Long-term exposure to air pollution and COVID-19 incidence: a multi-country study

17 Nov 2020

Abstract

Recently, 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 their inference regarding the effects of air pollution. In this study, we attempt to inform the debate about the long-term effect of air pollution on COVID-19 outcomes by looking at COVID-19 cases across Canada, Italy, England and the United States. A spatial ecological design was used to estimate the impact of air pollution on cases using contiguous small areas in each Country. We found that the air pollution effects vary across countries. The results from the US and Italy show that a 1 μgm^{-3} increase in long-term exposure to $PM_{2.5}$ increases the COVID-19 incidence rate by 12.6% (95% credible interval[CI] (7.4%, 18.1%) and 4.5% (95% CI (-0.1%, 9.2%), respectively, after adjusting for confounding and spatial autocorrelation. Additionally, we found that each 1% increase in visible minorities in a region increases COVID-19 incidence rate by 1.1%-2.8%.

Keywords: coronavirus disease, spatial model, epidemiology, multi-country, INLA

1 Introduction

The current outbreak of coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has led to 39 million cases and 1098 thousand deaths worldwide up to 16th Oct, 2020. Since the 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 effects on COVID-19 death/incidence rates.

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 micrometers) and NO_2 (nitrogen dioxide) 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, a recent study by Wu et al. (2020) found that an increase of 1 μgm^{-3} in the long-term average $PM_{2.5}$ was associated with a 11% (95% confidence interval, 6%, 17%) increase in the COVID-19 mortality rate in the US, while a research in England by Konstantinoudis et al. (2020) only found some evidence of an effect for NO_2 and large uncertainty for $PM_{2.5}$. 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%, however when controlling for ethnicity, air pollution exposure had no statistically significant impact on COVID-19 deaths. 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.

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 US. We focus on 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) people who have been long-term exposed to high pollution could have reduced natural defenses in their bodies, which makes it easier for COVID-19 or any other infection to be invasive. Specifically, for each country, we investigate whether long-term average exposure to air pollution ($PM_{2.5}$ 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 the potential confounders, including lung cancer incidence, unemployment rate and visible 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 of data 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-Data sources. 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 mainland United States, respectively. There are 219 counties in the United States having population greater than 300 thousand. More details are given in Table 1.

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 μgm^{-3} for $PM_{2.5}$, 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**	**USA**
Population	36.3	60.2	55.5	314.4
Areal units	93	107	149	3108
Total cases	194	373	560	7784
$PM_{2.5}$	(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)

2.1 COVID-19 data

For each country, the disease data comprise of the total confirmed COVID-19 cases in each areal unit (e.g., county for England, province for Italy). We denote Y_k as the reported number of COVID-19 cases for the kth 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 based on national incidence rate in sex_age groups by (3). We use standardized incidence ratio (SIR) given by $SIR_k = Y_k/E_k$, to measure the risk of disease, (for example, a SIR of 1.2 indicates a 20% increased risk of disease compared to that expected).

2.2 Air pollution data

In this study we consider $PM_{2.5}$ and NO_2 coming from two data sources: measurements from a monitoring network (Canada, Italy, England) and gridded modeled output (US). For Canada, Italy and England, the areal concentrations are estimated by taking the mean of all the measurements from those stations lying within a specific areal unit. As there are several areal units which are extremely small, we use a 5 kilometer buffer for assigning stations to units. For the $PM_{2.5}$ exposures in the US, areal concentrations are estimated by taking the mean of all the gridded estimates lying within a specific areal unit. The average of 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.

The NO_2 exposures in the US are obtained by interpolating ground monitoring data from 2017-2019, as the most recent gridded NO_2 data published dates from 2011. A geostatistical spatial model is fitted to ground NO_2 measurements, with the 2009-2011 gridded NO_2 and 2016-2018 gridded $PM_{2.5}$ as predictors.

Maps of $PM_{2.5}$ and NO_2 for each country are shown in Figure A.1 and A.2, while Figure A.3 presents their distributions across areal units. The latter shows that the long-term averaged $PM_{2.5}$ 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, ii) unemployment rate as a proxy for socio-economic deprivation, iii) the proportion of the population who belong to minority groups. The minority groups differed across countries, as: visible minorities, as defined by the Census, in Canada; individuals from Africa or Asia for Italy; the black and minority ethnic population for England; and the non-white population for the US. In addition, in Italy the index of material and social vulnerability was used in place of the unemployment, the former regarded as the better proxy of socio-economic deprivation. The maps of confounders for each country are shown in Figure A.1 and A.2, while Figure A.3 presents their distributions across areal units.

3 Methods

3.1 US NO_2 interpolation

A geostatistical model was fit to NO_2 monitoring data from 517 ground stations in the US, creating spatial predictions (or Kriged) estimates on a fine grid covering the lower-48

states. Writing X_i as the measurement from the ith ground station, the model used is

$$\begin{aligned} X_i \sim & \mathrm{N}[\lambda(s_i), \tau^2] \\ \log[\lambda(s)] = & W(s)\alpha + V(s) \\ \mathrm{cov}[V(s+h), V(s)] = & \sigma^2 \tilde{\rho}(|h|/\phi) \end{aligned} \tag{1}$$

where $V(s)$ is a Gaussian random field with spatial correlation function $\tilde{\rho}(\cdot)$. The spatial covariates $W(s)$ include an intercept, gridded $PM_{2.5}$ data, and gridded NO_2 data from 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 created using the R package `geostatsp` (Brown 2019; Brown 2015).

3.2 Statistical 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

$$\begin{aligned} Y_k \sim & \mathrm{Poisson}(E_k R_k) \\ \ln(R_k) = & X_k^\top \beta + U_k \\ U_k | U_\ell; \ell \neq k \sim & \mathrm{N}\left[\frac{\rho \cdot \mathrm{mean}\{U_j; j \sim k\}}{\rho + (1-\rho)/N_k}, \frac{\sigma^2/N_k}{\rho + (1-\rho)/N_k} \right] \end{aligned} \tag{2}$$

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

$$E_k = \sum_i \theta_i P_{ik} \tag{3}$$

where P_{ik} is the population of the ith 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 depen-

dence 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 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 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 $\beta \sim \mathrm{N}\left(0, 10^2 I\right)$.

The prior distributions for the spatial parameters are

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

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 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 its `besagproper2` model. 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 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

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

where $\tilde{X}_k$ 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 $PM_{2.5}$ (4.53 μgm^{-3}) and NO_2 (2.74 ppb) observations across countries.

4 Results

4.1 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 fractions. 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 and England and only slightly below the upper bound in Italy. The NO_2 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 rate for blacks is 62 per 10,000, compared with 23 per 10,000 for whites. In addition, Table 2 also suggests that the lung cancer mortality incidence rate was adversely associated with COVID-19 incidence in England, while the

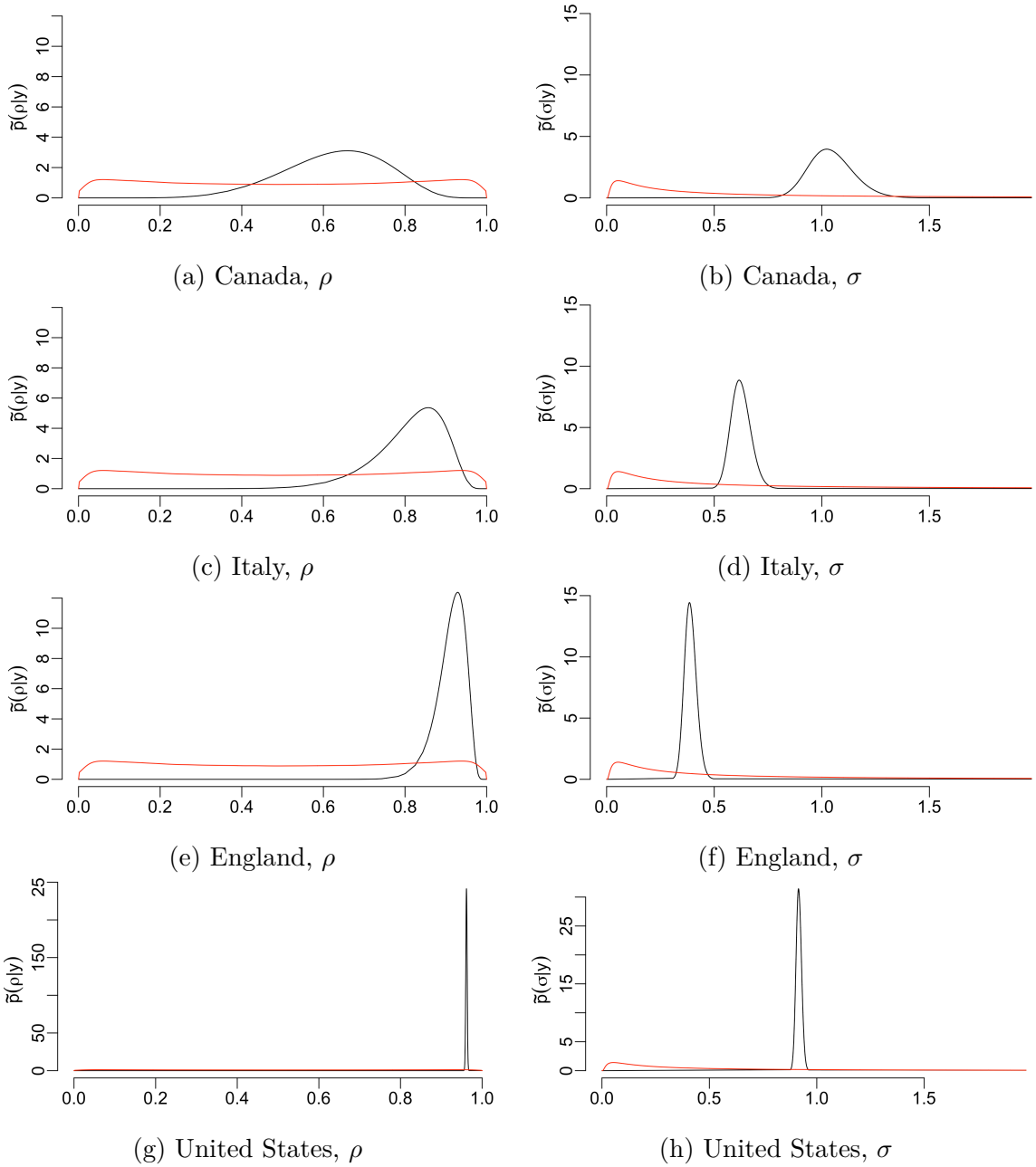
Table 2: Estimates and 95% intervals for effect sizes, spatial parameters and population-attributable fraction of incidence. Effects for $PM_{2.5}$, NO_2, lung cancer incidence, percent unemployed, and percent ethnic minorities are percent increase in relative risk for a one unit increase, a value of 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)
$PM_{2.5}$	-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)

unemployment is an important confounder in Canada, Italy 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 μgm^{-3} and NO_2 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 beneficial 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 NO_2 levels (see Figure A.3) are generally much higher than the PAF baseline levels.

Figure 1 shows both the prior and the posterior distributions of the spatial dependence parameter ρ and variance 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 health model (2).



(a) Canada, ρ (b) Canada, σ

(c) Italy, ρ (d) Italy, σ

(e) England, ρ (f) England, σ

(g) United States, ρ (h) United States, σ

Figure 1: Prior (red) and Posterior (black) for spatial dependence parameter ρ and spatial standard deviation parameter σ.

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. The estimate of spatial standard deviation σ in the table suggests that the random effects from England are much smaller than the others.

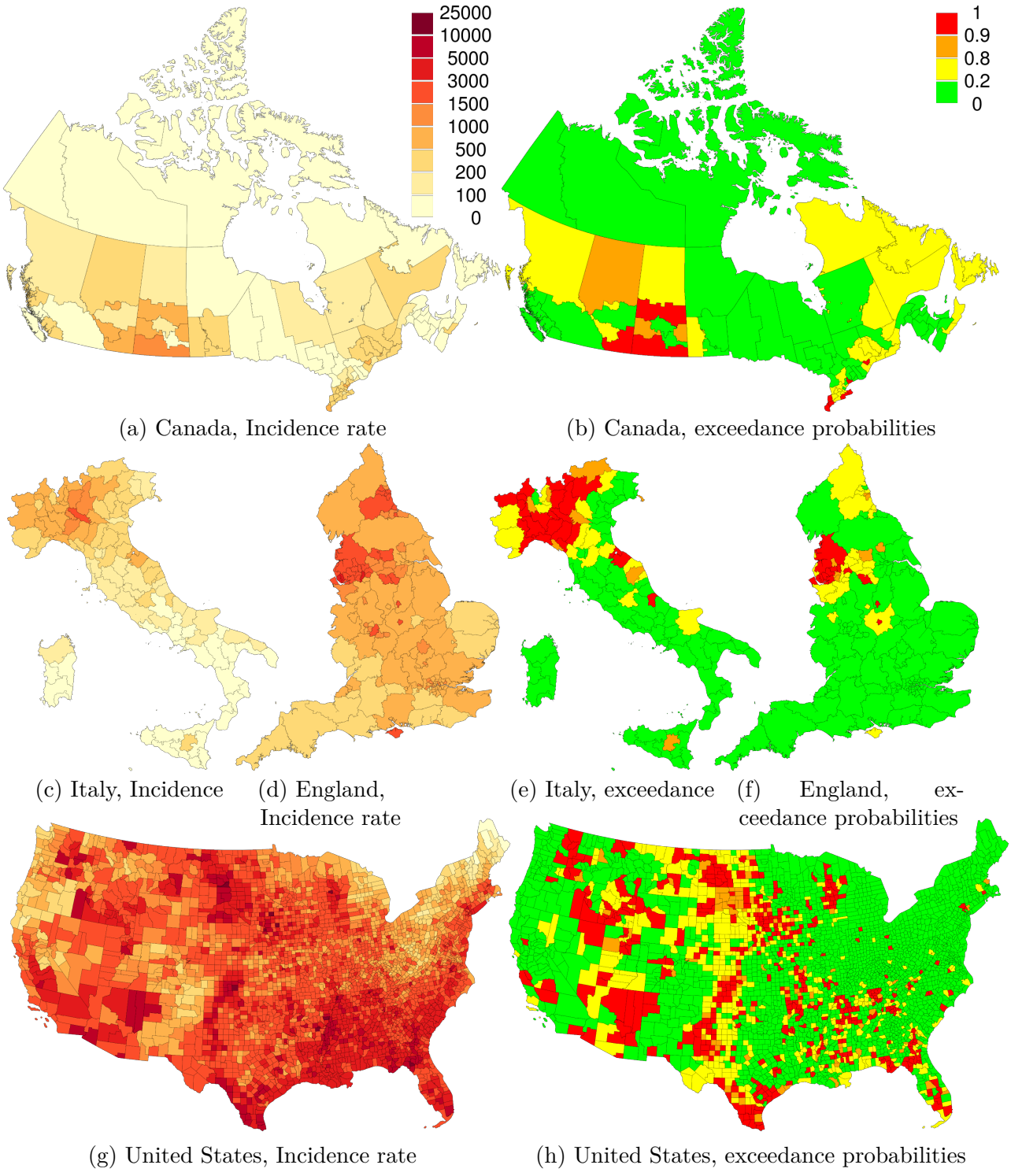
4.2 Spatial distribution

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 $\mathrm{E}\left(\bar{R}_k \mid Y\right)$ for each country are shown on the left panels of Figure 2. The right panels of Figure 2 show the probabilities of 50% excess risk $\Pr(R_k > 1.5 \mid Y)$, 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 1,500. Northern Italy has fairly high incidence, with the standardized incidence rate in northern provinces being about 3,000 with high probability. Similarly, the north west of England has high exceedance probability, with the standardized incidence rate in provinces being about 5,000.



(a) Canada, Incidence rate

(b) Canada, exceedance probabilities

(c) Italy, Incidence

(d) England, Incidence rate

(e) Italy, exceedance

(f) England, exceedance probabilities

(g) United States, Incidence rate

(h) United States, exceedance probabilities

Figure 2: Incidence rate per 100,000 people $\mathrm{E}\left(\bar{R}_k \mid Y\right)$, standardized to the EU standard population and probabilities of 50 % excess risk $\Pr(R_k > 1.5 \mid Y)$ for the four countries in the study.

5 Conclusion

The existing body of research on the impacts of air pollution on human health has linked $PM_{2.5}$ and NO_2 exposure to health damage, particularly respiratory (Bowatte et al. 2017; Lee et al. 2009) and lung diseases (Yu et al. 2003; Smith 2010; Järup et al. 1998; Das et al. 2008). Ogen (2020) reason that a body experiencing chronic respiratory stress due to air pollution has a diminished ability to defend itself from infections. It is therefore sensible hypothesis that long-term exposure to air pollution makes people more vulnerable to contracting COVID-19.

The analysis presented in this paper has found that the inferred relationship between long-term ambient exposures to $PM_{2.5}$ and NO_2 and COVID-19 incidence is inconsistent across the four countries examined. The results show that the air pollution health effects vary across countries and a 1 μgm^{-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%) and 4.5% (95% CI -0.1%, 9.2%) in the US and Italy, respectively, after adjusting for confounding and spatial autocorrelation, while the effect in England is more modest at 2.9% (-0.7%, 6.6%) and in Canada the data provide some evidence of a protective effect of $PM_{2.5}$, albeit characterised by large uncertainty.

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% to 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 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.

Acknowledgements

P. E. Brown is funded by the Natural Sciences and Engineering Research Council of Canada.

References

Andree, B. P. J. (2020). "Incidence of COVID-19 and Connections with Air Pollution Exposure: Evidence from the Netherlands". In: *medRxiv*. URL: https://www.medrxiv.org/content/early/2020/05/03/2020.04.27.20081562.

Berry, I., J.-P. Soucy, A. Tuite, and D. Fisman (2020). "Open access epidemiologic data and an interactive dashboard to monitor the COVID-19 outbreak in Canada". In: *Canadian Medical Association Journal*.

Blangiardo, M., M. Cameletti, G. Baio, and H. Rue (2013). "Spatial and spatio-temporal models with R-INLA". In: *Spatial and Spatio-temporal Epidemiology* 4, pp. 33–49.

Bowatte, G., B. Erbas, C. J. Lodge, L. D. Knibbs, L. C. Gurrin, G. B. Marks, P. S. Thomas, D. P. Johns, G. G. Giles, J. Hui, M. Dennekamp, J. L. Perret, M. J. Abramson, E. H. Walters, M. C. Matheson, and S. C. Dharmage (2017). "Traffic-related air pollution exposure over a 5-year period is associated with increased risk of asthma and poor lung function in middle age". In: *European Respiratory Journal* 50.4.

Brown, P. E. (2015). "Model-based geostatistics the easy way". In: *Journal of Statistical Software* 63.12, pp. 1–24.

Brown, P. E. (2019). *Geostatistical Modelling with Likelihood and Bayes*. R package version 1.7.8. URL: https://CRAN.R-project.org/package=geostatsp.

Cole, M., C. Ozgen, and E. Strobl (2020). "Air Pollution Exposure and Covid-19 in Dutch Municipalities". In: *Environ Resour Econ (Dordr)*, pp. 1–30.

Copat, C., A. Cristaldi, M. Fiore, A. Grasso, P. Zuccarello, S. Signorelli, G. Conti, and M. Ferrante (2020). "The role of air pollution (PM and NO_2) in COVID-19 spread and lethality: A systematic review". In: *Environmental research*.

Das, K., S. Das, and S. Dhundasi (Oct. 2008). "Nickel, its adverse health effects & oxidative stress". In: *The Indian journal of medical research* 128.4, pp. 412–425.

Eurostat (2020). "EU Standard Population (based on the EU and EFTA 2011-30 population projections)". In: URL: https://ec.europa.eu/eurostat/cache/metadata/Annexes/hlth_cdeath_esms_an1.pdf.

Haining, R., G. Li, R. Maheswaran, M. Blangiardo, J. Law, N. Best, and S. Richardson (2010). "Inference from ecological models: Estimating the relative risk of stroke from air pollution exposure using small area data". In: *Spatial and Spatio-temporal Epidemiology* 1.2 - 3, pp. 123–131.

Huang, G., D. Lee, and E. M. Scott (2018). "Multivariate space-time modelling of multiple air pollutants and their health effects accounting for exposure uncertainty". In: *Statistics in Medicine* 37.7, pp. 1134–1148.

Janes, H., F. Dominici, and S. Zeger (2007). "Trends in Air Pollution and Mortality: An Approach to the Assessment of Unmeasured Confounding". In: *Epidemiology* 18.4, pp. 416–423.

Järup, L., M. Berglund, C. G. Elinder, G. Nordberg, and M. Vanter (1998). "Health effects of cadmium exposure – a review of the literature and a risk estimate". In: *Scandinavian Journal of Work, Environment & Health* 24, pp. 1–51.

Killeen, B. D., J. Y. Wu, K. Shah, A. Zapaishchykova, P. Nikutta, A. Tamhane, S. Chakraborty, J. Wei, T. Gao, M. Thies, and M. Unberath (Apr. 2020). "A County-Level Dataset for Informing the United States' Response to COVID-19". In:

Konstantinoudis, G., T. Padellini, J. E. Bennett, B. Davies, M. Ezzati, and M. Blangiardo (2020). "Long-term exposure to air-pollution and COVID-19 mortality in England: a hierarchical spatial analysis". In: *medRxiv*. eprint: https://www.medrxiv.org/content/early/2020/08/11/2020.08.10.20171421.full.pdf.

Lee, D., C. Ferguson, and R. Mitchell (2009). "Air pollution and health in Scotland: a multicity study". In: *Biostatistics* 10.3, pp. 409–423.

Leroux, B., X. Lei, and N. Breslow (2000). "Estimation of disease rates in small areas: A new mixed model for spatial dependence". In: Springer-Verlag, New York. Chap. Statistical Models in Epidemiology, the Environment and Clinical Trials, Halloran, M and Berry, D (eds), pp. 179–191.

Lucas, J. (2020). "COVID-19 Canada Health Regions Shapefile". Version V3. In: URL: https://doi.org/10.5683/SP2/EIHG8O.

Mansournia, M. A. and D. G. Altman (2018). "Population attributable fraction". In: *BMJ* 360.

National Urban League (2020). *2020 State of Black America: Unmasked*. A National Urban League Publication.

Office for National Statistics (2020). *Does exposure to air pollution increase the risk of dying from the coronavirus (COVID-19)?* Last accessed 9 November 2020. URL: https://www.ons.gov.uk/economy/environmentalaccounts/articles/doesexposuretoairpollutionincreasetheriskofdyingfromthecoronaviruscovid19/2020-08-13.

Ogen, Y. (2020). "Assessing nitrogen dioxide (NO2) levels as a contributing factor to coronavirus (COVID-19) fatality". In: *Science of The Total Environment* 726, p. 138605.

Pope, I. C., R. Burnett, M. Thun, and et al (2002). "Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution". In: *Journal of the American Medical Association* 287.9, pp. 1132–1141.

Pozzer, A., F. Dominici, A. Haines, C. Witt, T. Münzel, and J. Lelieveld (2020). "Regional and global contributions of air pollution to risk of death from COVID-19". In: *Cardiovascular Research*.

Reich, B. J., S. Yang, Y. Guan, A. B. Giffin, M. J. Miller, and A. G. Rappold (2020). "A review of spatial causal inference methods for environmental and epidemiological applications". In: arXiv: 2007.02714 [stat.ME].

Rue, H., S. Martino, and N. Chopin (2009). "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71.2, pp. 319–392.

Smith, M. T. (2010). "Advances in Understanding Benzene Health Effects and Susceptibility". In: *Annual Review of Public Health* 31.1, pp. 133–148.

Wu, X., R. C. Nethery, M. B. Sabath, D. Braun, and F. Dominici (2020). "Air pollution and COVID-19 mortality in the United States: Strengths and limitations of an ecological regression analysis". In: *Science Advances* 6.45.

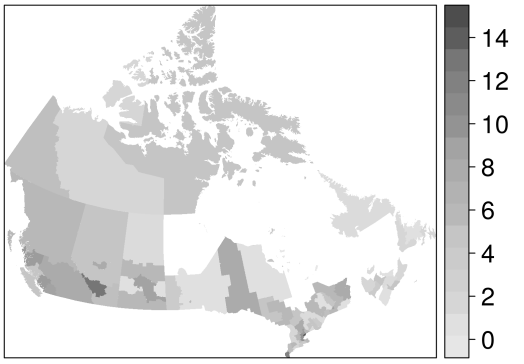
Yu, W. H., C. M. Harvey, and C. F. Harvey (2003). "Arsenic in groundwater in Bangladesh: A geostatistical and epidemiological framework for evaluating health effects and potential remedies". In: *Water Resources Research* 39.6.

Appendix–Data sources

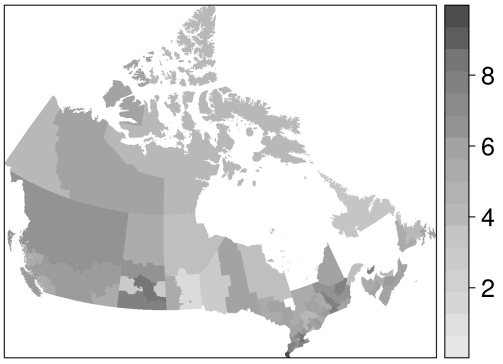
Table A.1: Data sources in this study.

Covariate	Description	Sources
Canada		
COVID-19	Github, Berry et al. (2020)	https://github.com/ishaberry/Covid19Canada
pm25	2014-2016, Environment Canada's NAPS	http://maps-cartes.ec.gc.ca/rnspa-naps/data.aspx?lang and Rural data
no2	2014-2016, Environment Canada's NAPS	http://maps-cartes.ec.gc.ca/rnspa-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
	2011,Statistics Canada	https://www150.statcan.gc.ca/n1/en/type/data?HPA=1
Unemployment		
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/EIHG8O
Italy		
COVID-19	R package covid19ita from Github	https://github.com/UBESP-DCTV/covid19ita/
pm25	2016-2018, European Environment Agency	https://www.eea.europa.eu/data-and-maps/data/aqereporting-8
no2	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/
pm25	2016-2018, European Environment Agency	https://www.eea.europa.eu/data-and-maps/data/aqereporting-8
no2	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
	2017/2018, Public Health England	https://www.localhealth.org.uk/#c=home
Unemployment		
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/
pm25	2016-2018, Atmospheric Composition Analysis Group	https://sites.wustl.edu/acag/
no2	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/incidencerates/index.php
	2018, Killeen et al. (2020)	https://www.kaggle.com/jieyingwu/covid19-us-countylevel-summaries
Unemployment		
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/

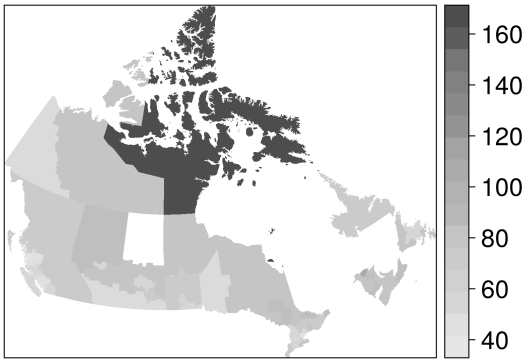
Appendix–Data description



(a) Canada, NO_2



(b) Canada, PM_{25}



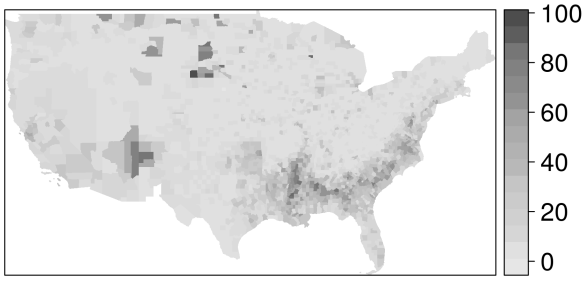
(c) Canada, Lung cancer

(d) Canada, Ethnicity

(e) United States, NO_2

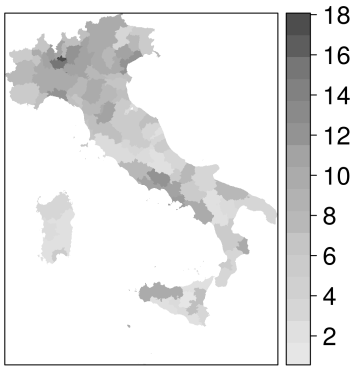
(f) United States, PM_{25}

(g) United States, Lung cancer

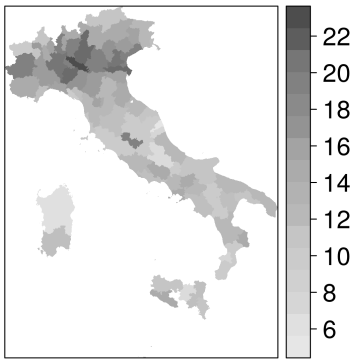


(h) United States, Ethnicity

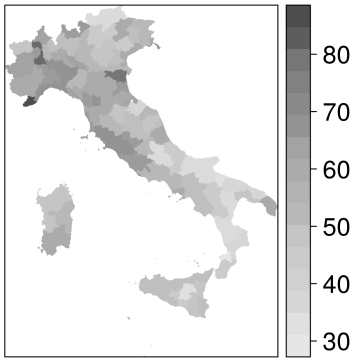
Figure A.1: Canada and United States data, including $PM_{2.5}$ (μgm^{-3}), NO_2 (ppb), Lung cancer (per 100,000) and Ethnicity (%).



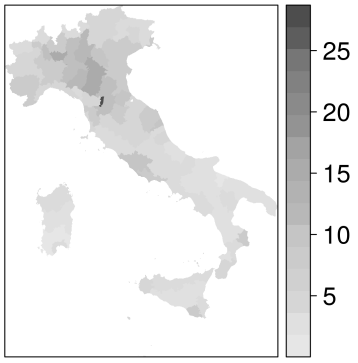
(a) Italy, NO_2



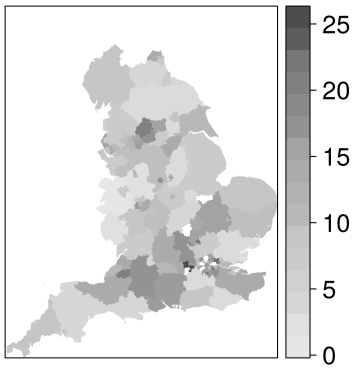
(b) Italy, PM_{25}



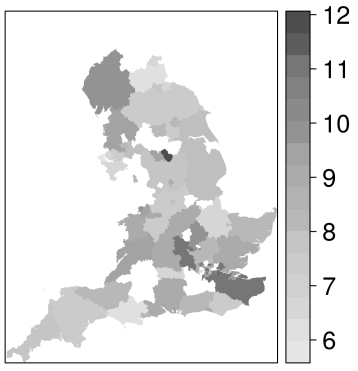
(c) Italy, Lung cancer



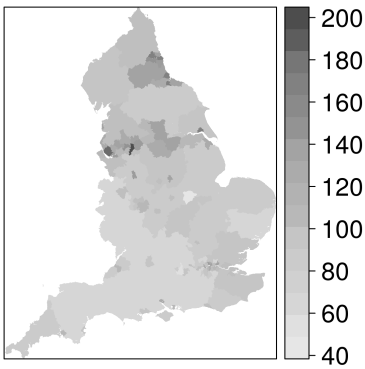
(d) Italy, Ethnicity



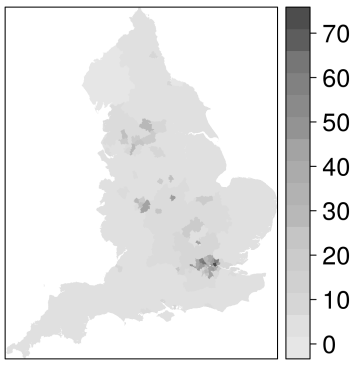
(e) England, NO_2



(f) England, PM_{25}

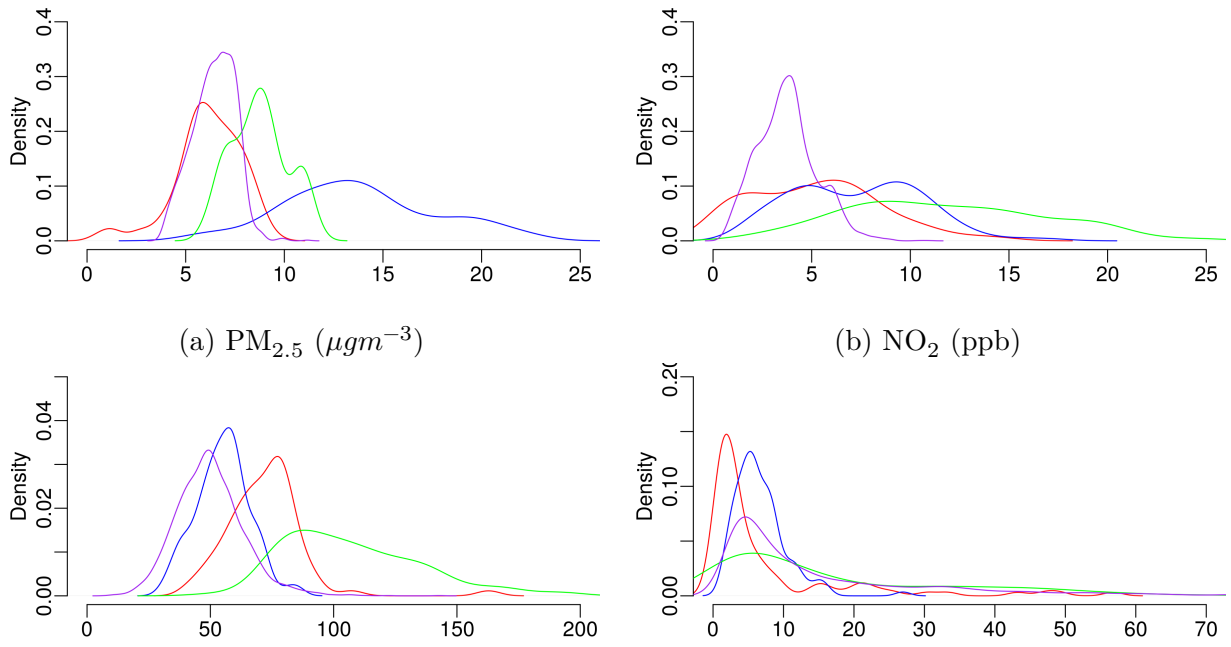


(g) England, Lung cancer



(h) England, Ethnicity

Figure A.2: Italy and England data, including $PM_{2.5}$ (μgm^{-3}), NO_2 (ppb), Lung cancer (per 100,000) and Ethnicity (%).



(a) $PM_{2.5}$ (μgm^{-3})

(b) NO_2 (ppb)

(c) Lung cancer (per 100,000)

(d) Ethnicity (%)

Figure A.3: Empirical distributions of covariates, Canada (red), Italy (blue), England (green), United States (purple).