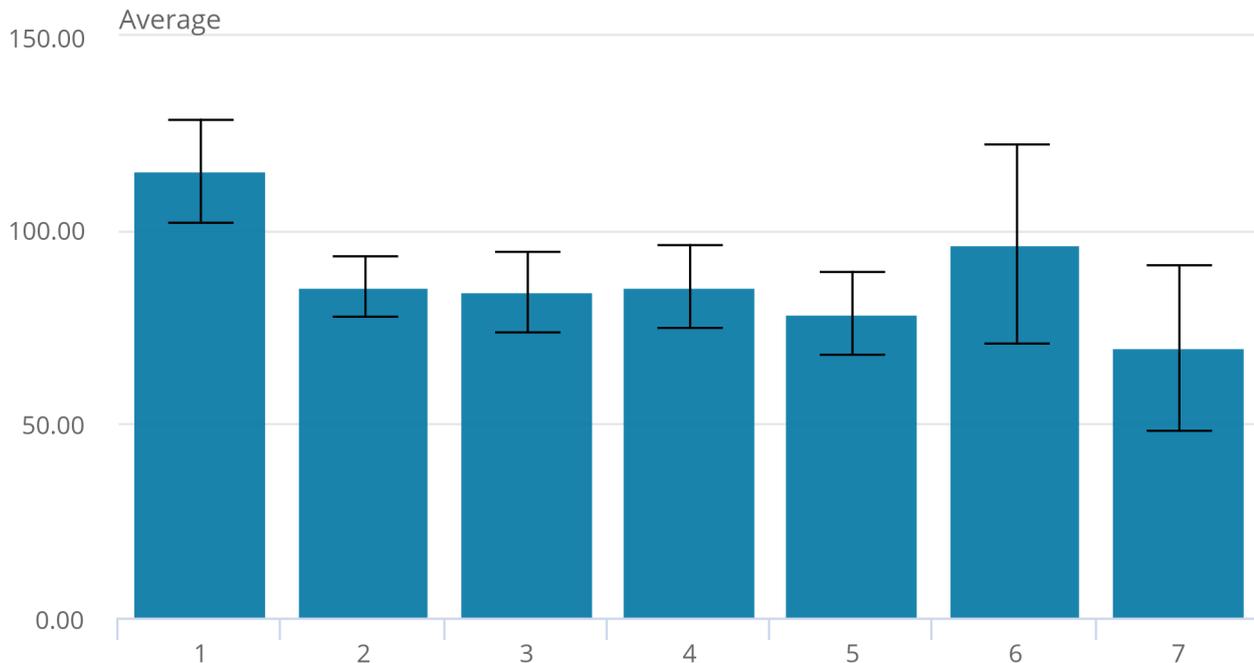
Source: Office for National Statistics – Coronavirus and the effect of air pollution on mortality in England

Figure 5 shows the average coronavirus (COVID-19) death rate by air pollution grouping. There is an apparent -- uncontrolled -- higher death rate in the highest air pollution group but no clear pattern among the remaining groups.

Ethnicity

The approach taken in this article is not well suited to teasing apart impacts on a specific ethnic group. There was significant correlation between the percentages of each ethnicity in the population with each other. The percentage of the White population had a coefficient of at least negative 0.94 for all other ethnicities. The lowest level of correlation was between the percentages of Black and Asian ethnic groups in a population (0.87). We therefore chose a single ethnicity to include in the control models based on its correlation with logit-adjusted deaths in Week 24 (week ending 12 June) 2020.

Table 2: Logit (age-standardised mortality rates (ASMR) for the coronavirus (COVID-19) in Week 24 2020) regressed on the percentage of the population from different (broad) ethnic groups

Ethnicity	Estimate	Lower	Upper	P
White	-2.27	-2.83	-1.71	2.20E-13
Mixed	17.84	12.65	23.03	1.90E-10
Asian	4.27	3.22	5.32	1.76E-13
Black	6.76	4.95	8.57	7.12E-12
Other	28.44	20.48	36.4	4.30E-11

Source: Office for National Statistics – Coronavirus and the effect of air pollution on mortality in England

Table 2 show that the percentage of the population of Asian ethnicity has the most significant correlation with logit-transformed and adjusted death rates in an otherwise uncontrolled model.

Weekly changes

We started by running a simple model of age-adjusted death rate against each pollutant with each model approach for every week since Week 11 (week ending 13 March) 2020. Figures 6 and 7 show that the correlations between $PM_{2.5}$, NO_2 and NO_x and COVID-19 mortality found early on in the infection were high, before falling rapidly and then appearing to level out; the negative binomial model shows similar outcomes.

Figure 6: The correlation between $PM_{2.5}$, NO_2 and NO_x and age-adjusted death rate fell from 15 March 2020 to early May as the total deaths increased

Changing weekly correlations between $PM_{2.5}$, NO_2 and NO_x and COVID-19 death rates based on logit model

Download this chart

[.XLSX](#)

Figure 7: The correlation between ozone and age-adjusted death rate is negative and increased from 15 March 2020 to early May as the total deaths increased

Changing weekly correlations between ozone and death rates based on logit model

Download this chart

[.XLSX](#)

At this stage, we decided to focus on $PM_{2.5}$ and NO_2. Ozone is negatively correlated with deaths and there is no reason to believe there is a negative causal relationship between ozone and COVID-19 mortality. NO_x is removed from further analysis since it is highly correlated with NO_2 (both in terms of concentrations and emissions sources) but with NO_2 having the slightly stronger correlation.

Figures 8 and 9 show how the rate of change in the correlation between $PM_{2.5}$ and death rates may have been affected by lockdown. The rate of change slows following lockdown as the rate of deaths begins to slow.

Figure 8: The rate of change in the correlation between exposure to $PM_{2.5}$ and age-adjusted death rates was fastest when the weekly death rate was highest and appeared to stabilise as the lockdown took hold

The correlation between exposure to $PM_{2.5}$ and logit death by week against total cumulative deaths with the lockdown shown as a dashed line in Week 13

Figure 8: The rate of change in the correlation between exposure to PM_{2.5} and age-adjusted death rates was fastest when the weekly death rate was highest and appeared to stabilise as the lockdown took hold

The correlation between exposure to PM_{2.5} and logit death by week against total cumulative deaths with the lockdown shown as a dashed line in Week 13
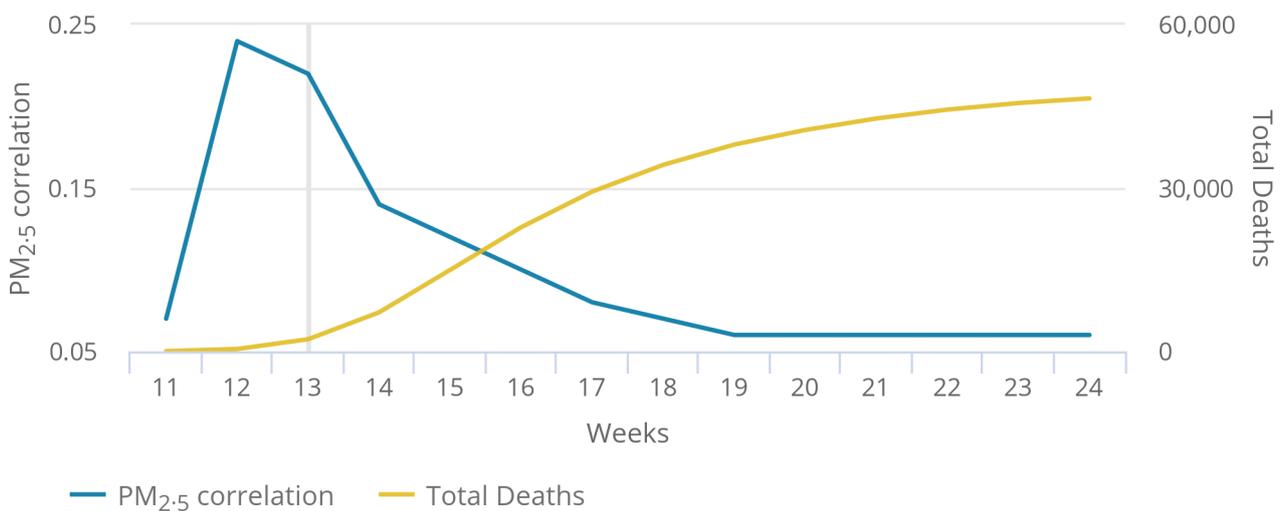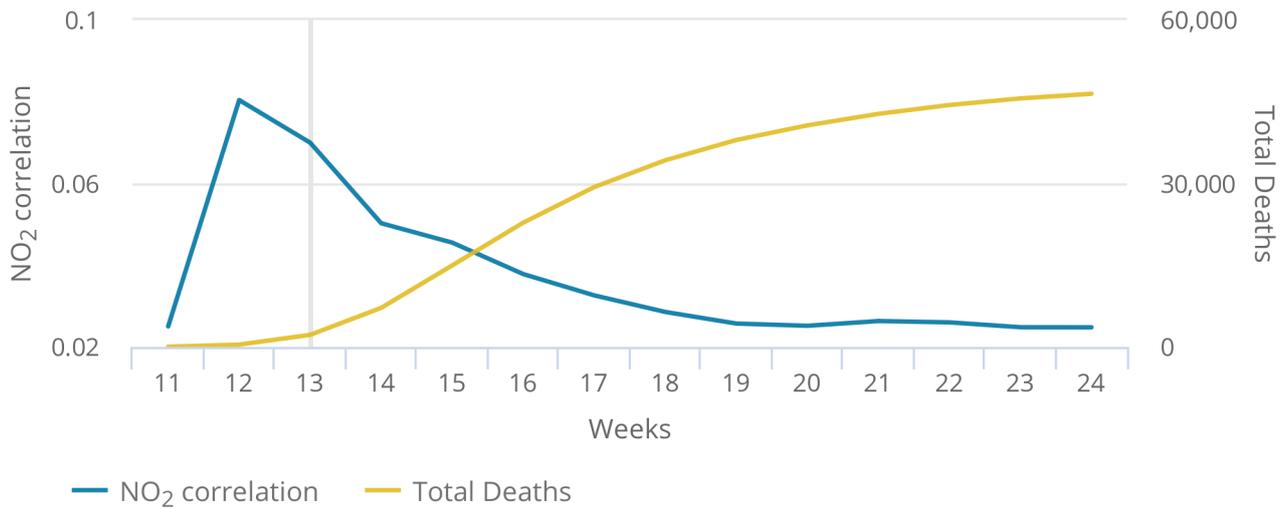

Source: Office for National Statistics – Coronavirus and the effect of air pollution on mortality in England

Figure 9: The rate of change in the correlation between exposure to NO$_2$ and age-adjusted death rates was fastest when the weekly death rate was highest and appeared to stabilise as the lockdown took hold

The correlation between exposure to NO$_2$ and logit death by week against total cumulative deaths with the lockdown shown as a dashed line in Week 13

Figure 9: The rate of change in the correlation between exposure to NO₂ and age-adjusted death rates was fastest when the weekly death rate was highest and appeared to stabilise as the lockdown took hold

The correlation between exposure to NO₂ and logit death by week against total cumulative deaths with the lockdown shown as a dashed line in Week 13

Source: Office for National Statistics – Coronavirus and the effect of air pollution on mortality in England

Air pollution impact with control variables

In this subsection, we build a model to control for confounding variables in a stepwise regression. The variables available for use were:

- hospital admission rate for cardiovascular COVID-19 co-morbidities

- hospital admission rate for all other COVID-19 co-morbidities

- Index of Multiple Deprivation (IMD) score (excluding the environmental domain)

- the percentage of the population who are female

- the percentage of the population of Asian ethnicity

- population density

- the estimated percentage of smokers in the population

Both stepwise approaches chose the following variables with an actual individual consumption (AIC) score of 110 (abbreviations used in tables in brackets):

- hospital admission rate for cardiovascular COVID-19 co-morbidities (cardio comorbidities)

- hospital admission rate for all other COVID-19 co-morbidities (other comorbidities)

- the percentage of the population of Asian ethnicity (Asian population)

- the estimated percentage of smokers in the population (smokers)

All but five of the 175 sample areas experienced some level of COVID-19-related mortality by Week 14 (week ending 3 April 2020), and so 170 were included in the analysis.

Tables 3 and 4 show the control model with the addition of $PM_{2.5}$ and NO_2 in turn.

Table 3: Model controlling for chosen confounding variables alongside 10-year average PM. exposure

	Estimate	Lower	Upper	P
(Intercept)	-8.74	-9.53	-7.95	<2e-16
Cardio comorbidities	7.65	1.13	14.18	0.02
Other comorbidities	0.03	0	0.05	0.03
Asian Population	2.86	1.12	4.6	0
Smokers	6.2	0.39	12	0.04
PM.	0.01	-0.03	0.06	0.56

Source: Office for National Statistics – Coronavirus and the effect of air pollution on mortality in England

Table 4: Model controlling for chosen confounding variables alongside 10-year average NO exposure

	Estimate	Lower	Upper	P
(Intercept)	-8.679	-9.302	-8.055	<2e-16
Cardio comorbidities	7.773	1.31	14.235	0.02
Other comorbidities	0.029	0.004	0.055	0.024
Asian Population	2.612	0.733	4.492	0.007
Smokers	5.967	0.193	11.741	0.044
NO	0.006	-0.009	0.022	0.413

Source: Office for National Statistics – Coronavirus and the effect of air pollution on mortality in England

Sensitivity testing

Tables 5 and 6 show that both $PM_{2.5}$ and NO_2 are affected by the removal of all other variables. However, the removal of comorbidities from the model (individually) further decreases the size and significance of the correlation with each air pollutant. The removal of the proportion of the population that are of Asian ethnicity from the model significantly increases both the estimated effect size and improves estimated significance. Interestingly, the correlation of $PM_{2.5}$ with deaths shifts to become negative when comorbidities are removed from the model. [1]

Table 5: Sensitivity testing on the effect on the correlation between PM. and death rate of removing covariates

Variable removed	Coefficient	P-value
None	0.013	0.56
Asian Population	0.072	2.41E-07
Cardio comorbidities	-0.001	0.97
Other comorbidities	-0.001	0.97
Smokers	0.008	0.71

Source: Office for National Statistics – Coronavirus and the effect of air pollution on mortality in England

Table 6: Sensitivity testing on the effect on the correlation between NO and death rate of removing covariates

Variable removed	Coefficient	P-value
None	0.006	0.41
Asian Population	0.024	4.84E-08
Cardio comorbidities	0.002	0.79
Other comorbidities	0.002	0.76
Smokers	0.029	0.02

Source: Office for National Statistics – Coronavirus and the effect of air pollution on mortality in England

Ethnicity and air pollution

Figures 10 and 11 show clear correlations between ethnicity and air pollution. Exposure to the pollutants NO_2 and $PM_{2.5}$ correlate with the percentage of the population that is Asian by 0.82 and 0.75 respectively, indicating very high collinearity.

Figure 10: Scatterplot of the proportion of the population that is BAME against average 10 year NO_2 concentration

There is a strong positive visual correlation between ethnicity and concentrations of NO_2

Download this chart

Figure 11: Scatterplot of the proportion of the population that is BAME against average 10 year NO$_2$ concentration

There is a strong positive visual correlation between ethnicity and concentrations of PM$_{2.5}$

Download this chart

Table 7 shows that the removal and inclusion of air pollutants does affect the correlation of ethnicity with COVID-19 death rates but that ethnicity remains highly significant in all cases.

Table 7: Correlation of the proportion of the population who are Asian with death rates with and without PM. and NO covariates

Model	Pollutant included	Coefficient	P-value
Logit	none	3.28	1.62E-09
Logit	PM.	2.86	1.57E-03
Logit	NO	2.61	7.14E-03

Source: Office for National Statistics – Coronavirus and the effect of air pollution on mortality in England

Impacts on mortality risk

In this subsection, we assume that the correlations we have found are real and significant and attempt to estimate what this means for mortality risk from COVID-19. An odds ratio of one equates to no effect, while any movement away from one indicates a percentage change from the baseline death rate, not an absolute change in the percentage of deaths expected.

By Week 24 (week ending 12 June) 2020, the coefficient estimates (ignoring reversals in correlation) run at between 0.01 and 0.07 for PM$_{2.5}$ and 0.006 to 0.02 for NO$_2$ across all models, including those without controls. The exponential of the coefficient gives us the odds ratio for the logit models.

We therefore found that a 1 µg m^{-3} change in 10-year exposure to PM$_{2.5}$ had odds ratios for COVID-19 mortality of between 1.01 (statistically insignificant in the controlled model) and 1.07 (removing ethnicity controls). The higher estimate is similar to that found by Wu et al. (2020) (1.08); while, in the fully controlled model from this article, the estimate has an upper-bounded coefficient of 0.06, indicating a significantly different estimate. The model lacking any confounding control variables estimated a slightly stronger effect than Wu et al. with a coefficient of 0.88. Coefficients for NO$_2$ indicate that a 1 µg m^{-3} change in 10-year exposure had odds ratios for COVID-19 mortality between 1.006 (statistically insignificant) and 1.02 (where ethnicity is removed from the model).

Notes for: Results

1. The correlation between 10-year PM$_{2.5}$ exposure and age cardiovascular and "other" comorbidities is negative 0.06 and negative 0.22.

6 . Conclusion

Our analysis does not discount the possibility of a correlation between PM$_{2.5}$ exposure and coronavirus (COVID-19) related mortality of a similar scale to that found by Wu et al. (2020). However, there is evidence to indicate that if there is a causative correlation, it is likely to have a lower level of effect than our higher-end estimates.

Our analysis of the effect of air pollution on COVID-19 is highly sensitive to the time during the pandemic at which the analysis is performed, likely because of the progressive spread of the disease outwards from urban, more polluted regions. In the period when the death rate remained high, a weekly analysis (controlling only for age and no other confounding variables) produces a decreasing degree of correlation with time.

While the early weeks were affected by the incomplete spread of infection, the later weeks were affected by the complicated impacts of lockdown on infections, which may have slowed the rate of decline in the correlation (without controlling for any confounding variables beyond age). Note also that Figure 5, while a crude visualisation, gives some indication that most of the trend is driven by higher rates of COVID-19 deaths in the most highly polluted group but with no clear trend in areas of lower pollution. That again might indicate that PM$_{2.5}$ and NO$_2$ in urban areas are acting as a proxy for the higher rates of infection in cities or other factors associated with area deprivation.

The behaviour of ozone in the analysis provides a further sense check on the hypothesis that we are largely observing PM$_{2.5}$ and NO$_2$ acting as proxies for increasingly urban areas. The geographic distribution of NO$_2$ and O$_3$ across the UK are more or less a mirror image of one another. Urban centres have high NO$_2$ and low O$_3$, because of local fast reactions between NO and O$_3$. In the suburban and rural environments, NO$_2$ is typically low (because it has reacted away through oxidation) and O$_3$ is high (because it is the end product). The linkage between urban NO, NO$_2$ and O$_3$ is discussed in the context of COVID-19 in the Department for Environment, Food and Rural Affairs (Defra) report on air quality changes during lockdown. The ozone anti-correlation in the early periods around Weeks 12 (week ending 20 March 2020) to 13 (week ending 27 March 2020) implies some kind of protective effect. A more likely interpretation is that ozone concentration in this period is acting as a proxy for living in the rural environment.

We cannot fully disentangle the impacts of air pollution from other factors that may be driving disparities in outcome for minority ethnic groups because of the high level of correlation between the ethnicity and PM$_{2.5}$ and NO$_2$ variables. This may be because air pollution is a significant factor in those disparities, but we can be reasonably certain here that it is not the only factor driving ethnic disparities. The higher-end estimate for PM$_{2.5}$ is taken from a model in which ethnicity is not controlled for, and it is likely an overestimate of the real effect. PM$_{2.5}$ and NO$_2$ in the absence of ethnicity in the model will act as a proxy for other issues disproportionately affecting ethnic minorities and driving higher death rates.

Most importantly, our analysis indicates that caution should be used in interpreting all data without very strong infection rate controls. Individual-level analysis with large datasets and strong infection rate data would be better placed to investigate these impacts but would take longer to produce. In addition, individual-level modelling could examine more acute exposure around the time of infection.

We recognise that the sampling approach used here is novel in this area of research. We consider it was appropriate to apply this approach given the lack of alternatives to deal with the significant challenges this type of analysis presents in these circumstances. The Committee on the Medical Effects of Air Pollutants (COMEAP) had concerns regarding the novel approach to sampling in terms of its comparability to other research and how sensitive our findings might be to the ways in which we broke up the population.

An option for future work would be to use alternate groupings to test the sensitivity of results to alternative sample groupings. That work might involve breaking up the sample by proportions of ethnic minorities in the population to potentially help tease apart ethnicity and air pollution impacts. If possible, we would also include additional confounding variables in the control models such as housing types. However, we would not expect that work to necessarily provide a clearer picture of the relationship between air pollution and COVID-19-related mortality rates; that will require detailed individual-level analysis to fully disentangle confounding variables and infection rate issues. Work on air pollution at an individual level is ongoing within the Office for National Statistics (ONS) for London only and is likely to be a more fruitful methodological approach.