



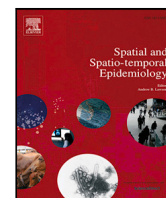
Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Contents lists available at ScienceDirect

Spatial and Spatio-temporal Epidemiology

journal homepage: www.elsevier.com/locate/sste

Original research

Long-term exposure to air pollution and COVID-19 incidence: A multi-country study[☆]

Guowen Huang^{a,b,*}, Marta Blangiardo^c, Patrick E. Brown^{a,b}, Monica Pirani^c^a Department of Statistical Sciences, University of Toronto, Toronto, ON, Canada^b Centre for Global Health Research, St Michael's Hospital, Toronto, ON, Canada^c MRC Centre for Environment and Health, Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK

ARTICLE INFO

Keywords:

Coronavirus disease
Spatial model
Epidemiology
Multi-country
INLA

ABSTRACT

The study of the impacts of air pollution on COVID-19 has gained increasing attention. However, most of the existing studies are based on a single country, with a high degree of variation in the results reported in different papers. We attempt to inform the debate about the long-term effects of air pollution on COVID-19 by conducting a multi-country analysis using a spatial ecological design, including Canada, Italy, England and the United States. The model allows the residual spatial autocorrelation after accounting for covariates. It is concluded that the effects of PM_{2.5} and NO₂ are inconsistent across countries. Specifically, NO₂ was not found to be an important factor affecting COVID-19 infection, while a large effect for PM_{2.5} in the US is not found in the other three countries. The Population Attributable Fraction for COVID-19 incidence ranges from 3.4% in Canada to 45.9% in Italy, although with considerable uncertainty in these estimates.

1. Introduction

The current outbreak of coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), led to 63 million cases and 1.4 million deaths worldwide by 30th Nov, 2020. Governments around the world imposed travel and business restrictions to slow the spread of COVID-19, which has impacted almost every part of life all over the world. The pandemic-imposed lockdown has had positive effects on air quality, as shown by AEA (2011) that the global energy-related carbon dioxide (CO₂) emissions fell by 5.8%. Similarly, Gautam (2020) compared the air quality data before and during the novel coronavirus pandemic, suggesting a significant reduction in the level of nitrogen dioxide (NO₂) in Asian and European countries due to COVID-19 lockdowns. However, since the long-term adverse effects of air pollution on human health (particularly respiratory and lung diseases) have been well documented (e.g., Pope et al., 2002; Janes et al., 2007; Haining et al., 2010; Huang et al., 2018), questions have been raised about its chronic exposure effects on COVID-19 death/incidence rates. As a result, there has been a steady growth in the investigation of air pollution impacts on COVID-19 in literature, with examples including Gupta et al. (2021), Gautam et al. (2021), Pozzer et al. (2020), and Wu et al. (2020).

A recent systematic review by Copat et al. (2020) pointed out the important contribution of long-term exposure to PM_{2.5} (atmospheric

particulate matter that has a diameter of less than 2.5 μm) and NO₂ on COVID-19 spread and lethality. Additionally, a large study by Pozzer et al. (2020) estimated that particulate air pollution contributed 15% (95% confidence interval 7%, 33%) to COVID-19 mortality worldwide and found that a significant fraction was attributable to anthropogenic sources, of which 50%–60% was related to fossil fuel. However, there is a high degree of variation in the literature so far regarding the effects of air pollution on COVID-19. For example, Andree (2020) investigated the relationship between PM_{2.5} and COVID-19 in the Netherlands and found that expected COVID-19 cases increased by almost 100% when pollution concentration increased by 20% (2 μgm⁻³), while a study by Cole et al. (2020) found that an increase of 1 μgm⁻³ in PM_{2.5} increased COVID-19 cases by about 7% in Dutch municipalities.

Rather than studying a single county, in this study we attempt to inform the debate about the long-term effects of air pollution on COVID-19 outcomes by looking at COVID-19 cases across different countries, namely Canada, Italy, England and the United States. We focus on reported COVID-19 cases because: (1) most existing studies investigated the air pollution effects on COVID-19 mortality, while there is only limited research focusing on the infection risk; (2) our hypothesis is that people who have been long-term exposed to high pollution could have reduced natural defences to respiratory infections; this in turn might make it easier for SARS-CoV-2 to bring on

[☆] Acknowledgements: P. E. Brown is funded by the Natural Sciences and Engineering Research Council of Canada.

* Corresponding author at: Department of Statistical Sciences, University of Toronto, Toronto, ON, Canada.

E-mail address: hgw0610209@gmail.com (G. Huang).

<https://doi.org/10.1016/j.sste.2021.100443>

Received 8 December 2020; Received in revised form 20 July 2021; Accepted 21 July 2021

Available online 11 August 2021

1877-5845/© 2021 Elsevier Ltd. All rights reserved.

Table 1

Data from Canada, Italy, England and mainland United States, with population in million and COVID-19 cases in thousand (up to 16th Oct, 2020). Covariate summary is shown by its 0.25 and 0.75 quantiles, with units $\mu\text{g m}^{-3}$ for $\text{PM}_{2.5}$, ppb for NO_2 , % for unemployment rate and visible minorities, cases per 100,000 for lung cancer mortality. In addition, in Italy the index of material and social vulnerability was used in place of the unemployment.

	Canada	Italy	England	United States
Population	36.3	60.2	55.5	314.4
Areal units	93	107	149	3108
Total cases	194	373	560	7784
$\text{PM}_{2.5}$	(5.37, 7.38)	(11.15, 16.12)	(7.59, 9.60)	(4.40, 6.90)
NO_2	(2.52, 7.25)	(4.65, 9.69)	(8.00, 15.28)	(2.68, 4.58)
Unemployment	(6.70, 10.00)	(98.07, 99.74)	(1.40, 2.60)	(3.10, 4.80)
Visible minorities	(1.60, 6.70)	(4.18, 8.22)	(4.00, 25.50)	(4.47, 19.87)
Lung cancer	(62.65, 79.33)	(47.95, 61.15)	(85.10, 125.80)	(40.70, 57.50)

COVID-19 infections and spread in the population. Specifically, for each country, we investigate whether long-term average exposure to air pollution ($\text{PM}_{2.5}$ and NO_2) increases the risk of COVID-19 infection, by comparing geographical contrasts in air pollution and incidence rates across small areas while accounting for potential confounders, including lung cancer incidence (a proxy for smoking), unemployment rate and proportions of visible or ethnic minorities.

The remainder of this paper is organized as follows. The data and its exploratory analysis are presented in Section 2, while the statistical methodology is outlined in Section 3. The results of the study are reported in Section 4, and the key conclusions are presented in Section 5.

2. Data

The data used in this study include COVID-19 cases, ambient air pollution concentrations and socio-economic confounders for four countries; a brief summary can be seen in Table 1. For each country, data were collected based on non-overlapping areal units which are 93 health regions, 107 provinces, 149 upper tier local authorities and 3108 counties for Canada, Italy, England and the mainland United States, respectively. Data sources and description are summarized in Appendix. The population of the mainland United States is 314.4 million which is about 9 times that of Canada and 5 times that of Italy or England. Although its area is slightly larger than the US, Canada has a much smaller number of areal units in this study. The median population for the areal units are 196, 386, 277 and 25 thousand for Canada, Italy, England and the mainland United States, respectively. There are 219 counties in the United States having population greater than 300,000. More details are given in Table 1.

2.1. COVID-19 data

For each country, the disease data comprise of the total confirmed COVID-19 cases in each areal unit (upper tier local authority for

Table 2

Estimates and 95% intervals for effect sizes, spatial parameters and population-attributable fraction of incidence. Effects for $\text{PM}_{2.5}$, NO_2 , lung cancer incidence, percent unemployed, and percent ethnic minorities are percent increase in relative risk for a one unit increase, and a value of 1.2 corresponds to a regression coefficient of $\log[1 + (1.2/100)]$.

	Canada		Italy		England		United States	
	Est	CI	Est	CI	Est	CI	Est	CI
Relative risk								
NO_2	5.0	(−3.8, 14.7)	2.7	(−0.4, 6.0)	0.5	(−0.5, 1.6)	2.3	(−3.2, 8.1)
$\text{PM}_{2.5}$	−10.9	(−23.9, 4.5)	0.5	(−2.5, 3.5)	2.9	(−0.7, 6.6)	12.6	(7.4, 18.1)
Lung cancer	−0.2	(−2.3, 2.0)	0.0	(−0.9, 0.9)	0.7	(0.5, 1.0)	0.0	(−0.2, 0.2)
Unemployment	−19.0	(−25.1, −12.4)	−7.4	(−20.8, 8.6)	4.5	(−3.0, 12.7)	−4.3	(−6.0, −2.5)
Visible minorities	2.8	(0.5, 5.1)	2.3	(−0.2, 5.0)	1.1	(0.7, 1.5)	1.3	(1.1, 1.5)
Spatial parameters								
Std deviation σ	1.1	(0.7, 1.6)	0.4	(0.3, 0.5)	0.2	(0.1, 0.2)	0.8	(0.8, 0.9)
Dependence ρ	0.6	(0.4, 0.8)	0.8	(0.6, 0.9)	0.9	(0.8, 1.0)	1.0	(1.0, 1.0)
Population attributable fraction								
PAF	3.4	(−59.9, 39.6)	45.9	(11.1, 66.3)	10.3	(1.6, 19.0)	26.1	(21.5, 30.2)

England, province for Italy, health region for Canada and county for the US) up to 16th Oct, 2020. We denote Y_k as the reported number of COVID-19 cases for the k th areal unit. As the number of case in an areal unit depends on its population, we calculate the expected number of cases in each areal unit (E_k) based on national incidence rate in sex and age groups by (3).

2.2. Air pollution data

The most recently available 3-year data are used to represent the long-term areal unit level pollution concentrations, that is 2014–2016 for Canada and 2016–2018 for the other three countries (2017–2019 for US NO_2).

In this study we consider $\text{PM}_{2.5}$ and NO_2 coming from two data sources (see Table A.1): measurements from a monitoring network (Canada, Italy, England, US NO_2) and gridded modelled output (US $\text{PM}_{2.5}$ and Canada). We based our analysis on directly measured monitoring data wherever possible. For Italy, England and 83 of the 93 regions in Canada, the air pollution monitoring networks are sufficiently dense, therefore simple averages of monitoring data from stations within 5 km buffer of each areal unit were computed.

The monitoring network in North America does not cover a large proportion of the counties in the US and a few more remote regions in Canada, and publicly available gridded data were used in these instances. For Canada, both $\text{PM}_{2.5}$ and NO_2 gridded data are available for recent years, and simple averages of values in grid cells in each of the remote region were used as exposures. For the US, gridded $\text{PM}_{2.5}$ data for recent years are available from Group (2021) and again averages within each county were computed. Gridded NO_2 data for the US was not available for recent years at the time of writing, and we created our own spatial interpolation of ground monitors as described in Section 3.1.

Maps of $\text{PM}_{2.5}$ and NO_2 for each country are shown in Figs. A.1–A.4, while Fig. A.5 presents their distributions across areal units. The latter shows that the long-term averaged $\text{PM}_{2.5}$ levels in Italy are much higher than the others, while the long-term averaged NO_2 levels in England are more uniformly distributed across areal units.

2.3. Socio-economic confounders

Confounders included in the model as areal-level covariates are: (i) the incidence of lung cancer mortality (cases per 100,000 people) as a proxy for smoking prevalence; (ii) unemployment rate as a proxy for socio-economic deprivation; and (iii) the proportion of the population who belong to minority groups. The minority groups differed across countries as reflected in the data available. In Canada the group used is visible minorities, as defined by the Census. For Italy we used individuals from Africa or Asia, the black and minority ethnic population was used for England, and the non-white population used in the US. In addition, in Italy the index of material and social vulnerability was

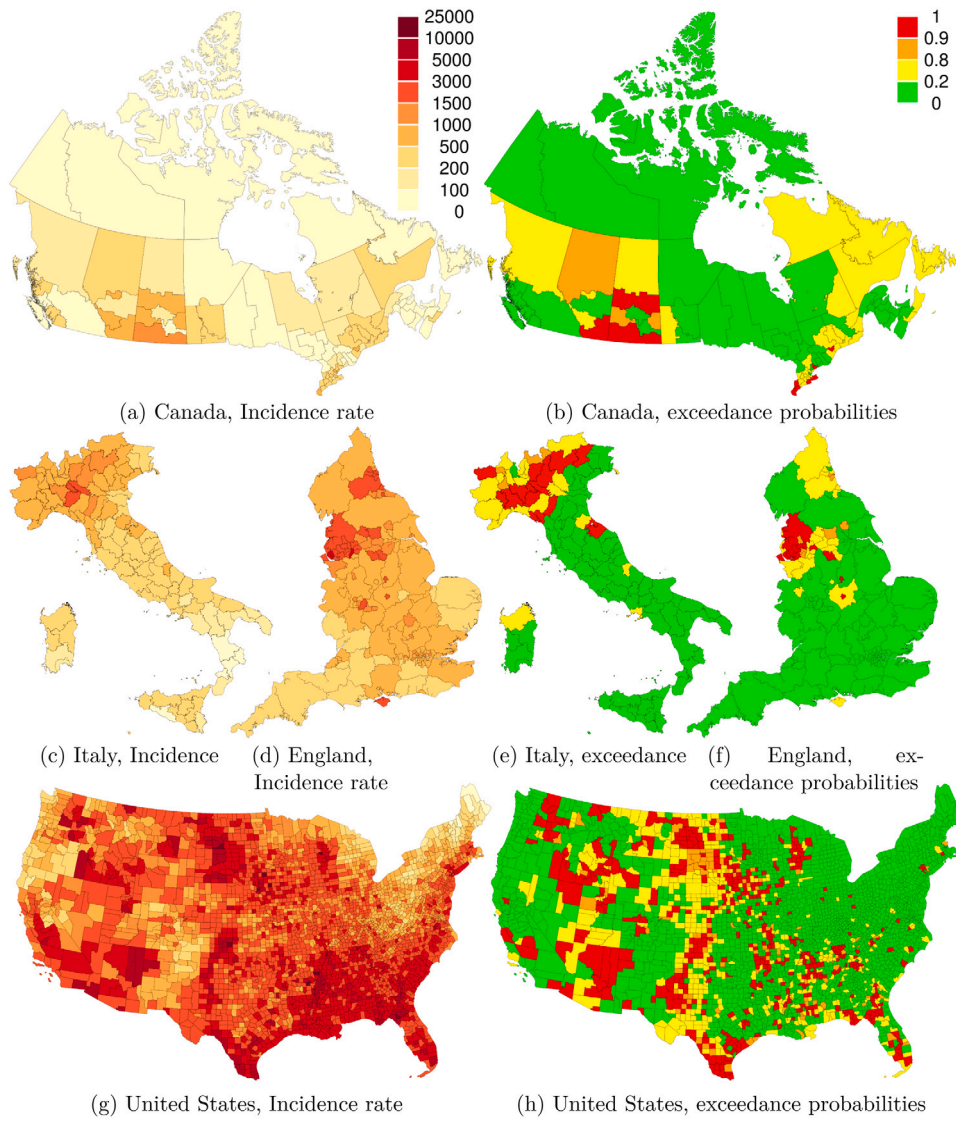


Fig. 1. Incidence rate per 100,000 people $E(\bar{R}_k | Y)$, standardized to the EU standard population and probabilities of 50% excess risk $\Pr(R_k > 1.5 | Y)$ for the four countries in the study.

used in place of the unemployment, the former believed to be a better proxy of socio-economic deprivation in the Italian context. The maps of confounders for each country are shown in Figs. A.1–A.4, while Fig. A.5 presents their distributions across areal units.

3. Methods

3.1. Air pollution interpolation

A geostatistical model was fit to NO_2 monitoring data (2017–2019) from 517 ground stations in the US, creating spatial prediction (or Kriged) estimates on a fine grid covering the lower-48 states. Writing X_i as the measurement from the i th ground station, the model used is

$$X_i \sim N[\lambda(s_i), \tau^2]$$

$$\log[\lambda(\mathbf{s})] = W(\mathbf{s})\alpha + V(\mathbf{s}) \quad (1)$$

$$\text{cov}[V(\mathbf{s} + \mathbf{h}), V(\mathbf{s})] = \sigma^2 \tilde{\rho}(|\mathbf{h}|/\phi)$$

where $V(\mathbf{s})$ is a Gaussian random field with spatial correlation function $\tilde{\rho}(\cdot)$ and \mathbf{h} is the separation vector. The spatial covariates $W(\mathbf{s})$ include an intercept, gridded $\text{PM}_{2.5}$ data, and gridded NO_2 data from the most recent year 2011. A Matérn correlation function with shape parameter

1 was used for $\tilde{\rho}(\cdot)$. Maximum Likelihood Estimates of model parameters were obtained and spatial predictions were created using the R package *geostatsp* (Brown, 2015).

3.2. COVID-19 incidence model

We use a spatial ecological design to estimate the impact of air pollution on counts of COVID-19 cases in small areas. Writing Y_k , E_k and R_k as the observed count, age–sex adjusted expected count, and relative risk for area k , the model is given by

$$Y_k \sim \text{Poisson}(E_k R_k)$$

$$\ln(R_k) = \mathbf{X}_k^T \boldsymbol{\beta} + U_k \quad (2)$$

$$U_k | U_{\ell}; \ell \neq k \sim N \left[\frac{\rho \cdot \text{mean}\{U_j; j \sim k\}}{\rho + (1 - \rho)/N_k}, \frac{\sigma^2/N_k}{\rho + (1 - \rho)/N_k} \right]$$

where $j \sim k$ indicates the regions j which share a common border with k and $N_k = \|j \sim k\|$ (the number of such regions). Expected counts are calculated as

$$E_k = \sum_i \theta_i P_{ik} \quad (3)$$

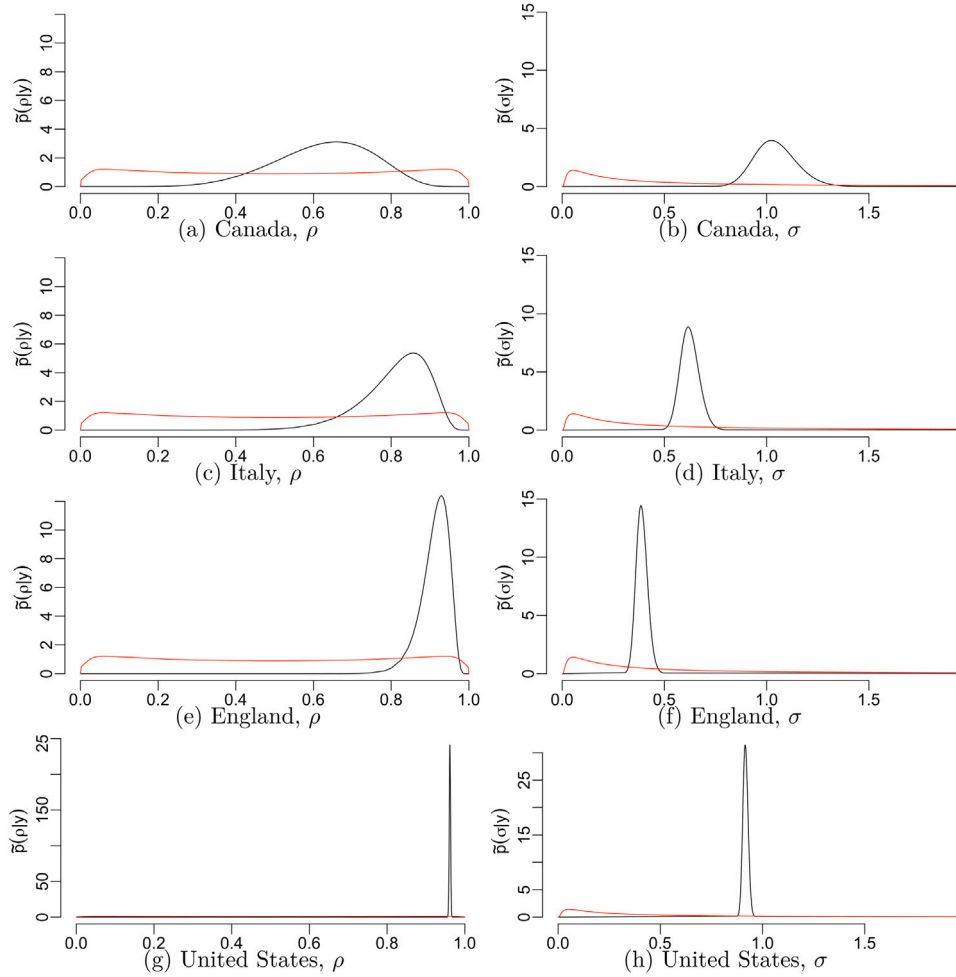


Fig. 2. Prior (red) and Posterior (black) for spatial dependence parameter ρ and spatial standard deviation parameter σ .

where P_{ik} is the population of the i th age–sex group in region k and θ_i are age–sex specific incidence rates estimated as the total observed cases in group i divided by the total population for group i .

The spatial random effect U_k follows a Leroux et al. (2000) model, with a spatial dependence parameter ρ and variance parameter σ^2 . When $\rho = 0$ the U_k are independent and Normally distributed with variance σ^2 , and the U_k surface will be rough in space with each region being unrelated to regions nearby. As ρ increases the value of U_k will become progressively more dependent on the mean of the U_j in neighbouring regions, and the surface becomes smoother. Setting $\rho = 1$ gives the standard spatial autoregressive model.

The explanatory variables in \mathbf{X}_k are ambient air pollutant concentrations and the confounders presented in the previous section. One interpretation of U_k is it accounts for unobserved confounders not included in the model, and we would expect $U_k = 0$ for area k if all the important risk factors were included in \mathbf{X}_k (and these effects were linear). The regression parameters β are assigned weakly informative zero-mean Gaussian priors with a large diagonal variance matrix $\beta \sim N(\mathbf{0}, 10^2 \mathbf{I})$.

The prior distributions for the spatial parameters are

$$\begin{aligned} \sigma &\sim \text{Exp}(\log 2) \text{ and} \\ \text{logit}(\rho) &\sim N(0, 1.8^2). \end{aligned} \quad (4)$$

The Exponential prior distribution of σ allows for, and encourages, small values and consequently a flat U_k surface. This would be expected if the covariates in \mathbf{X}_k were all correctly identified and accurately measured. The prior variance of 1.8^2 for ρ was chosen to make the prior for ρ reasonably uniform but discouraging the most extreme values.

The spatial models were implemented in the INLA software (Rue et al., 2009) and the Leroux CAR prior distribution for spatially structured random effects is available in R-INLA using the “besagpropr2” model. An alternative implementation of this prior distribution was first introduced in Ugarte et al. (2014) using the “generic1” model of R-INLA. The INLA, or Integrated Nested Laplace Approximation, methodology uses a computationally effective and extremely powerful alternative to implement Bayesian models, and is an increasingly popular analysis package in R. For any missing covariates, we treat the corresponding response as missing before fitting the INLA model. For details on how to fit spatial and spatio-temporal models with R and INLA, refer to Blangiardo et al. (2013).

3.3. Population attributable fraction

The population attributable fraction (or PAF, see Mansournia and Altman, 2018) provides a simple way of quantifying the combined impact of the two pollutants on COVID-19 incidence. For a given country, the PAF is defined as

$$\text{PAF} = 1 - \frac{\sum_k \exp(\tilde{\mathbf{X}}_k^\top \beta + U_k) E_k}{\sum_k R_k E_k}. \quad (5)$$

where $\tilde{\mathbf{X}}_k$ is the vector of covariates for region k where the values for pollutants have been replaced by reference values common across regions. Here the reference values are the 25% percentiles of $\text{PM}_{2.5}$ ($4.53 \mu\text{g m}^{-3}$) and NO_2 (2.74 ppb) observations across countries. By fitting model (2), the posterior samples for β and U_k are obtained, which then are used by (5) to obtain a posterior distribution of PAF to retain and propagate the uncertainty from the health model (2).

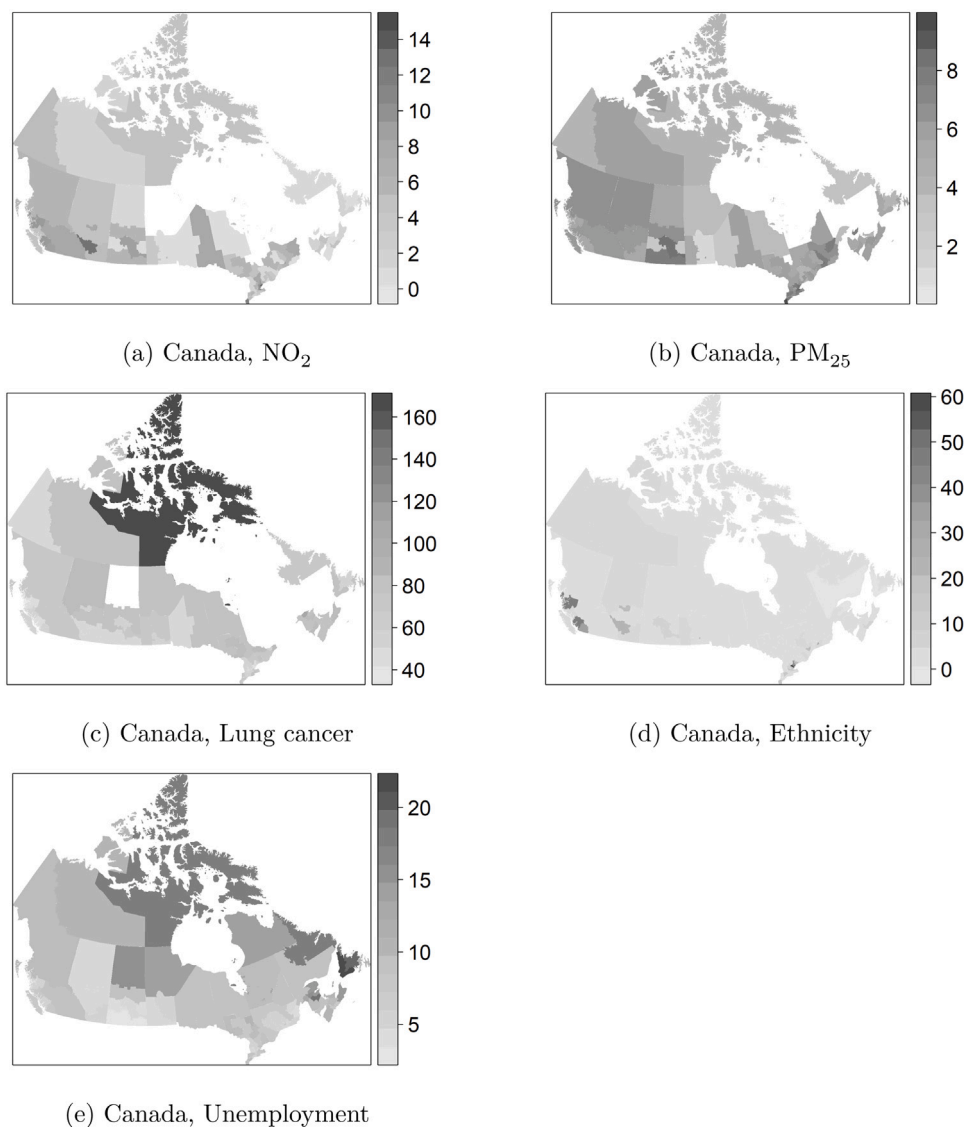


Fig. A.1. Canada data, including PM_{2.5} ($\mu\text{g}\cdot\text{m}^{-3}$), NO₂ (ppb), Lung cancer (per 100,000), Ethnicity (%) and Unemployment (%).

4. Results

4.1. Spatial distribution

In order to show comparable results across countries and small areas, we calculate the predicted incidence rate per 100,000 people, standardized to the EU standard population from Eurostat (2020). For area k , that is calculated by

$$\bar{R}_k = R_k \sum_i \theta_i \bar{P}_i,$$

where θ_i is the age–sex incidence rate from (3) and \bar{P}_i is the EU standard population for age–sex group i . Posterior means $E(\bar{R}_k | Y)$ for each country are shown on the left panels of Fig. 1. The right panels of Fig. 1 show the probabilities of 50% excess risk $\Pr(R_k > 1.5 | Y)$, which conveys some of the uncertainty in the estimates of R_k .

The predicted standardized incidence rate per 100,000 people in the US are much higher than the other countries, and some counties are as high as almost 25,000, or a quarter of the population. In contrast, its neighbour Canada has been more successful in controlling COVID-19, with the highest standardized incidence rate being about only 1500. Northern Italy has fairly high incidence, with the standardized incidence rate in northern provinces being about 3000 with high

probability. Similarly, the north west of England has high exceedance probability, with the standardized incidence rate in the upper tier local authorities being about 5000.

4.2. Air pollution effects

The main results of this study are presented in Table 2, which shows the relative risk based on a single unit increase in air pollution along with the spatial parameters and population attributable fraction. Note that the effects of PM_{2.5} are inconsistent across countries. We found some evidence of a protective effect of long-term exposure to PM_{2.5} in Canada, but with large uncertainty, while a harmful effect is present in the US, with more modest effects in Italy and England. The lower bound of the estimated US effect is well above the upper bounds for Canada, Italy and England. The NO₂ effects are more consistent, albeit very modest.

A 1% increase in visible minorities in a region increases COVID-19 incidence rate by 1.1%–2.8%. This adverse association is consistent with the results from the National Urban League (2020), which found the infection rates for black and Latino Americans are 62 and 73 per 10,000, respectively, compared with 23 per 10,000 for whites. As stated in National Urban League (2020), black and Latino Americans

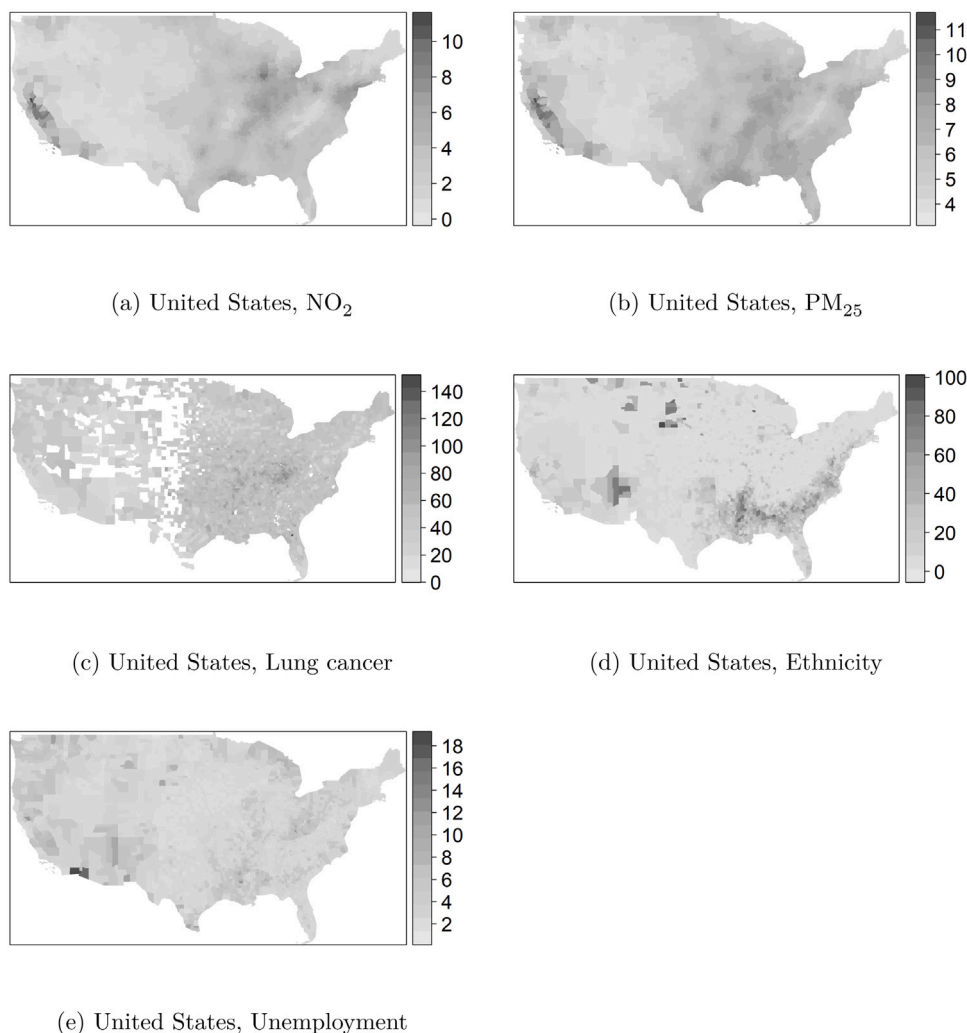


Fig. A.2. United States data, including PM_{2.5} (μgm^{-3}), NO₂ (ppb), Lung cancer (per 100,000), Ethnicity (%) and Unemployment (%).

are more likely to live in crowded housing conditions and to work in essential jobs which cannot be performed from home. In addition, Table 2 also suggests that the lung cancer mortality incidence rate was adversely associated with COVID-19 incidence in England, while the unemployment is an important confounder in Canada and the US.

The population attributable fractions from each country are shown in the last row of Table 2. The results suggest that if the long-term average PM_{2.5} drops to $4.53 \mu\text{gm}^{-3}$ and NO₂ drops to 2.74 ppb, the COVID-19 cases are likely to reduce 26.1% (95% CI 21.5%, 30.2%) for the US, 45.9% (95% CI 11.1%, 66.3%) for Italy and 10.3% (95% CI 1.6%, 19%) for England. The PAF from Canada study is neglectful with a wide credible interval, which is likely because of the estimate of PM_{2.5}. In addition, Table 2 shows that the PAF from the US is more certain than the other countries, as its 95% CI is much narrower, while the one from Italy is less certain with a very wide 95% CI. The highest magnitude estimate from Italy is mainly because its long-term PM_{2.5} and NO₂ levels (see Fig. A.5) are generally much higher than the PAF baseline levels.

Fig. 2 shows both the prior and the posterior distributions of the spatial dependence parameter ρ and standard deviation parameter σ from fitting the models in (2) to data from each country. Both parameters are well identified in all countries, although posterior distributions are narrowest for the US and its 3108 counties and widest for Canada's 93 health regions. The posteriors for ρ have little mass near zero, justifying the spatial random effect in the health model (2).

Table 2 also shows that the spatial dependence increases with the number of study areal units. For example, the estimated spatial parameter ρ is 0.6 for Canada while it is 1 for the US study. Higher value of ρ indicates a strong residual spatial autocorrelation after accounting for the known covariates.

For validation purposes, we ran an alternative commonly used convolution model proposed by Besag et al. (1991) with the parameterization from Riebler et al. (2016). We also performed a sensitivity analysis on the choice of prior distribution of model hyperparameters (σ, ρ) in the currently used Leroux CAR model. We considered exponential priors with $pr(\sigma > 1) = 0.2$ and $pr(\sigma > 1) = 0.8$. For ρ , we specified $\text{logit}(\rho) \sim N(0.75, 1.8^2)$ and $\text{logit}(\rho) \sim N(-0.75, 1.8^2)$ to represent either left or right shifted distributed of the prior, rather than the currently used 'rough uniform' distributed prior. The comparison results are presented in Table A.2, showing that the inferences are consistent across these various health models, which enhances the validity of the results.

5. Conclusion and discussion

Ogen (2020) reasons that a body experiencing chronic respiratory stress due to air pollution has a diminished ability to defend itself from infections. It is therefore a plausible hypothesis that long-term exposure to air pollution makes people more vulnerable to contracting COVID-19. We attempt to inform the debate about the long-term effects

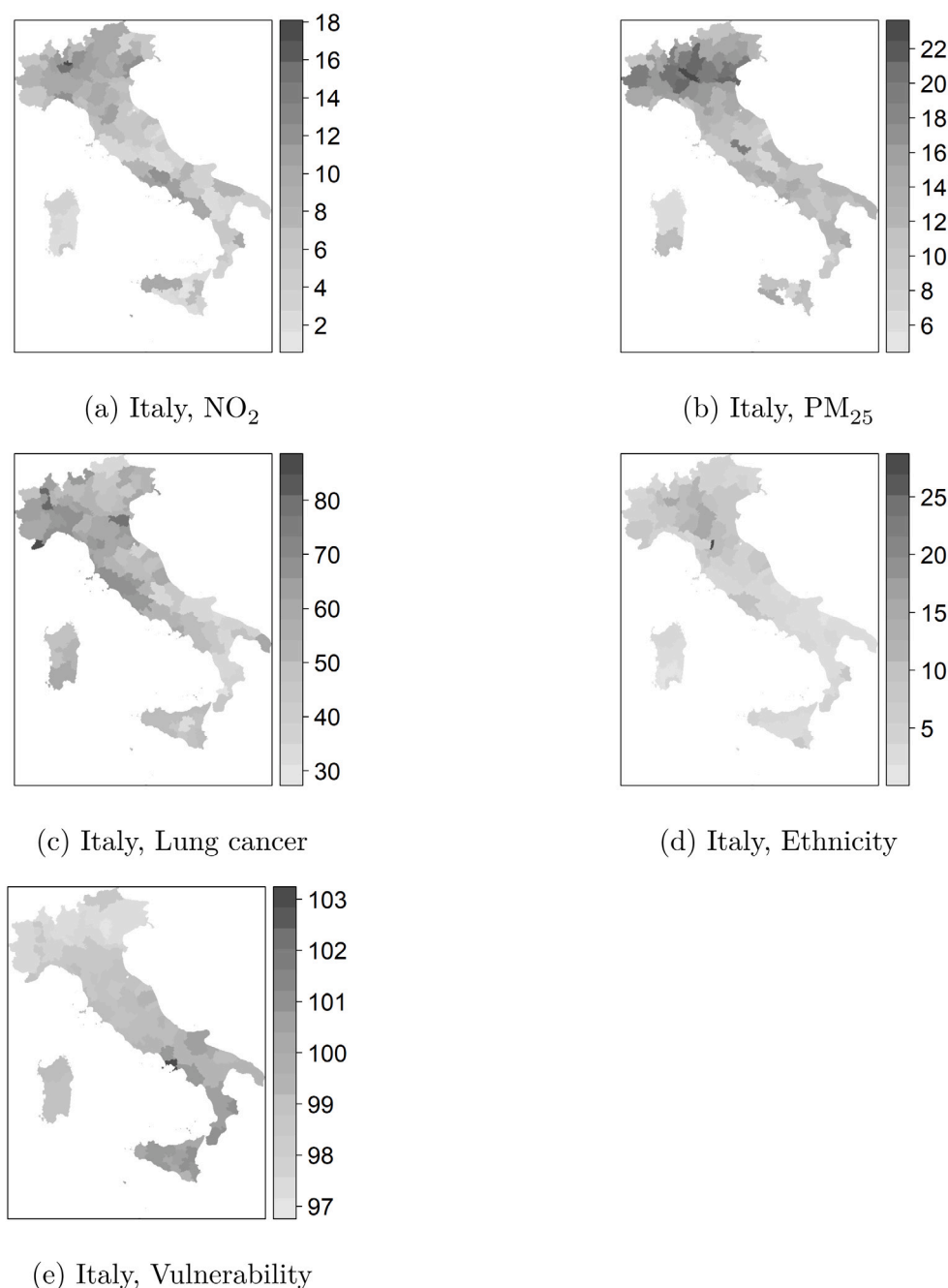


Fig. A.3. Italy data, including PM_{2.5} ($\mu\text{g}\cdot\text{m}^{-3}$), NO₂ (ppb), Lung cancer (per 100,000), Ethnicity (%) and Vulnerability.

of air pollution on COVID-19 by conducting a multi-country analysis, including Canada, Italy, England and the US.

The analysis presented in this paper has found that the inferred relationship between long-term ambient exposures to PM_{2.5} and NO₂ and COVID-19 incidence is inconsistent across the four countries examined. The results show that the relationship between air pollution exposure and COVID-19 incidence rate vary across countries, and a $1 \mu\text{g}\cdot\text{m}^{-3}$ increase in long-term exposure to PM_{2.5} increases the COVID-19 incidence rate by 12.6% (95% CI 7.4%, 18.1%) in the US, after adjusting for confounding and spatial autocorrelation, while the effects in Italy and England are more modest at 0.5% (95% CI -2.5%, 3.5%) and 2.9% (95% CI -0.7%, 6.6%), respectively, and in Canada the data provide some evidence of a protective effect of PM_{2.5}, albeit characterized by large uncertainty. This protective effect with a wide credible interval is likely caused by spatial confounding issues between

the spatial random effect and the explanatory variables included in the model, since the spatial confounding occurs when the covariates have a spatial pattern and are collinear with the spatial random effects (see, e.g., Adin et al., 2021; Hodges and Reich, 2010).

Our finding of the important impact of PM_{2.5} on COVID-19 in the US is consistent with a recent study by Wu et al. (2020), where they found that an increase of $1 \mu\text{g}\cdot\text{m}^{-3}$ in the long-term average PM_{2.5} was associated with an 11% (95% confidence interval 6%, 17%) increase in the COVID-19 mortality rate in the US. The finding of the modest (not significant) effects of NO₂ and PM_{2.5} in England tells the same story with the research by Konstantinou et al. (2021), where they only found some evidence of an effect for NO₂ and large uncertainty for PM_{2.5}. Similarly, an analysis performed by the UK's Office for National Statistics (2020) found that long-term exposure to PM_{2.5} could increase the risk of contracting and dying from COVID-19 by up to 7%,

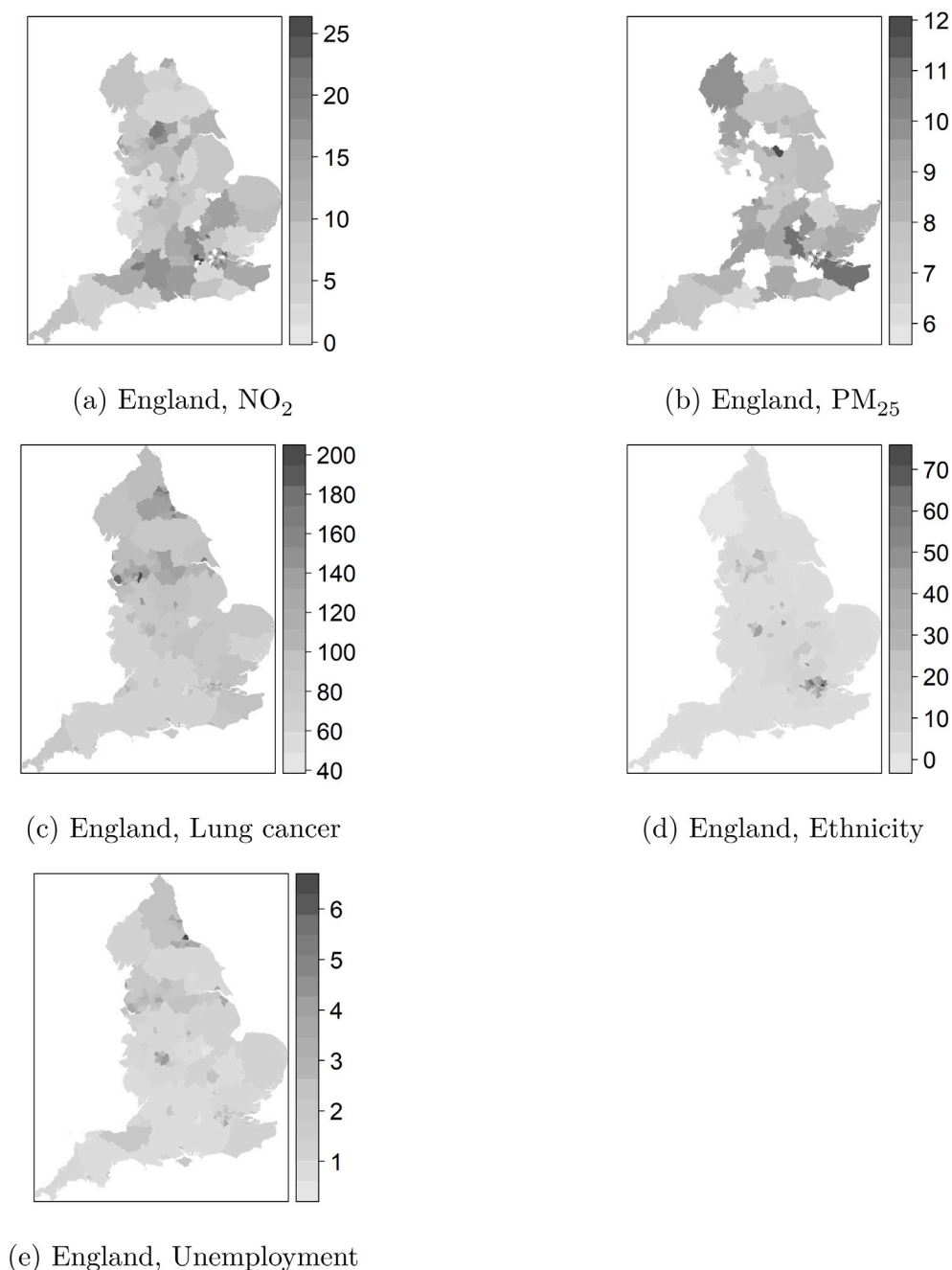


Fig. A.4. England data, including PM_{2.5} ($\mu\text{g m}^{-3}$), NO₂ (ppb), Lung cancer (per 100,000), Ethnicity (%) and Unemployment (%).

however when controlling for ethnicity, air pollution exposure had no statistically significant impact on COVID-19 deaths.

There are various reasons why COVID-19 prevalence varies across countries, and why socio-economic variables such as unemployment and ethnicity will influence COVID-19 prevalence more strongly in some countries than others. The biological hypothesis that air quality affects COVID-19 incidence through respiratory stress should, however, correspond to relative risks per unit increase being similar across countries. This lack of reproducibility should at a minimum lead the scientific community to view with some degree of caution the substantial relationships between air quality and COVID-19 reported by some studies. The population attributable fraction calculations in Table 2 estimate that in the US a quarter of cases can be ascribed to poor air quality, and the statistical error associated with this estimate is slight (95% CI 21%, 30%). If we were to take this result at face value, we

might conclude that lockdown measures can be relaxed in areas where air quality is high. The authors of this paper would not recommend such an action, as consistent and reproducible evidence supporting it is lacking.

One explanation for the inconsistent effects is the inherent limits of ecological studies where covariates refer to area-level, rather than individual-level, characteristics. Area-level measures are not always reliable proxies for individual-level exposures, individuals in a region are heterogeneous and the ecological fallacy can occur when there are complex dependencies and interactions at work at the individual-level. The relationship between individual-level and area-level exposures, and the measured values of the latter, could well be different in the four countries. The 3108 counties in the US provide higher-resolution spatial information than the 93–149 regions in the other countries, it is possible that the higher PM_{2.5} effects in the US result from a more

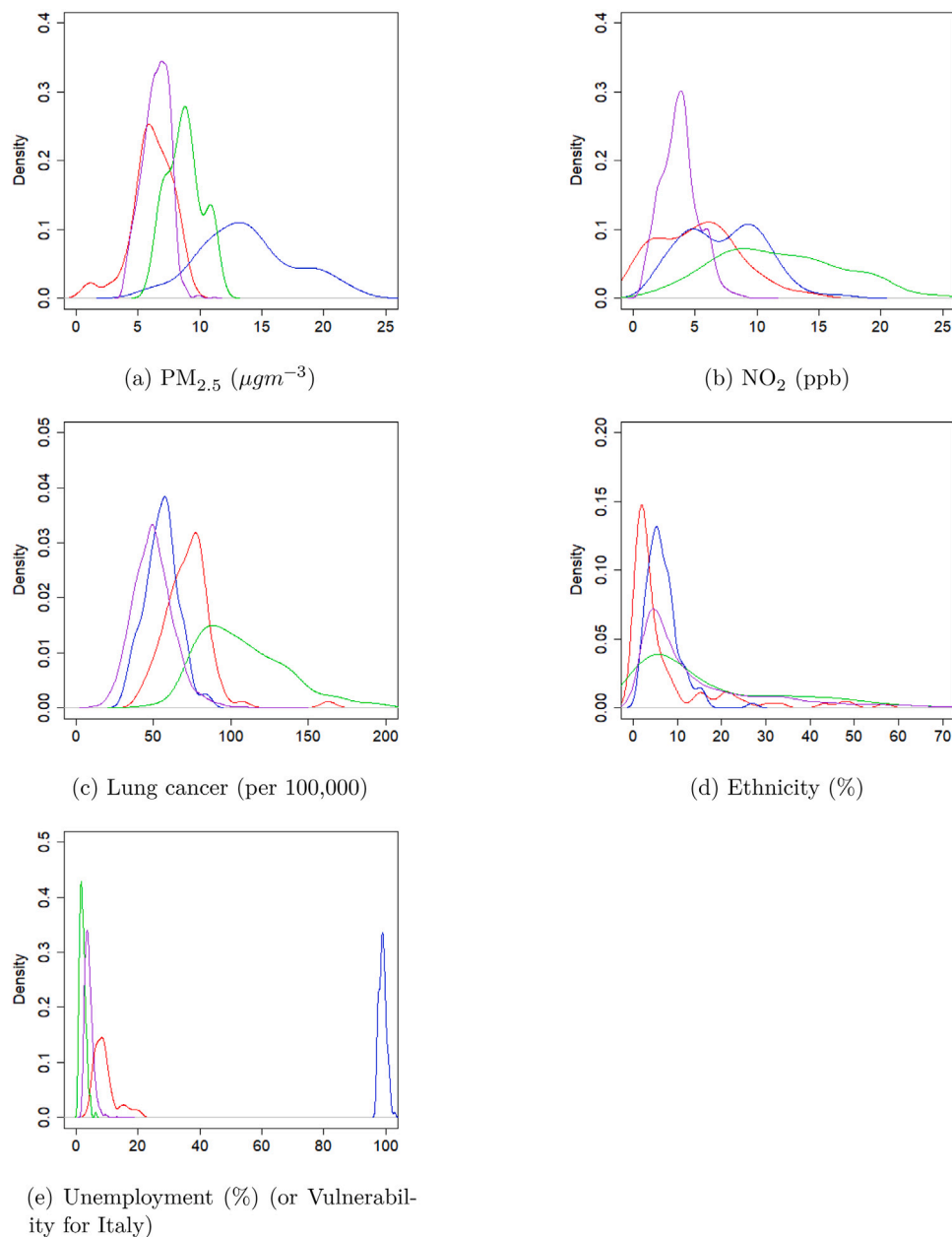


Fig. A.5. Empirical distributions of covariates, Canada (red), Italy (blue), England (green), United States (purple).

accurate exposure assessment and less attenuation of effects due to averaging.

Another explanation for the inconsistent results could be unmeasured confounders which are correlated with air quality to different degrees across countries. For example, $PM_{2.5}$ might be more strongly correlated with mobility and long-distance commerce in the more car-dependent United States than the other three countries. If mobility and commerce are in turn strong predictors of COVID-19 prevalence, failure to adjust for them as model covariates could cause a stronger inferred COVID-19/ $PM_{2.5}$ relationship in the US than elsewhere. Access to testing, adherence to control measures, and individuals in high-risk occupations might also be more strongly correlated with $PM_{2.5}$ in the US. Inferring causality from observational data is challenging and not always possible, and spatial observational data is particularly problematic (see Reich et al., 2020, for example). The lack of consistency in estimated $PM_{2.5}$ effect sizes across countries suggests that the inferred

relationships are not causal and additional unmeasured (and possibly individual-level) confounders are influencing the results.

This analysis has demonstrated the importance of replicating spatial analyses across multiple countries whenever possible. Spatial data on health outcomes and exposures is becoming increasingly available and the tools available for manipulating and managing these data make data acquisition a much simpler task than was the case previously. In the model building phase of an analysis, it is inevitable that results from several subtly different model formulations with different confounders will be known before a ‘final’ methodology is decided upon. Analysing several datasets with methodologies which are as close to identical as possible will help guard against overfitting and possible inherent bias towards positive results.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

The data sources used in this study are presented in Table A.1, while the results of prior sensitivity analysis of disease model are shown in Table A.2. The data description are shown as Figs. A.1–A.5.

Table A.1
Data sources in this study.

Covariate	Description	Sources
Canada		
COVID-19	Github, Berry et al. (2020)	https://github.com/ishaberry/Covid19Canada
PM _{2.5}	2014–2016, Environment Canada's NAPS; Health Canada	http://maps-cartes.ec.gc.ca/rmspa-naps/data.aspx?lang and Rural data
NO ₂	2014–2016, Environment Canada's NAPS; Health Canada	http://maps-cartes.ec.gc.ca/rmspa-naps/data.aspx?lang and Rural data
Lung cancer	2013/2015, Government of Canada	https://open.canada.ca/data/en/dataset/0112f88b-c08f-4a7a-8379-1c88b57c7412
Employment	2011, Statistics Canada	https://www150.statcan.gc.ca/n1/en/type/data?HPA=1
Ethnicity	2011, Statistics Canada	https://www150.statcan.gc.ca/n1/en/type/data?HPA=1
Population	2018, Statistics Canada	https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710013401
Shapefile	Scholars Portal Dataverse, Lucas (2020)	https://doi.org/10.5683/SP2/EIHG80
Italy		
COVID-19	R package covid19ita from Github	https://github.com/UBESP-DCTV/covid19ita/
PM _{2.5}	2016–2018, European Environment Agency	https://www.eea.europa.eu/data-and-maps/data/aqereporting-8
NO ₂	2016–2018, European Environment Agency	https://www.eea.europa.eu/data-and-maps/data/aqereporting-8
Lung cancer	Istituto Nazionale di Statistica	http://dati.istat.it/Index.aspx?DataSetCode=DCIS_CMORTE1_EV
Vulnerability	Istituto Nazionale di Statistica	http://ottomilacensus.istat.it/
Ethnicity	Istituto Nazionale di Statistica	http://dati.istat.it/Index.aspx?DataSetCode=DCIS_POPSTRRES1
Population	Istituto Nazionale di Statistica	http://demo.istat.it/index_e.html
Shapefile	Istituto Nazionale di Statistica	https://www.istat.it/it/archivio/222527
England		
COVID-19	GOV.UK	https://coronavirus.data.gov.uk/
PM _{2.5}	2016–2018, European Environment Agency	https://www.eea.europa.eu/data-and-maps/data/aqereporting-8
NO ₂	2016–2018, European Environment Agency	https://www.eea.europa.eu/data-and-maps/data/aqereporting-8
Lung cancer	2012–2016, Public Health England	https://www.localhealth.org.uk/#c=home
Employment	2017/2018, Public Health England	https://www.localhealth.org.uk/#c=home
Ethnicity	2011, Public Health England	https://www.localhealth.org.uk/#c=home
Population	2019, Office for National Statistics	https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration
Shapefile	Ministry of Housing, Communities and Local Government	http://data-communities.opendata.arcgis.com/search?q=Local%20Authority
United States		
COVID-19	USA FACTS	https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/
PM _{2.5}	2016–2018, Atmospheric Composition Analysis Group	https://sites.wustl.edu/acag/
NO ₂	2017–2019, EPA; 2009–2011, Atmospheric Composition Analysis Group	https://www.epa.gov/outdoor-air-quality-data ; https://sites.wustl.edu/acag/
Lung cancer	2012–2016, State Cancer Profiles	https://statecancerprofiles.cancer.gov/incidencrates/index.php
Employment	2018, Killeen et al. (2020)	https://www.kaggle.com/jieyingwu/covid19-us-countylevel-summaries
Ethnicity	2010, Killeen et al. (2020)	https://www.kaggle.com/jieyingwu/covid19-us-countylevel-summaries
Population	2019, USA FACTS	https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/
Shapefile	United States Census Bureau	https://www2.census.gov/geo/tiger/TIGER2019/COUNTY/

Table A.2
Prior sensitivity analysis, with estimates and 95% intervals for effect sizes with the prior from the manuscript, four additional priors, and an analysis with the convolution model parameterization of the spatial random effect from [Riebler et al. \(2016\)](#).

	Canada		Italy		England		United States	
	Est	CI	Est	CI	Est	CI	Est	CI
Prior from manuscript $pr(\sigma > 1) = 0.5, \text{logit}(\rho) \sim N(0, 1.8^2)$								
NO ₂	5.0	(−3.8, 14.7)	2.7	(−0.4, 6.0)	0.5	(−0.5, 1.6)	2.3	(−3.2, 8.1)
PM _{2.5}	−10.9	(−23.9, 4.5)	0.5	(−2.5, 3.5)	2.9	(−0.7, 6.6)	12.6	(7.4, 18.1)
Lung cancer	−0.2	(−2.3, 2.0)	0.0	(−0.9, 0.9)	0.7	(0.5, 1.0)	0.0	(−0.2, 0.2)
Unemployment	−19.0	(−25.1, −12.4)	−7.4	(−20.8, 8.6)	4.5	(−3.0, 12.7)	−4.3	(−6.0, −2.5)
Visible minorities	2.8	(0.5, 5.1)	2.3	(−0.2, 5.0)	1.1	(0.7, 1.5)	1.3	(1.1, 1.5)
$pr(\sigma > 1) = 0.2, \text{logit}(\rho) \sim N(0.75, 1.8^2)$								
NO ₂	4.9	(−3.6, 14.2)	2.5	(−0.5, 5.7)	0.5	(−0.4, 1.6)	2.5	(−3.1, 8.4)
PM _{2.5}	−11.8	(−24.4, 2.9)	0.4	(−2.5, 3.5)	3.0	(−0.5, 6.7)	12.5	(7.3, 18.0)
Lung cancer	0	(−2.1, 2.1)	−0.1	(−1.0, 0.7)	0.7	(0.5, 1.0)	0	(−0.2, 0.2)
Unemployment	−18.7	(−24.9, −12.1)	−4.2	(−18.0, 11.6)	4.2	(−3.2, 12.2)	−4.2	(−6.0, −2.5)
Visible minorities	2.9	(0.7, 5.1)	2.2	(−0.3, 4.8)	1.1	(0.7, 1.6)	1.3	(1.1, 1.5)
$pr(\sigma > 1) = 0.2, \text{logit}(\rho) \sim N(−0.75, 1.8^2)$								
NO ₂	5.1	(−3.9, 15.2)	3.0	(−0.2, 6.3)	0.5	(−0.5, 1.6)	2.1	(−3.4, 7.9)
PM _{2.5}	−9.7	(−23.1, 6.2)	0.5	(−2.6, 3.6)	2.8	(−0.9, 6.5)	12.8	(7.5, 18.2)
Lung cancer	−0.3	(−2.5, 1.9)	0.1	(−0.8, 1.0)	0.7	(0.5, 1.0)	0.0	(−0.2, 0.2)
Unemployment	−19.2	(−25.1, −12.8)	−11.0	(−23.5, 4.4)	4.8	(−2.8, 13.3)	−4.3	(−6.0, −2.5)
Visible minorities	2.7	(0.4, 5.0)	2.5	(−0.1, 5.2)	1.1	(0.6, 1.5)	1.3	(1.1, 1.5)
$pr(\sigma > 1) = 0.8, \text{logit}(\rho) \sim N(0.75, 1.8^2)$								
NO ₂	4.9	(−3.7, 14.4)	2.5	(−0.5, 5.7)	0.5	(−0.4, 1.6)	2.5	(−3.1, 8.3)
PM _{2.5}	−11.8	(−24.6, 3.2)	0.4	(−2.5, 3.5)	3.0	(−0.5, 6.7)	12.5	(7.3, 18.0)

(continued on next page)

Table A.2 (continued).

	Canada		Italy		England		United States	
	Est	CI	Est	CI	Est	CI	Est	CI
Lung cancer	0.0	(-2.1, 2.1)	-0.1	(-1.0, 0.8)	0.7	(0.5, 1.0)	0.0	(-0.2, 0.2)
Unemployment	-18.7	(-25.0, -12.0)	-4.2	(-18.0, 11.6)	4.2	(-3.2, 12.3)	-4.2	(-6.0, -2.5)
Visible minorities	2.9	(0.6, 5.1)	2.2	(-0.3, 4.8)	1.1	(0.7, 1.6)	1.3	(1.1, 1.5)
<i>pr</i> ($\sigma > 1$) = 0.8, <i>logit</i> (ρ) ~ <i>N</i> (-0.75, 1.8 ²)								
NO ₂	5.2	(-3.9, 15.2)	3.0	(-0.2, 6.3)	0.5	(-0.5, 1.6)	2.2	(-3.3, 8.0)
PM _{2.5}	-9.9	(-23.3, 6.0)	0.5	(-2.6, 3.6)	2.8	(-0.9, 6.5)	12.7	(7.5, 18.2)
Lung cancer	-0.3	(-2.4, 1.9)	0.1	(-0.8, 1.0)	0.7	(0.5, 1.0)	0.0	(-0.2, 0.2)
Unemployment	-19.2	(-25.2, -12.7)	-10.9	(-23.7, 4.9)	4.8	(-2.8, 13.3)	-4.3	(-6.0, -2.5)
Visible minorities	2.7	(0.4, 5.0)	2.5	(-0.1, 5.2)	1.1	(0.6, 1.5)	1.3	(1.1, 1.5)
Convolution model parameterization								
NO ₂	3.9	(-6.0, 15.2)	3.9	(0.7, 7.3)	0.6	(-0.4, 1.7)	3.2	(-2.3, 8.9)
PM _{2.5}	-8.6	(-23.7, 9.2)	1.9	(-1.2, 4.9)	3.3	(-0.3, 7.0)	12.0	(6.9, 17.4)
Lung cancer	0.1	(-2.2, 2.5)	0.0	(-0.9, 0.9)	0.7	(0.4, 0.9)	0.0	(-0.2, 0.2)
Unemployment	-20.6	(-26.9, -14.0)	-17.0	(-24.4, -8.6)	3.1	(-4.3, 10.9)	-4.1	(-5.8, -2.4)
Visible minorities	2.3	(-0.2, 4.9)	3.2	(0.5, 6.0)	1.2	(0.8, 1.7)	1.3	(1.1, 1.5)

References

- Adin, A., Goicoa, T., Hodges, J.S., Schnell, P.M., Ugarte, M.D., 2021. Alleviating confounding in spatio-temporal areal models with an application on crimes against women in India. *Stat. Model.* 1471082X2110154.
- AEA, 2011. UK modelling under the Air Quality Directive (2008/50/EC) for 2010 covering the following air quality pollutants: SO₂, NO_x, NO₂, PM₁₀, PM_{2.5}, lead, benzene, CO and ozone. http://uk-air.defra.gov.uk/assets/documents/reports/cat09/1204301513_AQD2010mapsrep_master_v0.pdf.
- Andree, B.P.J., 2020. Incidence of COVID-19 and connections with air pollution exposure: Evidence from the Netherlands. *MedRxiv*. Cold Spring Harbor Laboratory Press, URL: <https://www.medrxiv.org/content/early/2020/05/03/2020.04.27.20081562>.
- Berry, I., Soucy, J.-P.R., Tuite, A., Fisman, D., 2020. Open access epidemiologic data and an interactive dashboard to monitor the COVID-19 outbreak in Canada. *Canad. Med. Assoc. J.* 192 (15), E420.
- Besag, J., York, J., Mollie, A., 1991. Bayesian image restoration with two applications in spatial statistics. *Ann. Inst. Statist. Math.* 43, 1–59.
- Blangiardo, M., Cameletti, M., Baio, G., Rue, H., 2013. Spatial and spatio-temporal models with R-INLA. *Spat. Spatio-Temporal Epidemiol.* 4, 33–49.
- Brown, P.E., 2015. Model-based geostatistics the easy way. *J. Stat. Softw.* 63 (12), 1–24.
- Cole, M.A., Ozgen, C., Strobl, E., 2020. Air Pollution Exposure and Covid-19 in Dutch Municipalities. *Environ. Resour. Econ.* 76 (4), 581–610.
- Copat, C., Cristaldi, A., Fiore, M., Grasso, A., Zuccarello, P., Signorelli, S., Conti, G., Ferrante, M., 2020. The role of air pollution (PM and NO₂) in COVID-19 spread and lethality: A systematic review. *Environ. Res.* 191, 110129.
- Eurostat, 2020. EU standard population (based on the EU and EFTA 2011-30 population projections). URL: https://ec.europa.eu/eurostat/cache/metadata/Annexes/hlth_cdeath_esms_an1.pdf.
- Gautam, S., 2020. COVID-19: air pollution remains low as people stay at home. *Air Qual. Atmosp. Health* 13.
- Gautam, S., Samuel, C., Gautam, A.S., Kumar, S., 2021. Strong link between coronavirus count and bad air: a case study of India. *Environ. Dev. Sustainab.*
- Atmospheric Composition Analysis Group, 2021. Surface PM_{2.5}. <https://sites.wustl.edu/acag/datasets/surface-pm2-5/>.
- Gupta, A., Bherwani, H., Gautam, S., Anjum, S., Musugu, K., Kumar, N., Anshul, A., Kumar, R., 2021. Air pollution aggravating COVID-19 lethality? Exploration in Asian cities using statistical models. *Environ. Dev. Sustainab.* 23, 6408–6417.
- 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 (2-3), 123–131.
- Hodges, J.S., Reich, B.J., 2010. Adding spatially-correlated errors can mess up the fixed effect you love. *Amer. Statist.* 64 (4), 325–334.
- Huang, G., Lee, D., Scott, E.M., 2018. Multivariate space-time modelling of multiple air pollutants and their health effects accounting for exposure uncertainty. *Stat. Med.* 37 (7), 1134–1148.
- Janes, H., Dominici, F., Zeger, S.L., 2007. Trends in air pollution and mortality: an approach to the assessment of unmeasured confounding. *Epidemiology* 18 (4), 416–423.
- Killeen, B.D., Wu, J.Y., Shah, K., Zapaishchykova, A., Nikutta, P., Tamhane, A., Chakraborty, S., Wei, J., Gao, T., Thies, M., Unberath, M., 2020. A county-level dataset for informing the United States' response to COVID-19. *arXiv:2004.00756 [cs.CY]*.
- 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.
- 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. 179–191, Chap. Statistical Models in Epidemiology, the Environment and Clinical Trials, Halloran, M and Berry, D (eds).
- Lucas, J., 2020. COVID-19 Canada Health Regions Shapefile. *Scholars Portal Dataverse*, <http://dx.doi.org/10.5683/SP2/EIHG80>, (visited on 06/08/2021).
- Mansournia, M.A., Altman, D.G., 2018. Population attributable fraction. *BMJ* 360.
- National Urban League, 2020. 2020 State of Black America: Unmasked. A National Urban League Publication, URL: https://soba.iamempowered.com/?_ga=2.182044386.470730837.1597941243-1376958794.1597941243.
- Office for National Statistics, 2020. Does exposure to air pollution increase the risk of dying from the coronavirus (COVID-19)?
- Ogen, Y., 2020. Assessing nitrogen dioxide (NO₂) levels as a contributing factor to coronavirus (COVID-19) fatality. *Sci. Total Environ.* 726, 138605.
- Pope, I.C., Burnett, R., Thun, M., et al., 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA* 287 (9), 1132–1141.
- Pozzer, A., Dominici, F., Haines, A., Witt, C., Munzel, T., Lelieveld, J., 2020. Regional and global contributions of air pollution to risk of death from COVID-19. *Cardiovasc. Res.* 116, 2247–2253.
- Reich, B.J., Yang, S., Guan, Y., Giffin, A.B., Miller, M.J., Rappold, A.G., 2020. A review of spatial causal inference methods for environmental and epidemiological applications. *arXiv:2007.02714 [stat.ME]*.
- Riebler, A., Sørbye, S.H., Simpson, D., Rue, H., 2016. An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Stat. Methods Med. Res.* 25 (4), 1145–1165, PMID: 27566770.
- Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 71 (2), 319–392.
- Ugarte, M., Adin, A., Goicoa, T., Militino, A., 2014. On fitting spatio-temporal disease mapping models using approximate Bayesian inference. *Stat. Methods Med. Res.* 23 (6), 507–530.
- Wu, X., Nethery, R.C., Sabath, M.B., 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 (45).