Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Original Research

# Quantifying the impact of air pollution on Covid-19 hospitalisation and death rates in Scotland

Duncan Lee [a,*], Chris Robertson [b,c], Carole McRae [c], Jessica Baker [c]

[a] *School of Mathematics and Statistics, University of Glasgow, Glasgow, G12 8SQ, Scotland, United Kingdom*
[b] *Department of Mathematics and Statistics, University of Strathclyde, Glasgow, G1 1XH, Scotland, United Kingdom*
[c] *Public Health Scotland, Meridian Court, 5 Cadogan Street, Glasgow G2 6QE, Scotland, United Kingdom*

## ARTICLE INFO

## ABSTRACT

Better understanding the risk factors that exacerbate Covid-19 symptoms and lead to worse health outcomes is vitally important in the public health fight against the virus. One such risk factor that is currently under investigation is air pollution concentrations, with some studies finding statistically significant effects while other studies have found no consistent associations. The aim of this paper is to add to this global evidence base on the potential association between air pollution concentrations and Covid-19 hospitalisations and deaths, by presenting the first study on this topic at the small-area scale in Scotland, United Kingdom. Our study is one of the most comprehensive to date in terms of its temporal coverage, as it includes all hospitalisations and deaths in Scotland between 1$st$ March 2020 and 31$st$ July 2021. We quantify the effects of air pollution on Covid-19 outcomes using a small-area spatial ecological study design, with inference using Bayesian hierarchical models that allow for the residual spatial correlation present in the data. A key advantage of our study is its extensive sensitivity analyses, which examines the robustness of the results to our modelling assumptions. We find clear evidence that $PM_{2.5}$ concentrations are associated with hospital admissions, with a 1 $\mu gm^{-3}$ increase in concentrations being associated with between a 7.4% and a 9.3% increase in hospitalisations. In addition, we find some evidence that $PM_{2.5}$ concentrations are associated with deaths, with a 1 $\mu gm^{-3}$ increase in concentrations being associated with between a 2.9% and a 10.3% increase in deaths.

## 1. Introduction

Covid-19 has caused worldwide health and economic devastation, and was declared a global pandemic by the World Health Organisation (WHO) on 11$th$ March 2020. The disease originated in Wuhan in the People's Republic of China in December 2019, and subsequently spread across the world in the following months. As of 1$st$ March 2022 there have been over 437 million cases worldwide, and over 5.9 million of those individuals have sadly died from the virus (Johns Hopkins Coronavirus Resource Centre, https://coronavirus.jhu.edu/map.html). There is thus a rapidly growing research literature focusing on the Covid-19 pandemic, including: (i) modelling the spread of the pandemic and its impacts on healthcare (Remuzzi and Remuzzi, 2020 and Lee et al., 2021); (ii) identifying the factors that make people at a higher risk of displaying severe symptoms (Conticini et al., 2020 and Wu et al., 2020); (iii) identifying the wider health impacts of the pandemic (Douglas et al., 2020); and (iv) developing real-time surveillance systems to track disease incidence (Dong et al., 2020).

One important epidemiological question about the pandemic is the role that air pollution may play in exacerbating Covid-19 symptoms, potentially leading to greater numbers of hospital admissions and deaths from the virus in more polluted areas. Air pollution is well known to be a global public health problem, with the WHO linking seven million deaths to it each year worldwide (World Health Organisation, 2016). One of the main health outcomes associated with air pollution is respiratory disease (COMEAP, 2020), and thus it is biologically plausible that air pollution may deleteriously affect Covid-19 sufferers, resulting in worse health outcomes. However, the evidence on this to date is somewhat mixed, with Wu et al. (2020), Coker et al. (2020), Berg et al. (2021) and Mendy et al. (2021) all finding significant effects of $PM_{2.5}$ on Covid-19 deaths and morbidities. In contrast, Konstantinoudis et al. (2021) and Sun et al. (2021) found no significant effects of $NO_2$ or $PM_{2.5}$ on Covid-19 deaths. Additionally, in a recent multi-country study (Huang et al., 2021) found significant

* Corresponding author.
*E-mail addresses:* Duncan.Lee@glasgow.ac.uk (D. Lee), chris.robertson@strath.ac.uk (C. Robertson), carole.mcrae@phs.scot (C. McRae), jessica.baker2@phs.scot (J. Baker).

**Table 1**
Summary of the spatial distribution of the data across the Intermediate Zones for the entire 17 month duration of the study. SIMD denotes the Scottish index of multiple deprivation.

| Variable | Percentile | | | | |
|---|---|---|---|---|---|
| | 0% | 25% | 50% | 75% | 100% |
| **Hospitalisations** | | | | | |
| Count | 0 | 12 | 21 | 31 | 104 |
| SMR | 0.00 | 0.54 | 0.93 | 1.41 | 3.66 |
| **Deaths** | | | | | |
| Count | 0 | 3 | 6 | 11 | 51 |
| SMR | 0.00 | 0.45 | 0.85 | 1.43 | 5.81 |
| **Air pollutants** | | | | | |
| $NO_2$ ($\mu g\ m^{-3}$) | 1.32 | 5.20 | 8.66 | 12.04 | 31.48 |
| $PM_{2.5}$ ($\mu g\ m^{-3}$) | 2.49 | 4.95 | 5.56 | 6.07 | 7.74 |
| $PM_{10}$ ($\mu g\ m^{-3}$) | 4.46 | 8.15 | 9.16 | 9.96 | 13.22 |
| **Covariates** | | | | | |
| Covid-19 case rates (per 100,000 during the study) | 0 | 3773 | 5647 | 7516 | 15,744 |
| % vaccinated (2 doses) by April 2021 | 6.6 | 25.1 | 30.2 | 35.8 | 61.8 |
| SIMD - employment domain | 0.5 | 5.1 | 8.3 | 12.6 | 30.9 |
| Carehome places (number) | 0 | 0 | 0 | 52 | 312 |
| Population density (people per hectare) | 0.0 | 3.6 | 27.1 | 39.7 | 236.5 |
| Black (%) | 0.0 | 0.1 | 0.2 | 0.6 | 26.6 |
| Chinese (%) | 0.0 | 0.1 | 0.3 | 0.6 | 11.9 |
| Indian/Pakistani/Bangladeshi (%) | 0.0 | 0.3 | 0.7 | 1.7 | 48.5 |

effects of $PM_{2.5}$ on Covid-19 incidence in the USA but not in Canada, England or Italy.

This paper adds to this limited global evidence base, by presenting a new study quantifying the effects of air pollution on Covid-19 outcomes in Scotland, United Kingdom, which as of 1*st* March 2022 has seen over 1,390,000 cases and over 10,670 deaths (Johns Hopkins Coronavirus Resource Centre) due to Covid-19. The impact of air pollution on health more generally is a continuing priority for the Scottish government, because it recently published its second Cleaner Air for Scotland (CAFS) strategy in July 2021 (Scottish Government, 2021). Thus understanding the role that air pollution may play on adverse health outcomes due to Covid-19 is an important topic for the Scottish Government, and as no study has yet examined this at the small-area scale, this is the evidence gap that this paper fills.

Our study is based on small-area count data summarising Covid-19 hospitalisations and deaths between March 2020 and July 2021, which thus makes it one of the most up-to-date and comprehensive studies in terms of its temporal duration (e.g. Wu et al., 2020 used data up to 18*th* June 2020 while Konstantinoudis et al., 2021 used data up to 30*th* June 2020). These small-area count data are modelled using Poisson log-linear models, where the spatial variation in Covid-19 risk is explained by air pollution concentrations, a range of other important confounding factors, and random effects allowing for residual spatial autocorrelation. When presenting the study results we conduct a wide-ranging sensitivity analysis to our modelling assumptions, thus quantifying the robustness of our results to model misspecification. The study region and data are presented in Section 2, while the statistical models we use are presented in Section 3. The results of the study are presented in Section 4, while Section 5 concludes the paper highlighting our key findings.

## 2. Data

The study region is Scotland, which for the purposes of this study has been partitioned into $K = 1279$ small areal units called Intermediate Zones (IZ) that have an average population of around 4000 people. IZs are a Scottish Government developed geography for the distribution of small-area statistics, and further details can be found at https://www.spatialdata.gov.scot/geonetwork/srv/api/records/389 787c0-697d-4824-9ca9-9ce8cb79d6f5. The time period for our study is from 1*st* March 2020 until 31*st* July 2021, a duration of 17 months. We use a spatial rather than a spatio-temporal ecological design for our study, where the data are aggregated over time to yield one summary measure for each variable in each IZ representing the entire study duration. We adopt this design because having a, say monthly, spatio-temporal design would result in around 1 hospitalisation and less than 1 death on average in each spatio-temporal unit (see Table 1), which would thus make regression effects hard to estimate due to the very small numbers of disease events in each spatio-temporal unit.

### 2.1. Covid-19 hospitalisation and death data

There are two outcome variables in this study, namely the numbers of Covid-19 hospitalisations and deaths that occurred from the population living in each IZ between 1*st* March 2020 and 31*st* July 2021. The hospitalisation data were obtained directly from Public Health Scotland, while the death data were obtained from the National Records for Scotland (https://www.nrscotland.gov.uk/covid19stats).

The numbers of Covid-19 hospitalisations and deaths in an IZ will depend on the size and age–sex demographics of its population, which we adjust for using indirect standardisation. Specifically, we compute the expected numbers of Covid-19 hospitalisation and deaths in each IZ, based on the assumption that Scotland-wide age–sex specific hospitalisation/death rates apply in each IZ. National age–sex specific Covid-19 hospitalisation/death rates for the study period by five year age group were obtained from the sources listed above, while age–sex specific population totals for 2019 (the closest year of data available) were obtained from the Scottish Government. The expected number of hospitalisations/deaths in each IZ is computed by multiplying the number of people in each age–sex group by the national Covid-19 hospitalisation or death rate for that group, before summing over groups to obtain the final expected count separately for hospitalisations and deaths.

An exploratory measure of Covid-19 risk in each IZ is the standardised morbidity/mortality (death) ratio (SMR), which is the observed number of hospitalisations/deaths divided by the expected numbers of hospitalisations/deaths. An SMR of 1 corresponds to an average risk area relative to the national hospitalisation/death rates, values less than one correspond to low risk areas, while values greater than one correspond to high risk areas. Specifically, a value of 1.2 corresponds to a 20% increased risk compared to the national average. Maps of the spatial distributions of the SMR are displayed in panels A (hospitalisation) and B (death) of Fig. 1, while a numerical summary of their quantiles across the IZs is given in Table 1. Both these displays highlight the large amount of noise in the SMR, with values ranging between (0, 3.66) for hospitalisations and (0, 5.81) for deaths. The increased
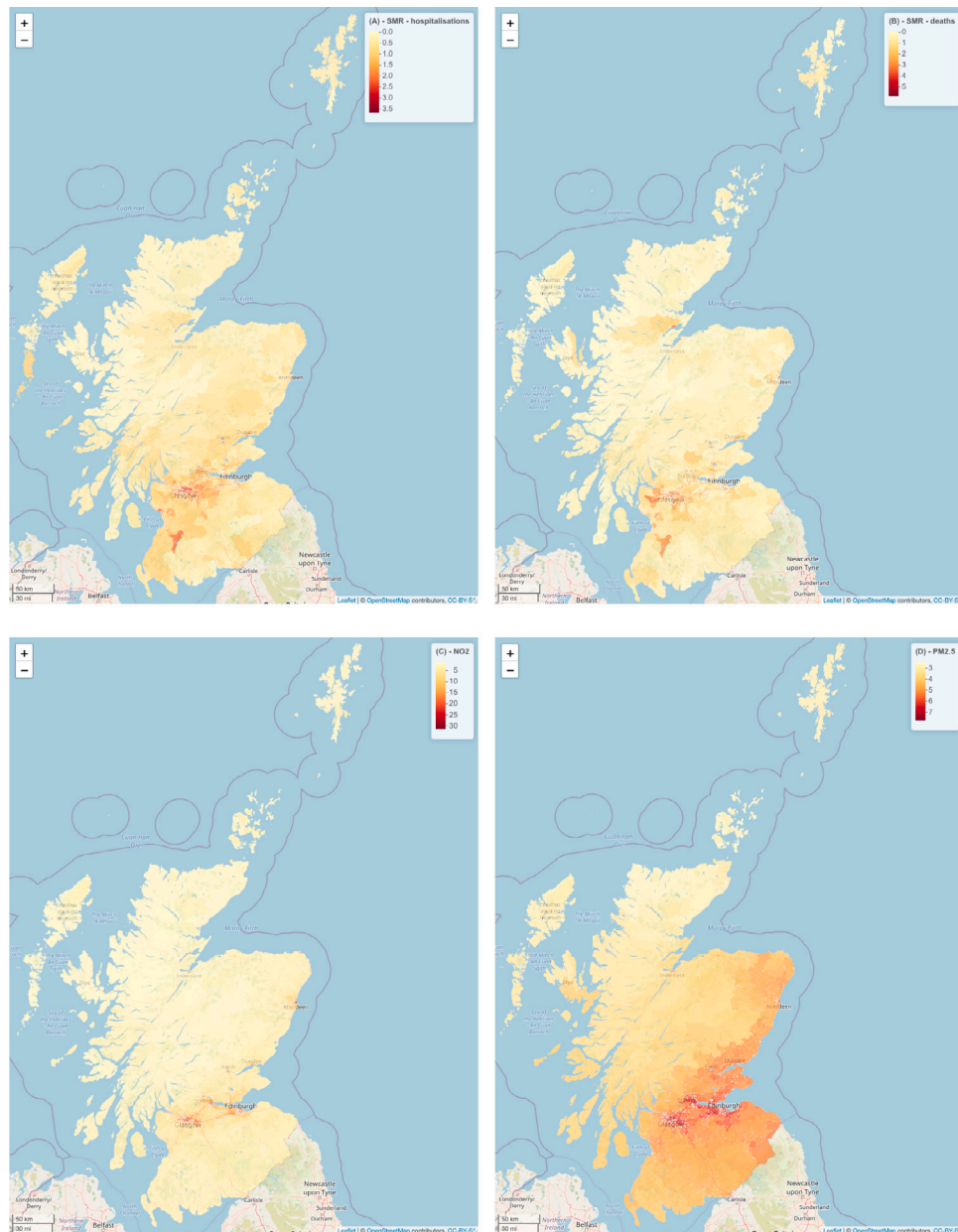
**Fig. 1.** Maps of the study data. Panels (A) and (B) show the SMR for Covid-19 hospitalisations and deaths respectively, while panels (C) and (D) depict $NO_2$ and $PM_{2.5}$ concentrations respectively.

amount of noise in the death SMRs is due to them being based on smaller numbers of events compared to the hospitalisation SMRs, with median sizes of 21 hospitalisations and 6 deaths in an IZ over the study period. The SMRs also show little in the way of a clear spatial trend across Scotland, which is also likely to be caused by the large amount of noise.

### 2.2. Air pollution data

The network of air pollution monitors in Scotland is spatially sparse (see http://www.scottishairquality.co.uk), and thus does not provide a measurement for each of the $K = 1279$ IZs in this study. Therefore, in common with Haining et al. (2010) and Lee et al. (2019) we utilise modelled concentrations instead, specifically annual averages for 2019 from the Pollution Climate Mapping (PCM) model (https://uk-air.defra.gov.uk/data/pcm-data) developed for the Department for the Environment, Food and Rural Affairs (DEFRA). We use concentrations for 2019 rather than for 2020, because the latter are artificially

lower than *normal* concentration levels. This is because lockdowns and other mobility restrictions were introduced in Scotland during 2020, which reduced transport usage and hence pollution concentrations for a sizeable proportion of that year. A full lockdown was implemented between 23*rd* March–29*th* May, while a range of other local restrictions occurred across Scotland between June and December. However, to assess the effect of using 2019 rather than 2020 concentrations we also fit models based on the 2020 concentrations as a sensitivity analysis.

The PCM model produces estimated concentrations on a 1 km$^2$ grid, which is spatially misaligned with the irregularly shaped IZs. The problem of spatial misalignment between multiple data sets is known as the *change of support problem* (Gelfand et al., 2001), and is a common challenge in spatial epidemiology. For example, Berrocal et al. (2010) and Forlani et al. (2020) proposed approaches for dealing with spatial misalignment between multiple air pollution data sets, specifically between point-level measurements and grid-based modelled estimates. In contrast, a range of approaches have been proposed for spatially re-aligning air pollution measurements to an irregular areal

unit geography, which is the spatial footprint that population-level disease data are typically available at. When the pollution data used in a study are point-level measurements then geostatistical Kriging is often used (e.g. Huang et al., 2021), while if modelled data with complete spatial coverage are used then simple averaging (e.g. Haining et al., 2010 and Lee et al., 2019) or area-based interpolation (Berg et al., 2021) are most commonly adopted.

In this study as the pollution estimates from the PCM model have complete spatial coverage of the study region, we use the simple averaging approach in common with the authors outlined above. Specifically, each 1 $km^2$ gridded concentration has an associated centroid (central point), and we estimate the pollution concentration in an IZ as the mean of the grid square concentrations whose centroids lie within the IZ. Any IZ that does not contain a grid square centroid is assigned the pollution concentration from the nearest grid square.

In this study we consider concentrations of nitrogen dioxide ($NO_2$) and both coarse ($PM_{10}$) and fine ($PM_{2.5}$) particulate matter, all of which are measured in $\mu g\ m^{-3}$. We choose these pollutants because they are the ones responsible for 35 out of the 36 air quality management areas in Scotland, for details see http://www.scottishairquality.scot/laqm/aqma. The spatial distributions of these pollutants are presented in Fig. 1 and Table 1, with the former displaying $NO_2$ (panel C) and $PM_{2.5}$ (panel D) concentrations. The figure and table show that high $NO_2$ concentrations are almost exclusively observed in the main urban centres of Aberdeen, Dundee, Edinburgh and Glasgow, with the remaining rural areas exhibiting very low concentrations. In contrast, $PM_{2.5}$ concentrations are more evenly distributed spatially, albeit with a clear east–west trend, with higher concentrations in the east being caused by transboundary pollution coming from continental Europe.

### 2.3. Covariate data

We have also collected data on a number of confounding factors that will likely influence Covid-19 hospitalisation and death rates, which are summarised in Table 1. The main confounding factor when modelling Covid-19 death rates is the numbers of Covid-19 cases, and data are available from Public Health Scotland (https://www.opendata.nhs.scot/dataset/covid-19-in-scotland) on the rate of cases per 100,000 people over the 17 month duration of the study. Additionally, we have data on Covid-19 vaccination rates, which will likely affect hospitalisation and death rates during the latter part of the study period. The first vaccinations were given in Scotland in December 2020, so while they could not affect the Covid-19 outcomes that occurred between March 2020 and December 2020, they may have had an affect after that time. However, it is not clear what an appropriate cutoff date should be for measuring vaccination rates in this study, because if one chooses an early date, such as 1st January 2021, then most areas will have a very small proportion of their population vaccinated by then and hence the variable will likely have little effect. In contrast, if one chooses a cutoff date of 31st July 2021 (the end of the study), then vaccination rates on this date are unlikely to have affected the hospitalisations/deaths that occurred during the study. Therefore we compute the percentage of the adult population (defined as people aged 20 or over because population counts are only available for 5-year age bands) who were either single vaccinated or double vaccinated by the end of February 2021 and April 2021, and determine which is the best predictor via a model building exercise.

Another important confounding factor in spatial epidemiological studies is socio-economic deprivation, with poorer populations tending to exhibit a greater burden of adverse health compared to more affluent areas. In Scotland in 2020 socio-economic deprivation is measured by the Scottish index of multiple deprivation (SIMD, https://simd.scot/), which is a composite index comprising indicators in the domains of geographical access to services, crime, education, employment, health, housing and income. However, as our outcome variables are health related we do not consider indicators from the health domain. Full

details of the individual indicators in the SIMD are given in Section 1 of the supplementary material accompanying this paper. The index and its indicators are constructed at the Data Zone (DZ) scale, which is a smaller small-area geography that nests exactly within the IZs used here. Thus each individual indicator is aggregated to the IZ scale by population weighted averaging. However, there are 9 incidences of missing values across these SIMD indicators, which are estimated by a spatially weighted average of the values in neighbouring areas (those that they share a common border with).

We also consider a number of other covariates for possible inclusion in the model, because existing studies have shown that they may affect rates of Covid-19 ill health or death. The first is population density computed as the number of people per hectare in each IZ, and areas with higher population densities are likely to have higher Covid-19 hospitalisation/death rates due to the likely increased mixing of the population as they live closer together. Secondly, care homes for elderly people are known to have been substantially impacted by Covid-19 ill health throughout the pandemic. Therefore we obtained data for each IZ on the number of registered places in care homes for older adults in November 2020 (the middle of the study period) from the Care Inspectorate (https://www.careinspectorate.com/). Additionally, we have data on ethnicity from the 2011 Scottish census, with more recent data not available because the census is only performed every ten years and the 2021 census was pushed back to 2022 due to Covid-19. For each IZ we have data on the percentage of people who are from the following ethnic groups taken from the census: (i) black; (ii) Chinese descent; (iii) Indian/Pakistani/Bangladeshi descent; and (iv) white.

### 2.4. Exploratory analysis

We first built simple Poisson log-linear regression models separately for the hospitalisation and death count data sets without allowing for spatial autocorrelation, as this enables us to determine whether the covariates capture all of the spatial autocorrelation present in the data. We began by including the expected number of hospitalisations/deaths as an offset term in each model, as well as adding in the covariates relating to Covid-19 case rates, population density and the number of carehome places. All of these covariates were significantly related to both the hospitalisation and death counts, and were hence retained in the model.

Next we considered how to account for socio-economic deprivation in the model, and after ignoring the health domain the SIMD comprises 20 indicators of socio-economic deprivation from six domains, which have pairwise correlations ranging between −0.83 and 0.98. Therefore we included variables relating to each of the six remaining domains in separate models, and represent socio-economic deprivation by the domain that minimises the AIC. For the Crime, Employment and Income domains that are measured by only one indicator, we use that indicator as the domain specific measure of socio-economic deprivation. For the remaining domains we apply principal component analysis (PCA) separately to the indicators in that domain, and for simplicity use the first principal component (PC) as the domain specific measure of socio-economic deprivation. These first principal components account for 75% (Access), 99% (Education) and 95% (Housing) of the variation in the domain specific indicators, suggesting that large proportions of the variation in each domain can be captured by just 1 PC. The employment domain produced the best fitting model for both outcomes as measured by AIC, and is the one that is retained in both cases.

Then we added in the four vaccination rate variables in separate models due to their high collinearity, and overall the percentage of the adult population who were double vaccinated by the end of April 2021 best fitted the data and was hence used in the final models for both disease outcomes. Finally we added in the ethnicity covariates, specifically the percentages of each IZ's population who were: (i) black; (ii) from Indian/Pakistani / Bangladeshi descent; and (iii) from Chinese

descent. Note, the percentage of each IZ's population who were white was highly correlated with these three variables and was hence not included. A single air pollutant was also added to the model at this stage, and multiple pollutants are not considered in a single model in the main set of results due to their high pairwise correlations (between 0.66 and 0.91).

The residuals from these final covariate only models were then assessed for the presence of overdispersion and spatial autocorrelation. The residuals exhibited substantial overdispersion for both outcomes, with overdispersion parameters of $\hat{\xi} = 2.13$ (hospitalisations) and $\hat{\xi} = 1.96$ (deaths) respectively, where $\text{Var}[Y_k] = \xi \mathbb{E}[Y_k]$. They also exhibited significant spatial autocorrelation, which was assessed via a Moran's I test (Moran, 1950) based on a neighbourhood matrix (see next section for details) defined by the commonly used border sharing rule. The residuals had observed Moran's I statistics of 0.26 for hospitalisations and 0.10 for deaths, with corresponding p-values of 0.0001 against a null hypothesis of independence (based on a permutation test with 10,000 random permutations) in both cases, suggesting that the data contain residual spatial autocorrelation after covariate adjustment.

## 3. Statistical models

Let $(Y_k, e_k)$ respectively denote the observed and expected numbers of hospitalisations or deaths from the population living in the $k$th IZ, where $k = 1, \ldots, K(= 1279)$. We model hospitalisations and deaths separately because this ensures that correlations are not induced between their risk estimates, as would be the case if a multivariate CAR type model (e.g. Gelfand and Vounatsou, 2003) was used. These between outcome correlations may affect the estimated covariate effects and their standard errors, and we do not want the death rates, for example, to influence the relationship between air pollution and hospitalisation rates. We fit the same univariate spatial models outlined below to the hospitalisation and death outcomes separately, and inference is based in a Bayesian paradigm using Integrated Nested Laplace Approximations (INLA) via the R package INLA (Rue et al., 2009).

### 3.1. Level 1 - Data likelihood model

As the response variable $\mathbf{Y} = (Y_1, \ldots, Y_K)$ is a vector of spatially aggregated counts for the $K$ IZs, the first level of our Bayesian hierarchical model is the Poisson log-linear specification given by

$$Y_k \sim \text{Poisson}(e_k \theta_k) \quad \text{for } k = 1, \ldots, K \tag{1}$$
$$\ln(\theta_k) = \mathbf{x}_k^\top \boldsymbol{\beta} + \phi_k.$$

Here $\theta_k$ denotes the risk (rate) of hospitalisation or death in the $k$th IZ relative to the expected count $e_k$, and is on the same scale as the SMR described above for interpretation purposes. The spatial variation in $\theta_k$ is modelled by a $p \times 1$ vector of covariates, including air pollution concentrations, $\{\mathbf{x}_k\}$, with regression parameters $\boldsymbol{\beta}$, and a spatial random effect $\phi_k$. The air pollution concentrations are taken from the spatially re-aligned PCM model output described in the previous section, and for consistency with the existing air pollution and Covid-19 studies (e.g. Wu et al., 2020, Huang et al., 2021 and Konstantinoudis et al., 2021) we assume they are known concentrations. Additionally, these concentrations are used by the UK government advisory Committee on the Medical Effects of Air Pollutants (COMEAP) in their reports (see for example COMEAP, 2010), which validates our use of them in this study as known concentrations. We note that an alternative modelling approach would be to allow for uncertainty in the estimated pollution concentrations, using approaches similar to those proposed by Blangiardo et al. (2016), Lee et al. (2016), Huang et al. (2018) and Cameletti et al. (2019). The next section describes the models we fit for the spatial random effects $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_K)$, while the following section discusses prior specification.

### 3.2. Level 2 - Spatial random effects model

The random effects $\boldsymbol{\phi}$ allow for any residual spatial autocorrelation present in the data after covariate adjustment, which is caused by factors such as unmeasured confounding. The spatial autocorrelation structure in the random effects is controlled by a $K \times K$ binary neighbourhood matrix $\mathbf{W}$, which specifies which pairs of areas are spatially close together, i.e. are defined to be *neighbours*. If two areas $(k, j)$ are defined to be *neighbours* then their random effects $(\phi_k, \phi_j)$ are modelled as partially autocorrelated, while if they are defined not to be neighbours then $(\phi_k, \phi_j)$ are modelled as conditionally independent. In this paper we consider the two commonly used specifications of $\mathbf{W}$ described below, which allows us to assess the sensitivity of our results to this modelling choice.

**Border sharing -** The border sharing rule is the most commonly used method for specifying $\mathbf{W}$, and $w_{kj} = 1$ if areal units $(k, j)$ share a common border and $w_{kj} = 0$ otherwise, with $w_{kk} = 0 \, \forall \, k$. However, in our data there are 6 island IZs that have no neighbours under this definition, which we rectify by assigning the mainland IZ that is closest to each of these island IZs as a neighbour. The matrix is further adjusted by adding in 5 additional neighbour relations linking island groups to the mainland, which ensures the neighbourhood graph has only one connected component.

**Nearest neighbour -** The border sharing rule results in vastly different numbers of neighbours across the set of IZs (between 1 and 26), so we compare this to the $d$-nearest neighbour rule where every IZ has a similar number of neighbours. Specifically, area $j$ is a neighbour of area $k$, and hence $w_{kj} = 1$, if it is one of the $d$ nearest to it in terms of inter-centroid distances, with $w_{kk} = 0 \, \forall \, k$ as before. This matrix is asymmetric, and is made symmetric by if $w_{kj} = 1$ and $w_{jk} = 0$ then setting $w_{jk} = 1$. Here we choose $d = 4$ because it has an almost identical mean number of neighbours for each IZ as the border sharing specification (5.162 compared to 5.163). However, these specifications only agree on around 60% of their neighbours, meaning they are somewhat different specifications of spatial closeness. Finally, the matrix is adjusted by adding in 2 additional neighbour relations to ensure the neighbourhood graph has only one connected component.

The adjustments described above to ensure the neighbourhood graphs each have one connected component ensures that multiple identifiability constraints (one for each connected component) are not required when fitting the model. A number of conditional autoregressive (CAR) priors have been proposed for modelling spatial autocorrelation in areal unit data using a neighbourhood matrix, and here we consider the following three specifications to see how robust our results are to this modelling choice.

**Leroux -** The CAR prior proposed by Leroux et al. (2000) for globally smooth spatial autocorrelation.

**Modified BYM -** The CAR prior proposed by Simpson et al. (2017) for globally smooth spatial autocorrelation whose scale is not affected by $\mathbf{W}$.

**LCAR -** The CAR prior proposed by Lee and Mitchell (2013) for locally smooth spatial autocorrelation.

A full mathematical specification of each of these random effects priors is given in Section 2 of the supplementary material accompanying this paper.

### 3.3. Level 3 - Prior distributions

The regression parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ are assigned independent weakly informative zero-mean Gaussian prior distributions with a large variance, i.e. $\beta_j \sim N(0, 100000)$ for $j = 1, \ldots, p$, to ensure the data play the dominant role in estimating their values. This choice of prior distribution for the regression parameters is commonly used in the literature, but the choice of prior distributions for the precision and spatial dependence parameters $(\tau, \rho)$ described in Section 2 of the supplementary material is less straightforward. Therefore we fit the three hierarchical spatial models described above to the data with two different choices of prior distributions for $(\tau, \rho)$, so that its impact on the results can be assessed. These two sets of prior distributions are given by:

**Main -** $\ln(\tau) \sim \text{log-gamma}(1, 0.01)$ and $\ln\left(\frac{\rho}{1-\rho}\right) \sim N(0, 100)$, which are chosen to be weakly informative, thus letting the data speak for themselves.

**Alternative -** $\ln(\tau) \sim \text{log-gamma}(1, 0.0005)$ and $\ln\left(\frac{\rho}{1-\rho}\right) \sim N(0, 2.22)$, which are chosen as they are the default priors suggested for the Leroux CAR model by the INLA software that we use for inference.

## 4. Study results

The results of our study are presented below, and in all cases inference is based on integrated nested Laplace approximations using the full Laplace approximation. To illustrate the sensitivity of our results to model specification, we compare models with all possible combinations of the following three factors:

- **Spatial random effects priors** - the three CAR priors for the spatial random effects listed in Section 3.2 denoted by **Leroux**, **Modified BYM**, and **LCAR**.
- **Neighbourhood matrix** - the two specifications of the neighbourhood matrix described in Section 3.2, namely the border sharing rule denoted **Border** and the 4 nearest neighbour rule denoted **4NN**.
- **Priors** - the two sets of prior distributions for $(\tau, \rho)$ outlined in Section 3.3 and denoted by **Main** and **Alternative**.

In all cases the models include the following covariates, which were chosen following the model building exercise outlined in Section 2.4: Covid-19 case rates per 100,000 people over the study period; number of carehome places; population density; the percentage of working age people defined to be employment deprived; the percentage of the adult population who were double vaccinated by the end of April 2021; and the percentages of people who are black, Chinese, or from Indian/Pakistani/Bangladeshi descent. Finally, each model includes one of the three air pollutants ($NO_2$, $PM_{2.5}$, $PM_{10}$). Note, multiple pollutants are not included in the same model in this main analysis due to their high collinearity. The next sub-section describes the overall fit of each model (4.1), while subsequent sub-sections present the results of the air pollution effects (4.2), the other covariate effects (4.3), and outlines the further sensitivity analyses presented in the supplementary material accompanying this paper (4.4).

### 4.1. Overall model fit

The overall fit to the data of each model specification is summarised by the deviance information criterion (DIC, Spiegelhalter et al., 2002) in Table 2, where smaller values represent a better fitting model. Also presented in the table is the effective number of independent parameters $p.d$, where smaller values represent a less complex model. The values in the table relate to models including $PM_{2.5}$ as the only pollutant. The table shows that for both the hospitalisation and death

**Table 2**

Overall fit of the different models to each data set as measured by the deviance information criterion (DIC) and the effective number of independent parameters ($p.d$) in brackets.

| W matrix | Priors | Random effects model | | |
|---|---|---|---|---|
| | | Leroux | Modified BYM | LCAR |
| **Hospitalisation data** | | | | |
| Border | Main | 7917 (405) | 7914 (439) | 7907 (401) |
| | Alternative | 7917 (414) | 7911 (451) | 7908 (409) |
| 4NN | Main | 7962 (363) | 7938 (447) | 7944 (361) |
| | Alternative | 7956 (382) | 7937 (451) | 7942 (376) |
| **Death data** | | | | |
| Border | Main | 6427 (439) | 6378 (509) | 6412 (436) |
| | Alternative | 6426 (454) | 6378 (509) | 6409 (446) |
| 4NN | Main | 6418 (503) | 6391 (513) | 6403 (466) |
| | Alternative | 6417 (503) | 6391 (512) | 6402 (470) |

outcomes changing the model has relatively little impact on the DIC, with maximum reductions of 55 for hospitalisations and 49 for deaths between the best and worst fitting models. The LCAR prior fits the hospitalisation data best while the modified BYM CAR prior fits the death data best, although as described above the differences in DICs are not large. The effect of changing the specification of **W** is smaller for the death outcome and slightly larger for the hospitalisation outcome, and for the latter the border sharing specification consistently best fits the data. Finally, the choice of priors for $(\tau, \rho)$ have almost no impact on model fit for either disease outcome as measured by the DIC.

### 4.2. Pollution-health effects

The estimated relative risks between each pollutant and disease outcome for all model specifications are displayed in Tables 3 and 4, which respectively relate to the hospitalisation and death outcomes. In all cases these relative risks are estimated from single pollutant models, and relate to a 5 μg m$^{-3}$ increase in $NO_2$ and a 1 μg m$^{-3}$ increase in both $PM_{2.5}$ and $PM_{10}$ as these are realistic increases in the concentrations that might occur in Scotland (see Table 1).

For the hospitalisation outcome the estimated relative risks are fairly consistent across the different models, suggesting the results are relatively robust to model misspecification. The main finding is that elevated $PM_{2.5}$ concentrations are consistently statistically significantly associated with Covid-19 hospitalisations, with a 1 μg m$^{-3}$ increase in concentrations being associated with between a 7.4% and a 9.3% increased risk. To further examine the consistency of these estimated associations between the models we summarise their entire posterior distributions via density strips in Section 3 of the supplementary material accompanying this paper.

Increasing $PM_{10}$ concentrations by 1 μg m$^{-3}$ is estimated to increase the risk of hospitalisation by between 1.9% and 2.9%, with half of these estimated associations (those based on the border sharing **W**) being just statistically significant at the 5% level while the other half (those based on the 4 nearest neighbour **W**) are nearly statistically significant. Finally, the estimated relative risks for $NO_2$ correspond to between a 1.6% and a 2.8% increased risk for a 5 μg m$^{-3}$ increase in concentrations for all models considered, but these effects are not statistically significant as all the 95% credible intervals contain the null risk of one.

For the death outcome the estimated relative risks are more variable across the different model specifications for $NO_2$ and $PM_{2.5}$ compared to the corresponding results for hospitalisations. This is most likely to be because the death outcome has many fewer events in each IZ compared to the hospitalisation outcome (6 vs. 21 on average), leading to less precise and consistent inference and a stronger impact of the prior distributions. For $PM_{2.5}$ a 1 μg m$^{-3}$ increase in concentrations is associated with between a 2.9% and a 10.3% increased risk, and

**Table 3**
Estimated relative risks and 95% credible intervals for the effects of each pollutant on Covid-19 hospitalisation rates using 2019 pollution data. The relative risks relate to a 5 μg m$^{-3}$ increase in NO$_2$ and a 1 μg m$^{-3}$ increase in both PM$_{2.5}$ and PM$_{10}$.

| Pollutant | W matrix | Priors | Random effects model | | |
|---|---|---|---|---|---|
| | | | Leroux | Modified BYM | LCAR |
| NO$_2$ | Border | Main | 1.016 (0.980, 1.053) | 1.020 (0.984, 1.056) | 1.020 (0.985, 1.056) |
| | | Alternative | 1.018 (0.982, 1.056) | 1.020 (0.984, 1.057) | 1.022 (0.987, 1.059) |
| | 4NN | Main | 1.021 (0.985, 1.058) | 1.021 (0.985, 1.057) | 1.025 (0.985, 1.058) |
| | | Alternative | 1.024 (0.988, 1.062) | 1.021 (0.986, 1.058) | 1.028 (0.991, 1.065) |
| PM$_{2.5}$ | Border | Main | 1.082 (1.030, 1.136) | 1.090 (1.038, 1.143) | 1.086 (1.034, 1.140) |
| | | Alternative | 1.089 (1.037, 1.144) | 1.090 (1.039, 1.144) | 1.093 (1.041, 1.147) |
| | 4NN | Main | 1.074 (1.022, 1.128) | 1.082 (1.029, 1.137) | 1.080 (1.028, 1.135) |
| | | Alternative | 1.086 (1.033, 1.142) | 1.083 (1.030, 1.138) | 1.089 (1.036, 1.144) |
| PM$_{10}$ | Border | Main | 1.025 (1.002, 1.048) | 1.029 (1.006, 1.052) | 1.025 (1.003, 1.048) |
| | | Alternative | 1.026 (1.004, 1.050) | 1.029 (1.006, 1.052) | 1.027 (1.004, 1.050) |
| | 4NN | Main | 1.019 (0.996, 1.042) | 1.021 (0.998, 1.045) | 1.019 (0.997, 1.043) |
| | | Alternative | 1.021 (0.998, 1.045) | 1.021 (0.998, 1.045) | 1.021 (0.998, 1.044) |

**Table 4**
Estimated relative risks and 95% credible intervals for the effects of each pollutant on Covid-19 death rates using 2019 pollution data. The relative risks relate to a 5 μg m$^{-3}$ increase in NO$_2$ and a 1 μg m$^{-3}$ increase in both PM$_{2.5}$ and PM$_{10}$.

| Pollutant | W matrix | Priors | Random effects model | | |
|---|---|---|---|---|---|
| | | | Leroux | Modified BYM | LCAR |
| NO$_2$ | Border | Main | 1.023 (0.958, 1.091) | 1.017 (0.958, 1.080) | 1.027 (0.963, 1.095) |
| | | Alternative | 1.032 (0.967, 1.101) | 1.017 (0.958, 1.080) | 1.034 (0.970, 1.102) |
| | 4NN | Main | 1.058 (0.996, 1.124) | 1.018 (0.959, 1.081) | 1.049 (0.985, 1.116) |
| | | Alternative | 1.059 (0.997, 1.124) | 1.018 (0.959, 1.081) | 1.052 (0.988, 1.118) |
| PM$_{2.5}$ | Border | Main | 1.035 (0.940, 1.140) | 1.029 (0.948, 1.117) | 1.051 (0.956, 1.152) |
| | | Alternative | 1.068 (0.971, 1.168) | 1.029 (0.948, 1.117) | 1.077 (0.982, 1.173) |
| | 4NN | Main | 1.102 (1.028, 1.179) | 1.029 (0.944, 1.121) | 1.091 (1.005, 1.175) |
| | | Alternative | 1.103 (1.030, 1.180) | 1.029 (0.944, 1.120) | 1.096 (1.014, 1.177) |
| PM$_{10}$ | Border | Main | 1.002 (0.963, 1.043) | 1.006 (0.968, 1.044) | 1.005 (0.966, 1.046) |
| | | Alternative | 1.009 (0.969, 1.050) | 1.006 (0.968, 1.044) | 1.010 (0.970, 1.050) |
| | 4NN | Main | 1.018 (0.981, 1.056) | 0.999 (0.960, 1.039) | 1.008 (0.966, 1.048) |
| | | Alternative | 1.019 (0.983, 1.056) | 0.999 (0.960, 1.039) | 1.011 (0.971, 1.050) |

4 of these 12 associations are statistically significant at the 5% level. The corresponding risks for NO$_2$ range between 1.7% and 5.9% for a 5 μg m$^{-3}$ increase in concentrations, although in all cases these effects are not statistically significant. Finally, the effects of PM$_{10}$ are consistent across all 12 models, with non-significant relative risks close to one in all cases.

### 4.3. Non-pollution covariate effects

The estimated relative risks and 95% credible intervals for the non-pollution covariates are presented in Tables 5 and 6, which respectively relate to hospitalisations and deaths. In all cases the relative risks relate to realistic increases in each covariate, which are given in brackets in column 1 of the tables. The tables present relative risks from models with the three types of spatial CAR prior assumed for the random effects that allow for spatial autocorrelation. However, for brevity all results relate to the border sharing neighbourhood matrix with the main priors for ($\tau, \rho$) and PM$_{2.5}$ being the pollutant included in the model. The tables show that in all cases the estimated relative risks are almost identical across the three spatial correlation structures, suggesting that the results are robust to this choice.

The tables show that unsurprisingly increased Covid-19 case rates are significantly associated with increased hospitalisations and deaths, as increasing the case rate per 100,000 people over the entire study duration by 1000 is estimated to increase hospitalisations by around 12% and deaths by around 13.5%. Covid-19 vaccination rates, as measured by the percentage of adults who were double vaccinated by the end of April 2021, are significantly associated with reduced deaths, with a 5% increase in vaccination rates estimated to lead to a 4% reduction in the risk of death. This relatively small effect size occurs because a sizeable proportion of the deaths occurred before the vaccination programme

began, thus attenuating its effect size. In contrast, vaccinations appear to have no effect on the risk of hospitalisation after adjusting for the other covariates, as the estimated relative risks in Table 5 are close to the null risk of one and are not significant at the 5% level. This may be because the cohort of people who were vaccinated by April 2021 were the mostly vulnerable individuals in health terms (either in the highest age groups or people with underlying health conditions), and hence vaccination reduced their likelihood of death but not of being hospitalised. Alternatively, the lack of an effect may be due to a sizeable proportion of the study duration occurring before vaccination began, which means we should treat the results of this variable with extreme caution.

In addition to Covid-19 case rates, the other main driver of elevated risks of hospitalisation and death is socio-economic deprivation. A 5% increase in the percentage of the working age population who are defined to be employment deprived is significantly associated with a 16% increased risk of hospitalisation and a 8.5% increased risk of death. The number of care home places is also significantly associated with Covid-19 deaths, because if an area increases its care home places by 10 the estimated risk of death increases by around 4%. In contrast, care homes appear to exhibit no association with hospitalisation risk, as the relative risks are all close to one. Population density appears to have no significant associations with either hospitalisation or death once the other covariate effects are accounted for, as all the estimated relative risks are close to one. Finally, ethnicity does not appear overall to have a strong association with Covid-19 hospitalisations or deaths, as 14 of the 18 estimated associations are not significant at the 5% level. An exception to this is for the risk of hospitalisation for people from Chinese descent, as for every 1% increase in an areas' Chinese population, the relative risk reduces by between 2.3% and 2.6%. Additionally, people who are black appear to have a small increased risk of both

**Table 5**

Estimated relative risks and 95% credible intervals for the effects of the non-pollution covariates on Covid-19 hospitalisation. The relative risks relate to the increases given in brackets in the first column of the table. Note, the final row of the table relates to the % of the population in each IZ that are from Indian, Pakistani or Bangladeshi origin.

| Covariate | Random effects model | | |
|---|---|---|---|
| | Leroux | Modified BYM | LCAR |
| Case rates (1000) | 1.117 (1.105, 1.129) | 1.120 (1.108, 1.132) | 1.122 (1.110, 1.134) |
| Carehome places (10) | 0.998 (0.995, 1.001) | 0.998 (0.995, 1.001) | 0.998 (0.995, 1.001) |
| Population density (10) | 1.004 (0.996, 1.012) | 1.004 (0.996, 1.012) | 1.004 (0.996, 1.012) |
| % Employment deprivation (5) | 1.161 (1.141, 1.181) | 1.159 (1.138, 1.179) | 1.159 (1.139, 1.179) |
| % Vaccinated (5) | 1.006 (0.991, 1.022) | 1.004 (0.989, 1.020) | 1.007 (0.992, 1.022) |
| % Black (1) | 1.013 (0.998, 1.028) | 1.015 (1.000, 1.030) | 1.014 (0.999, 1.028) |
| % Chinese (1) | 0.977 (0.956, 0.998) | 0.974 (0.953, 0.996) | 0.975 (0.954, 0.996) |
| % Ind/Pak/Ban (1) | 1.001 (0.994, 1.008) | 1.001 (0.994, 1.008) | 1.001 (0.994, 1.007) |

**Table 6**

Estimated relative risks and 95% credible intervals for the effects of the non-pollution covariates on Covid-19 death. The relative risks relate to the increases given in brackets in the first column of the table. Note, the final row of the table relates to the % of the population in each IZ that are from Indian, Pakistani or Bangladeshi origin.

| Covariate | Random effects model | | |
|---|---|---|---|
| | Leroux | Modified BYM | LCAR |
| Case rates (1000) | 1.134 (1.112, 1.156) | 1.139 (1.118, 1.161) | 1.135 (1.114, 1.157) |
| Carehome places (10) | 1.040 (1.034, 1.045) | 1.040 (1.035, 1.046) | 1.040 (1.035, 1.045) |
| Population density (10) | 1.005 (0.991, 1.019) | 1.008 (0.993, 1.022) | 1.004 (0.990, 1.019) |
| % Employment deprivation (5) | 1.086 (1.053, 1.120) | 1.084 (1.050, 1.120) | 1.086 (1.053, 1.120) |
| % Vaccinated (5) | 0.963 (0.938, 0.990) | 0.960 (0.935, 0.986) | 0.963 (0.937, 0.989) |
| % Black (1) | 1.023 (0.996, 1.051) | 1.023 (0.995, 1.053) | 1.023 (0.995, 1.050) |
| % Chinese (1) | 0.990 (0.951, 1.030) | 0.986 (0.947, 1.026) | 0.990 (0.951, 1.029) |
| % Ind/Pak/Ban (1) | 0.994 (0.981, 1.007) | 0.994 (0.982, 1.006) | 0.994 (0.982, 1.007) |

hospitalisation and death, although the results are not quite statistically significant. However, given the only ethnicity data available relate to 2011 these results should again be interpreted with caution.

### 4.4. Further sensitivity analyses

We have conducted a number of further sensitivity analyses to assess the robustness of our results, and they are presented in Sections 4 to 7 of the supplementary material accompanying this paper. The first of these relates to how we estimate pollution concentrations, and in the main study we have used annual average pollution concentrations for 2019 to represent population exposure, because as previously discussed it is likely to be more representative of *normal* concentrations than annual average concentrations in 2020. However, in Section 4 of the supplementary material we present estimated relative risks and 95% credible intervals based on the 2020 annual average pollution concentrations as a sensitivity analysis, and the results show broadly similar messages to those presented in Tables 3 and 4.

Section 5 of the supplementary material investigates the potential for spatially varying effects across Scotland, specifically estimating separate effects for the health boards containing the four largest cities in Scotland (Grampian–Aberdeen; Tayside–Dundee; Lothian–Edinburgh; and Greater Glasgow and Clyde–Glasgow). Additionally, the models fitted so far in this study consider each pollutant separately in single pollutant models, because this is the norm in the existing air pollution and Covid-19 literature. However, we extend this in Section 6 of the supplementary material by estimating the joint effects of $NO_2$ and $PM_{2.5}$ / $PM_{10}$, because individuals are simultaneously exposed to multiple different pollutants. Finally, as the estimated $PM_{2.5}$ association with hospitalisations is the strongest of all those considered, we examine how this association varies by age group and period of the study in Section 7 of the supplementary material.

### 5. Discussion

This paper presents the first study examining the potential effects that air pollution may have on worsening Covid-19 outcomes at the small-area scale in Scotland, and adds to the relatively small international evidence base on this topic that includes (Coker et al., 2020; Wu et al., 2020; Berg et al., 2021; Konstantinoudis et al., 2021; Sun et al., 2021). The major strength of our study is the extensive sensitivity analysis we have performed, which allows us to quantify the robustness of our results to model misspecification. This sensitivity analysis includes changing the residual spatial autocorrelation structure (both **W** and the random effects model), the prior distributions assumed, the estimation of air pollution concentrations, the potential for regional variation in effect sizes, the simultaneous effects of multiple pollutants, and how the effects vary with age and vaccination.

The major limitation of our study is its ecological (population-level) design, as the unit of inference is the group of individuals who live in each IZ rather than having data for each individual. Ecological-level studies are prone to ecological bias and the ecological fallacy (Wakefield and Salway, 2001), the latter of which is where one incorrectly assumes that the estimated group-level association holds at the individual-level. This estimated group-level association does not in any way equate to individual-level cause-and-effect, because it may be affected by factors such as within-area variation in either the exposure or the confounders. We used this ecological study design because the data required for an individual-level study were not available for confidentiality reasons, and this ecological approach has been predominant in the air pollution and Covid-19 literature to date (see for example Konstantinoudis et al., 2021 and Berg et al., 2021, while Mendy et al., 2021 is a notable exception). Thus our findings outlined below should be treated as indicative associations, rather than evidence of individual-level cause-and-effect.

Our main finding is that fine particulate matter air pollution measured as $PM_{2.5}$ is significantly associated with increased risks of hospitalisation due to Covid-19, as all 12 estimated relative risks for this pollutant-outcome pair are statistically significant at the 5% level. The size of the estimated relative risks are also fairly consistent, ranging between a 7.4% and a 9.3% increase in risk for each $1\mu g \ m^{-3}$ increase in concentrations. This compares to estimates of around 18% in Ohio from Mendy et al. (2021) in an individual-level study, and 26% in Colorado from Berg et al. (2021) in a spatial ecological study similar to this one.

Similarly, when considering death as the outcome measure, Coker et al. (2020) and Wu et al. (2020) report significant increased risks of 9% and 11% for each $1 \, \mu g \, m^{-3}$ increase in $PM_{2.5}$ concentrations in northern Italy and the USA respectively. In contrast, Konstantinoudis et al. (2021) report insignificant results in England, but this may be due to their use of estimated (using a downscaling model) rather than known Covid-19 death counts at the small-area scale. It may also be due to the small number of deaths they had in each areal unit, which numbered just over 1 on average. Our present study has an average of only 6 deaths in each IZ, and this could also be the reason why 8 of the 12 estimated relative risks between $PM_{2.5}$ and Covid-19 deaths were not statistically significant. Our estimated relative risks ranged between around a 2.9% and a 10.3% increased risk, but the small numbers of deaths caused these estimates to have relatively wide and hence non-significant 95% credible intervals in the majority of cases.

Overall then this study corroborates the majority of the earlier research findings situated in different countries, which adds weight to the international evidence that fine particulate matter air pollution worsens Covid-19 outcomes. The sizes of these estimates across the different studies is however quite variable, which could be due to numerous factors including the varying concentrations of air pollution and the small numbers of Covid-19 events in each areal unit making the estimates slightly unstable. Thus an important area of future work will be to conduct a meta-analysis of these studies, which will allow a more precise global effect estimate to be produced that borrows strength across the individual studies. A further area of future work will be to investigate the impact of allowing for pollution uncertainty in the disease models, for example using an approach similar to that proposed by Cameletti et al. (2019).

Our overall conclusions around the effects of the other two pollutants on Covid-19 ill health are less clear cut. For example, $PM_{10}$ appears to have a borderline significant association with hospitalisations (6 out of 12 associations were significant), with increased risks of between 2% and 3% for a $1 \, \mu g \, m^{-3}$ increase in concentrations. In contrast, no effects are observed with Covid-19 death, as all 12 estimated relative risks are very close to one and non-significant. For $NO_2$ the increased risks are estimated to be around 1%–3% for hospitalisations and between 2% and 6% for deaths, but in all cases they are not significant at the 5% level. The effects of $NO_2$ and $PM_{10}$ on Covid-19 outcomes appear to have been examined less frequently than those of $PM_{2.5}$, and of the studies referenced here only (Konstantinoudis et al., 2021) has considered $NO_2$ while none considered $PM_{10}$. This suggests that future studies are needed to examine the effects that both these pollutants may have on Covid-19 outcomes, to provide more conclusive evidence on their effects than have been possible to draw here.

### Declaration of competing interest

### Acknowledgements

### Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.sste.2022.100523.

### References

Berg, K., Romer Present, P., Richardson, K., 2021. Long-term air pollution and other risk factors associated with COVID-19 at the census tract level in colorado. Environ. Pollut. 287, 117584.

Berrocal, V., Gelfand, A., Holland, D., 2010. A spatio-temporal downscaler for output from numerical models. J. Agric. Biol. Environ. Statist. 15, 176–197.

Blangiardo, M., Finazzi, F., Cameletti, M., 2016. Two-stage Bayesian model to evaluate the effect of air pollution on chronic respiratory diseases using drug prescriptions. Spat. Spatio-Temporal Epidemiol. 18, 1–12.

Cameletti, M., Gómez-Rubio, V., Blangiardo, M., 2019. Bayesian modelling for spatially misaligned health and air pollution data through the INLA-SPDE approach. Spat. Statist. 31, 100353.

Coker, E., Cavalli, L., Fabrizi, E., Guastella, G., Lippo, E., Parisi, M., Pontarollo, N., Rizzati, M., Varacca, A., Vergalli, S., 2020. The effects of air pollution on COVID19 related mortality in northern Italy. Environ. Resour. Econ. 76, 611–634.

COMEAP, 2010. The mortality effects of long-term exposure to particulate air pollution in the United Kingdom. https://www.gov.uk/government/collections/comeap-reports.

COMEAP, 2020. Summary of COMEAP recommendations for the quantification of health effects associated with air pollutants. https://www.gov.uk/government/collections/comeap-reports.

Conticini, E., Frediani, B., Caro, D., 2020. Can atmospheric pollution be considered a co-factor in extremely high level of SARS-CoV-2 lethality in Northern Italy? Environ. Pollut. 261, 114465.

Dong, E., Du, H., Gardner, L., 2020. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect. Dis. 20, 533–534.

Douglas, M., Katikireddi, S., Taulbut, M., McKee, M., McCartney, G., 2020. Mitigating the wider health effects of Covid-19 pandemic response. Br. Med. J. 369, m1557.

Forlani, C., Bhatt, S., Cameletti, M., Krainski, E., Blangiardo, M., 2020. A joint Bayesian space–time model to integrate spatially misaligned air pollution data in R-INLA. Environmetrics 31, e2644.

Gelfand, A., Vounatsou, P., 2003. Proper multivariate conditional autoregressive models for spatial data analysis. Biostatistics 4 (1), 11–15.

Gelfand, A., Zhu, L., Carlin, B., 2001. On the change of support problem for spatio-temporal data. Biostatistics 2, 31–45.

Haining, R., Li, G., Maheswaran, R., Blangiardo, M., Law, J., Best, N., Richardson, S., 2010. Inference from ecological models: estimating the relative risk of stroke from air pollution exposure using small area data. Spat. Spatio-Temporal Epidemiol. 1, 123–131.

Huang, G., Blangiardo, M., Brown, P., Pirani, M., 2021. Long-term exposure to air pollution and COVID-19 incidence: A multi-country study. Spat. Spatio-Temporal Epidemiol. 39, 100443.

Huang, G., Lee, D., Scott, M., 2018. Multivariate space-time modelling of multiple air pollutants and their health effects accounting for exposure uncertainty. Stat. Med. 37, 1134–1148.

Konstantinoudis, G., Padellini, T., Bennett, J., Davies, B., Ezzati, M., Blangiardo, M., 2021. Long-term exposure to air-pollution and COVID-19 mortality in England: a hierarchical spatial analysis. Environ. Int. 146, 106316.

Lee, D., Mitchell, R., 2013. Locally adaptive spatial smoothing using conditional auto-regressive models. J. R. Stat. Soc. Ser. C. Appl. Stat. 62, 593–608.

Lee, D., Mukhopadhyay, S., Rushworth, A., Sahu, S., 2016. A rigorous statistical framework for spatio-temporal pollution prediction and estimation of its long-term impact on health. Biostatistics 18, 370–385.

Lee, D., Robertson, C., Marques, D., 2021. Quantifying the small-area spatio-temporal dynamics of the Covid-19 pandemic in Scotland during a period with limited testing capacity. Spat. Statist. 100508.

Lee, D., Robertson, C., Ramsay, C., Gillespie, C., Napier, G., 2019. Estimating the health impact of air pollution in Scotland, and the resulting benefits of reducing concentrations in city centres. Spat. Spatio-Temporal Epidemiol. 29, 85–96.

Leroux, B., Lei, X., Breslow, N., 2000. Estimation of Disease Rates in Small Areas: A New Mixed Model for Spatial Dependence. Springer-Verlag, New York, pp. 135–178.

Mendy, A., Wu, X., Keller, J., Fassler, C., Apewokin, S., Mersha, T., Xie, C., Pinney, S., 2021. Air pollution and the pandemic: Long-term PM2.5 exposure and disease severity in COVID-19 patients. Respirology.

Moran, P., 1950. Notes on continuous stochastic phenomena. Biometrika 37, 17–23.

Remuzzi, A., Remuzzi, G., 2020. COVID-19 And Italy: what next? Lancet 395, 1225–1228.

Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). J. R. Statist. Soc. Ser. B 71, 319–392.

Scottish Government, 2021. Cleaner air for Scotland 2 – Towards a better place for everyone. https://www.gov.scot/publications/cleaner-air-scotland-2-towards-better-place-everyone/.

Simpson, D., Rue, H., Riebler, A., Martins, T., Sarbye, S., 2017. Penalising model component complexity: A principled, practical approach to constructing priors. Statist. Sci. 32, 1–28.

Spiegelhalter, D., Best, N., Carlin, B., Van der Linde, A., 2002. Bayesian measures of model complexity and fit. J. R. Statist. Soc. B 64, 583–639.

Sun, Y., Hu, X., Xie, J., 2021. Spatial inequalities of COVID-19 mortality rate in relation to socioeconomic and environmental factors across England. Sci. Total Environ. 143595, 533–534.

Wakefield, J., Salway, R., 2001. A statistical framework for ecological and aggregate studies. J. R. Statist. Soc.: Ser. A (Statist. Soc.) 164, 119–137.

World Health Organisation, 2016. Ambient air pollution: A global assessment of exposure and burden of disease. http://apps.who.int/iris/bitstream/10665/250141/1/9789241511353-eng.pdf.

Wu, X., Nethery, R., Sabath, M., Braun, D., Dominici, F., 2020. Air pollution and COVID-19 mortality in the United States: Strengths and limitations of an ecological regression analysis. Sci. Adv. 6, eabd4049.