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.
Bayesian profile regression to study the ecologic associations of correlated environmental exposures with excess mortality risk during the first year of the Covid-19 epidemic in Lombardy, Italy

Eric S. Coker a,*, John Molitor b, Silvia Liverani c, James Martin a, Paolo Maranzano d, Nicola Pontarollo e, Sergio Vergalli f

a Department of Environmental and Global Health, University of Florida, 1225 Center Dr, Gainesville, FL, 32610, United States
b College of Public Health and Human Sciences, Oregon State University, Milam Hall 157, 2520 SW Campus Way, Corvallis, OR, 97331, United States
c School of Mathematical Sciences, Queen Mary University of London, Mile End Road London E1 4NS, United Kingdom
d Department of Economics, Management and Statistics of the University of Milano-Bicocca (Unimib), Piazza dell’Ateneo Nuovo, 1 - 20126, Milano, Italy
e Department of Economics and Management, Università degli Studi di Brescia, Brescia, Via S. Faustino 74/B, 25122, Brescia, Italy
f Department of Agricultural Economics, Università Cattolica Del Sacro Cuore, Piacenza, Via Emilia Parmense, 29122, Piacenza PC, Italy

ARTICLE INFO

Keywords:
Covid-19
Air pollution
Inequalities
Social determinants of health
Mortality

ABSTRACT

Many countries, including Italy, have experienced significant social and spatial inequalities in mortality during the Covid-19 pandemic. This study applies a multiple exposures framework to investigate how joint place-based factors influence spatial inequalities of excess mortality during the first year of the Covid – 19 pandemic in the Lombardy region of Italy. For the Lombardy region, we integrated municipality-level data on all-cause mortality between 2015 and 2020 with 13 spatial covariates, including 5-year average concentrations of six air pollutants, the average temperature in 2020, and multiple socio-demographic factors, and health facilities per capita. Using the clustering algorithm Bayesian profile regression, we fit spatial covariates jointly to identify clusters of municipalities with similar exposure profiles and estimated associations between clusters and excess mortality in 2020. Cluster analysis resulted in 13 clusters. Controlling for spatial autocorrelation of excess mortality and health-protective agency, two clusters had significantly elevated excess mortality than the rest of Lombardy. Municipalities in these highest-risk clusters are in Bergamo, Brescia, and Cremona provinces. The highest risk cluster (C11) had the highest long-term particulate matter air pollution levels (PM2.5 and PM10) and significantly elevated NO2 and CO air pollutants, temperature, proportion ≤18 years, and male-to-female ratio. This cluster is significantly lower for income and ≥65 years. The other high-risk cluster, Cluster 10 (C10), is elevated significantly for ozone but significantly lower for other air pollutants. Covariates with elevated levels for C10 include proportion 65 years or older and a male-to-female ratio. Cluster 10 is significantly lower for income, temperature, per capita health facilities, ≤18 years, and population density. Our results suggest that joint built, natural, and socio-demographic factors influenced spatial inequalities of excess mortality in Lombardy in 2020. Studies must apply a multiple exposures framework to guide policy decisions addressing the complex and multi-dimensional nature of spatial inequalities of Covid-19-related mortality.

1. Introduction

Throughout the Covid-19 pandemic, caused by the global spread of the SARS-CoV-2 novel coronavirus, significant within-region and between-region spatial heterogeneity of COVID-19-related deaths has been reported (Konstantinouidis et al., 2022). Despite widespread implementation of public health interventions, which are shown to have prevented or delayed the spread or severity of the virus among the global population (Hsiang et al., 2020), the pandemic is ongoing, with some countries still experiencing relatively high Covid-19 case fatality rates (Johns Hopkins University, 2022; Ritchie et al., 2020). As of September 20, 2022, the ongoing COVID-19 global pandemic has led to...
over half a billion (612, 791, 170) confirmed positive SARS-Cov-2 cases and contributed to over 6 million (6,529,108) deaths (Johns Hopkins University, 2022).

Several studies have already reported extensively on the significant spatial inequalities of Covid-19 cases and excess mortality in Italy (Biggeri et al., 2020; Blangiardo et al., 2020; Coker et al., 2020, 2020, 2020; Conti et al., 2020; De Angelis et al., 2021; Della Rossa et al., 2020; Gibertoni et al., 2021; Kezjar and Lusa, 2020; Mannucci et al., 2020; Michelozzi et al., 2020; Mingione et al., 2021; Sandrini et al., 2020; Scortichini et al., 2021). A similar pattern has emerged for other countries in Europe. For instance, using hierarchical spatiotemporal estimation, Konstantinouidou et al. (2022) used weekly data on excess mortality during the pandemic in the regions (NUTS-2 level) of 5 European countries: Italy, France, England, Spain, and Switzerland. These countries exhibited significant yet variable spatial heterogeneity between and within themselves. In the case of Switzerland, mortality excesses were highly localized, whereas, in Greece, the spread was more homogeneous. The significant spatial heterogeneity of Covid-19 case fatality rate suggests that a confluence of spatial determinants of health can make some sub-populations more vulnerable to severe Covid-19 outcomes. Extensive epidemiological data has emerged showing that different place-based factors, such as underlying distributions of age and sex of the population, access to health care or health facilities, ambient temperature, and air pollution, appear to explain some of the observed spatial heterogeneity of Covid-19-related mortality (Aloisi et al., 2022; Biggeri et al., 2020; Blangiardo et al., 2020; Calderón-Larrañaga et al., 2020; Coker et al., 2020; Conti et al., 2020; De Angelis et al., 2021; Gibertoni et al., 2021; Kezjar and Lusa, 2020; Mannucci et al., 2020; Michelozzi et al., 2020; Sandrini et al., 2020; Scortichini et al., 2021; Ye et al., 2021).

An analysis (Islam et al., 2021) of global mortality data from the first year of the COVID-19 pandemic (2020) indicates that Italy ranked second only to the United States in excess mortality. During the first year of the pandemic, the Lombardy region of Northern Italy had some of the highest regional excess mortality rates in Italy, making it a remarkable case study of an area disproportionately impacted early on during the COVID-19 pandemic (Blangiardo et al., 2020; Buonanno et al., 2020; Coker et al., 2020; Modì et al., 2021). Coker et al. (2020) analyzed data for Northern Italy. They showed that higher levels of long-term ambient fine particulate matter (PM2.5) air pollution were associated with increased excess mortality during the early period of the COVID-19 pandemic. Other researchers have since corroborated this association in Italy (Aloisi et al., 2022; De Angelis et al., 2021; Ye et al., 2021) and other parts of the world (Bourdrel et al., 2021; Katoto et al., 2021; Kogevinas et al., 2021; Tchicaya et al., 2021). However, exposure to ambient PM2.5 does not occur in isolation from other co-occurring air pollutants. Instead, human exposure occurs as a mixture of chemicals (pollutant profiles) found in ambient air. Therefore, it is likely that no single pollutant—like ambient PM2.5—contributes alone (Collivignarelli et al., 2021) to driving spatial inequalities of COVID-19 mortality (Kogevinas et al., 2021). Moreover, spatial variability in co-exposure to social, built, and other natural environmental factors may make some sub-populations more susceptible to air pollutants’ health effects. Hence, there is a need to explore the spatial patterns in COVID-19-related excess mortality associated with air pollution, especially in the case of co-occurring social and environmental determinants of health.

A comprehensive assessment of the environment should incorporate the natural, built, and social environments, determining their combined effects on COVID-19 mortality is not trivial. The conventional multiple variable regression approach taken thus far—when analyzing spatial associations of air pollutants on COVID-19-related excess mortality—has typically modeled a given pollutant’s independent effects, adjusting for other co-pollutants or place-based social determinants of health. The various methodological challenges with adjusting for multiple co-exposures in a conventional regression framework include high potential correlations between co-exposures, leading to unstable effect estimates and difficulty identifying the “bad actors”. In addition, such high dimensional data results in the estimation of many model parameters, including testing multiple interaction effects. Cluster analysis can mitigate some of these methodological challenges mentioned above. First, cluster analysis exploits multiple correlated covariates by meaningfully grouping similar observations (or regions) for a specific set of predictors (Xu and Wunsch, 2010). Additionally, clusters identified from a cluster analysis can be mapped (Coker et al., 2016, 2018). Hence, we can investigate the spatial patterning of clusters of exposure profiles (Hoover et al., 2018) in a study area to identify sub-populations most vulnerable (Coker et al., 2018) to COVID-19 mortality (Das et al., 2021; Liu et al., 2021). And while it remains essential to identify individual place-based factors driving spatial inequalities of COVID-19 mortality, the reality is that multiple environmental factors can be highly correlated and interdependent, and vary regionally. This reality suggests that any policy decisions intended to mitigate the public health impacts of Covid-19 must evaluate the different co-exposure profiles that are either harmful or protective for important public health outcomes. Hence, there is a need to identify and describe observable co-exposure profiles and analyze how these different profiles can influence COVID-19 spatial inequalities.

Our study builds on our previous work regarding the association between ambient PM2.5 and COVID-19 mortality in Northern Italy. This new study identifies the combination of place-based factors (“exposure profiles”), including multiple ambient air pollutants, ambient temperature, and other contextual factors that could further explain within-region variability of COVID-19-related excess mortality, specifically in the Lombardy region of Northern Italy. Here, rather than estimate the health effects of air pollution conditional on other social determinants of health, we explore the joint patterning of air pollution with these contextual factors and assess ecologic associations between combined exposures and municipality-level excess mortality during 2020 in the Lombardy region. This data-driven approach thus deviates from the conventional multivariable regression approaches that have thus far examined the Covid-19 pandemic by testing for independent effects of air pollutants. We exploit statistical analysis techniques, including Bayesian cluster and spatial regression analysis methods that explore the joint effects of multiple environmental exposures and their spatial patterning.

2. Methods

2.1. Study area and population

Our study is restricted to the 12 provinces of the Lombardy region located in Northern Italy. Our analysis of this study area totals 1506 municipalities. Among these provinces is Milano, which incorporates Italy’s most heavily industrialized and populated city of Milan (Zoran et al., 2020). Italy’s Lombardy region encompasses approximately 24,000 km² with almost 10 million inhabitants, making it Italy’s most populous region and Europe’s third most populous. Lombardy is also a significant driver of multiple sectors of Italy’s economy, contributing up to 21.9% of the national Gross Domestic Product (GDP) (Regione Lombardia, 2020). We focus our analysis on the Lombardy region of Northern Italy because it represents a wide range of urban and rural contexts adjacent to one another and is thus interdependent in economic activity and shared governance. For example, while Lombardy is Italy’s most heavily industrialized area, it also includes more rural, agricultural-based provinces such as Lodi province, located in the Po River plain and borders Milano province (Cattaneo et al., 2011; Regione Lombardia, 2020). Lombardy’s diverse setting thus provides a unique opportunity to study place-based determinants of COVID-19-related excess mortality in a region that is spatially congruent, economically and governmentally interdependent, and relatively homogenous ethnically.
2.2. Covariate data

Our study utilizes area-level statistical and spatial data from various publicly available databases. The selection of spatial covariates was based on data suggested in the literature. We assessed several environmental factors, including long-term ambient concentrations of six air pollutants (previous 5-year average) (EEA, and annualized surface ambient temperature in 2020 (ERAS, 2021, p. 5). We included other municipality indicators available from the Italian National Institute of Statistics (ISTAT Statistics, 2021), including demographic (18 years or younger, 65 years or older, proportion male) and taxable income by the municipality. Other contextual GIS data were included, such as population density (Center for International Earth Science Information Network-CIESIN-Columbia University, 2017) and health facility location data available from the Global Healthsites Mapping Project (healthsites.io, 2021). Raster data of population density was averaged to the boundary of each municipality. Health facility data was restricted to only those sites classified as clinics, hospitals, or doctor offices. A per capita value was derived by dividing each municipality’s total number of facilities by the municipality population size in 2020. We assessed these place-based (ecologic) factors at the municipality level. We rescaled the respective values into Z-scores before fitting the Bayesian Profile Regression analysis (described below).

2.3. Exposure assessment: air pollution data and spatial interpolation estimation

Air pollution monitoring data from 2015 to 2019 for Italy and neighboring countries were obtained from the website of the European Environment Agency (EEA). These data consist of hourly measurements of PM2.5, PM10, SO2, NO2, CO, and O3 from air monitoring locations. The hourly readings were first aggregated to daily mean concentrations. However, if any monitor had missing data for any hour of a day, that entire day was considered missing and was imputed using inverse distance weighting. This interpolation technique uses neighboring monitors with measurement values to estimate the value of a location without data. We then calculated the average daily pollutant concentrations for each monitoring location using each day from 2015 to 2019. Some monitoring locations had a collocated air monitor, and in such cases, we averaged the daily values of the collocated monitors.

Using the long-term pollutant concentrations from each monitoring location, we applied kriging as a spatial interpolation method to estimate long-term pollutant concentrations as a continuous surface over the entire study area. We limited interpolation input data to the monitoring sites within 150 km of the Lombardy region (Fig. 1). The critical feature of kriging is that it exploits the spatial correlation between sampled points as the basis for estimating values at un-sampled locations, giving more weight to locations closer to an interpolation area of interest while also accounting for the clustering of sampling points. Here we used the functionality from the R package ‘gstat’ (Pebesma, 2004) to perform kriging. We specified an exponential model and allowed the optimization process to select the best parameters for the variogram. This allowed us to create a spatially continuous prediction surface of the concentration of each pollutant at a spatial resolution of 1 km. Finally, to derive a long-term pollutant concentration estimate for each municipality, we averaged the values of all pixels that fell within the municipality’s boundaries. The maps in Fig. 1 indicate the locations of each monitoring station and the distribution of spatially-krigged long-term pollutant concentrations.

Fig. 1. Five-year average concentrations (2015–2019) for each air pollutant and air monitoring location (red dots). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
air pollutant concentrations for each pollutant within the Lombardy region.

2.4. Outcome: all-cause excess mortality

We use excess mortality as our study outcome of interest to serve as a proxy for mortality attributed to the COVID-19 pandemic. Given that deaths attributable to COVID-19 have not been uniformly captured in Italy’s vital statistics records, excess mortality is a robust indicator of COVID-19 mortality tied to the pandemic situation (Buonanno et al., 2020; Coker et al., 2020). Here, we use all-cause mortality for 2020–2019; Coker et al., 2020). Here, we use all-cause mortality for 2020 (pandemic year) as the modeled outcome. Since we use mortality data from all of 2020, we can capture both of Italy’s COVID-19 epidemic waves (late February to early June [first wave] and then late October to the end of December [second wave]). To determine excess mortality in each municipality, we first use the number of deaths in the previous five years (2015–2019) and divide this number by the annual population size in these same five years to derive an expected mortality rate for each municipality. We then use this expected mortality rate to calculate the expected number of deaths in 2020 by multiplying the expected mortality rate by the population size of each municipality for the year 2020. This estimate of expected mortality was then used as the offset in the statistical analysis described in turn (De Angelis et al., 2021).

2.5. Statistical analysis

2.5.1. Bayesian Profile Regression

We applied a semi-parametric clustering method known as Bayesian profile regression (BPR) (Molitor et al., 2010) to investigate the ecologic relationship of joint distributions of multiple air pollutants, ambient temperature, and other social and demographic factors with COVID-19-related excess mortality. BPR was initially developed by Molitor and colleagues (Molitor et al., 2010) and further advanced by Liverani and colleagues with the PReMiuM package for R (Liverani et al., 2015). BPR can link potentially collinear covariates with a health outcome (response) through cluster membership. Extensions of the BPR model allow for spatially structured error terms (Liverani et al., 2016). Clustering with BPR is implemented using a Dirichlet process mixture model, which means there is no need to specify the number of clusters before fitting the clustering algorithm. Markov chain Monte Carlo (MCMC) sampling (Gilks et al., 1995) is used to fit the model, which outputs a different clustering or “partition” of the data at each iteration of the sampler. The MCMC algorithm therefore meaningfully propagates uncertainty into the clustering. A dissimilarity matrix and partition around medoids (PAM) approach is further applied in BPR to partition individuals (in our case, municipalities) into clusters based on BPR’s rich MCMC output.

The model estimates associations between the response variable (excess mortality) with clusters of exposure profiles using the following general formula:

$$\log( RR) = \theta_x + \beta w_i + u_i,$$

where $\beta = (\beta_1, \ldots, \beta_p)$ represents the regression parameter coefficients related to confounding covariates $w_i$ where $w_i = (w_{i1}, \ldots, w_{ip})$, $\theta_x$ is a random cluster effect with $x_i$ indexing the cluster to which individual $i$ belongs, and $u_i$ represents the spatially structured error term (Coker et al., 2018; Lavigne et al., 2020; Liverani et al., 2016). For a more detailed explanation of the Poisson count response sub-model with extra response variation, the reader is referred to (Liverani et al., 2015) where the full model specification is described in detail.

The objective of BPR in this study is to cluster the joint distribution (exposure profiles) of the contextual/air pollutant values jointly with the outcome, the total deaths for 2020, into the model. We modeled excess mortality (response) with a Negative Binomial distribution—allowing for over-dispersion—and offset excess mortality counts using the log of each municipality’s expected all-cause mortality in 2020. The covariates in the clustering portion of this joint analysis are six air pollutants (PM2.5, PM10, SO2, NO2, CO, and O3) and several correlated contextual factors such as age demographics (proportion 65 years or older, proportion 18 years or younger, male-to-female ratio), population density, health facilities per capita, income per capita, and average ambient temperature.

Regional Health Protection Agency (HPA) ($w_i$) was also fit as a possible spatial confounder for the spatial relationships between cluster and excess mortality (Conti et al., 2020). In 2015, Italy established the Health Protection Agencies (HPAs), that is, administrative subdivisions (districts) of the regions with the task of implementing regional planning regarding the provision of health care (e.g., hospitals and territorial medical facilities) and social and health care (e.g., care of the elderly) through public and private parties (subject to regulated contracts), and collection of local health registry data. Lombardy is divided into 8 ATSs, each with its own operational and economic autonomy but subject to regional guidelines. We therefore control for potential district heterogeneity arising from this autonomy by including in the model..

We also fit an intrinsic spatial conditional autoregressive (ICAR) term ($u_i$) as a spatial random effect in our BPR model. The ICAR model assumes complete spatial autocorrelation of the outcome, i.e., excess mortality, between a defined set of neighbors (Besag et al., 1991; Liverani et al., 2016). The neighbors’ adjacency weights matrix employed in our analysis is structured as $k = 4$ nearest neighbors. At each iteration of the MCMC algorithm, the risk of all-cause excess mortality risk for each cluster of exposure profiles is estimated. From the posterior distribution, we calculated the adjusted central tendencies and 95% credible intervals of excess mortality incidence rate ratios (IRR) relative to the baseline excess mortality rate of the study population. Here, we calculated quantities to represent credible intervals (equal-tailed) because it provides some indication of the levels of uncertainty of risk and covariate levels within cluster.

2.5.2. Supplementary regression analysis

A separate negative binomial hierarchical regression model was fit to supplement our primary BPR analysis. All covariates, including the HPA fixed effect, were fit jointly as model parameters to include independent associations with municipality-level excess mortality in Lombardy in 2020. However, a random slope was fit for ambient PM2.5 to determine if there is a spatially varying association between PM2.5 and excess mortality at the province level. The reader is referred to the Supplemental Materials to view the modeling results from this supplementary analysis. This model was fit using the glmer.nb function with the ‘lme4’ package in R.

3. Results

3.1. Overview

In 2020, there were 10, 027, 314 people and 136,249 deaths in Lombardy, resulting in an overall mortality rate of 13.58 per 1000 people. The previous five-year (2015–2019) annual average number of deaths was 99,771 for Lombardy. Accounting for changes in population size for 2020, the expected number of deaths was 99,791 for Lombardy. This resulted in an estimate of 36,458 excess deaths in 2020 (36.5% increase) for the Lombardy region (standardized mortality ratio (SMR) = 1.37), which is consistent with other estimates of 2020 excess mortality for Lombardy (Maruotti et al., 2021). The spatial distribution of municipality-level unadjusted SMRs of all-cause mortality for the study area is mapped in Fig. 2A. A Global Moran’s I test with inverse distance weighting shows evidence of Global spatial autocorrelation for municipality-level SMRs in 2020 (Moran’s I index = 0.114; p < 0.0001). Evidence of local positive spatial clustering of SMRs using a Local Moran’s I is mapped in Fig. 2C. As expected, we observe significant spatial variability of SMRs at the municipality level for Lombardy, with evidence for local spatial clustering of high SMRs occurring mostly in Bergamo, Brescia, and
Cremona provinces. The Pearson’s correlations between the spatial covariates included in the cluster analysis ranged between −0.85 and 0.96 (see Fig. S1, Supplemental Materials).

3.2. Bayesian Profile Regression

BPR analysis resulted in 13 representative clusters of contextual/air pollutant exposure profiles (Fig. 3). Statistical summaries of these clusters in terms of each cluster’s number of municipalities, population size, and different mortality measures are presented in Table 1. The spatial pattern of these 13 spatial exposure profile clusters is further mapped in Fig. 3, indicating that most clusters geographically straddle multiple provinces and are largely spatially congruous.

Fig. 4a summarizes incidence rate ratios (IRRs) of excess mortality for each cluster compared to the overall population mean (baseline) excess mortality. Clusters significantly associated with higher excess mortality relative to the baseline include Cluster 10 (IRR: 1.11 [95%CrI: 1.07, 1.15]) and Cluster 11 (IRR: 1.11 [95%CrI: 1.07, 1.17]). Conversely, Cluster 2 (IRR: 0.92 [95%CrI: 0.89, 0.95]), Cluster 8 (IRR: 0.91 [95%CrI: 0.88, 0.95]), Cluster 9 (IRR: 0.90 [95%CrI: 0.84, 0.96]) and Cluster 13 (IRR: 0.88 [95%CrI: 0.80, 0.96]) are associated with significantly lower excess mortality risk relative to baseline risk. The map in Fig. 4b highlights the spatial pattern of clusters with low (blue), moderate (green), and high (red) excess mortality risk. The municipalities within the two highest-risk clusters (C11 and C10) are located in Bergamo, Brescia, and Cremona provinces. Municipalities of C11 are located in Bergamo, Brescia and Cremona by 42.5%, 23.9% and 33.6% respectively. Municipalities of C10 are located in Bergamo and Brescia by 67.5% and 32.5%, respectively.

3.2.1. Exposure profiles of high-risk clusters

In Fig. 5, we summarize the posterior distributions of the air pollution exposure profiles for high-risk clusters.
pollutant and contextual values (scaled) for each representative cluster. Here, we characterize the exposure profiles of the high-risk clusters by comparing the individual covariate levels for each cluster relative to the overall (baseline) covariate levels in Lombardy. As seen in Fig. 5, the 10th and 90th percentiles are used to indicate if a variable, within a particular cluster, is significantly different from the population baseline average for that variable. Red boxplots indicate a high level, green indicates a typical level, and blue indicates a low level. The highest risk cluster (C11) has the highest air pollution levels for PM2.5 and PM10. C11 is also significantly elevated for NO2 and CO air pollutants compared to baseline levels but significantly lower for SO2 and O3. C11 is significantly elevated for the contextual clusters for temperature, proportion 18 years or younger, and male to female ratio. Conversely, C11 is significantly lower for income and 65 years or older. Air pollution levels for the other high-risk cluster, C10, are elevated considerably for O3 only and significantly lower for other air pollutants. Contextual variables with elevated levels for C10 include the proportion 65 years or older and the male to female ratio. All other contextual variables are

---

**Table 1**

Population size and mortality measures (Long-term Average, Observed, Expected, and Excess) for the Lombardy study population and clusters derived from Bayesian Profile Regression.

| Number of Municipalities | Population (2020) | Average Number of Deaths (2015-2019) | Observed Number of Deaths (2020) | Expected Number of Deaths (2020) | Excess Number of Deaths (2020) | SMR (2020) |
|-------------------------|-------------------|--------------------------------------|----------------------------------|-------------------------------|-------------------------------|------------|
| Total Clusters          | 1506              | 10,027,314                           | 99,771                           | 136,249                       | 99,791                        | 36,458     | 1.37       |
| 1                       | 203               | 2,215,543                            | 19,805                           | 28,566                        | 19,810                        | 8756       | 1.44       |
| 2                       | 64                | 1,128,021                            | 10,106                           | 13,776                        | 10,087                        | 3689       | 1.37       |
| 3                       | 187               | 1,159,080                            | 11,571                           | 15,122                        | 11,506                        | 3616       | 1.31       |
| 4                       | 84                | 226,336                              | 2580                             | 3267                          | 2544                          | 723        | 1.28       |
| 5                       | 80                | 2,470,377                            | 24,646                           | 32,171                        | 25,122                        | 7049       | 1.28       |
| 6                       | 132               | 121,900                              | 1507                             | 2189                          | 1477                          | 712        | 1.48       |
| 7                       | 92                | 220,210                              | 2509                             | 3243                          | 2487                          | 756        | 1.30       |
| 8                       | 91                | 423,309                              | 4458                             | 5930                          | 4434                          | 1496       | 1.34       |
| 9                       | 102               | 186,642                              | 2790                             | 3658                          | 2734                          | 924        | 1.34       |
| 10                      | 153               | 454,110                              | 4831                             | 7000                          | 4763                          | 2237       | 1.47       |
| 11                      | 154               | 999,749                              | 10,246                           | 14,375                        | 10,175                        | 4200       | 1.41       |
| 12                      | 93                | 179,271                              | 2198                             | 3175                          | 2148                          | 1027       | 1.48       |
| 13                      | 71                | 242,606                              | 2524                             | 3777                          | 2504                          | 1273       | 1.51       |

* Expected deaths are calculated using the number of deaths in the previous five years (2015-2019) and dividing this number by the population size in these same five years to derive an expected mortality rate for each municipality. This expected mortality rate was used to calculate the expected number of deaths in 2020 by multiplying the expected mortality rate by the population size of each municipality for the year 2020.

b SMR = Standardized Mortality Ratio (observed deaths/expected deaths) without adjustment.
significantly lower for C10, including temperature, per capita health facilities, income, 18 years or younger, and population density.

4. Discussion

The objective of this study is to estimate the multivariate relationship between multiple placed-based (ecologic) environmental factors and all-cause excess mortality in the first year of the pandemic (2020) for the Lombardy region of Northern Italy. We applied Bayesian profile regression (BPR) as a clustering tool to achieve this objective. The ecologic environmental factors assessed encompassed various dimensions of the environment such as natural (air pollutants and temperature), built (health facilities), and socio-demographic (age, sex, and income) factors. Bayesian profile regression (BPR) jointly detected clusters of exposure profiles and modeled the spatial relationship of these clusters with excess mortality after accounting for spatial autocorrelation of excess mortality and spatial confounding. In summary, two of the clusters identified in our analysis (C10 and C11) exhibited significantly higher adjusted excess mortality risk in 2020. The highest adjusted excess mortality risk cluster (C11) was characterized by significantly elevated long-term air pollutant levels (PM$_{2.5}$, PM$_{10}$, NO$_2$, and CO), temperature, proportion 18 years or younger, and male to female ratio, and significantly lower taxable income and proportion 65 years or older. The other high-risk cluster, C10, was characterized by significantly elevated O$_3$, 65 years or older, male to female ratio, significantly lower temperature, per capita health facilities, taxable income, proportion 18 years or younger, and population density. While clusters C10 and C11 exhibited distinct exposure profiles, the populations in these clusters had very similar elevated excess mortality risk levels. Our study underscores that the combination of local place-based factors, such as natural environment (air pollution and temperature), social environment (SES), and built environment (proximity to health facilities), vary spatially and influence spatial patterning of COVID-19-related mortality risk during the first year of the pandemic in Italy’s Lombardy region.

Our findings must be evaluated in light of emerging data from other spatially explicit analyses that show complex spatial relationships between COVID-19 outcomes and various environmental factors. A study by (Middya and Roy, 2021) explored spatial relationships between multiple environmental factors and COVID-19 mortality across districts of India. Comparing modeling results between geographically weighted regression (GWR) and global estimation modeling (e.g., ordinary least squares regression), GWR performed better at predicting COVID-19 mortality, and there was significant spatial heterogeneity in spatial effects on COVID-19 for ambient PM$_{2.5}$, socioeconomic status (SES), and older populations. Another study in India (Das et al., 2021) investigated the relationship between spatial clustering of COVID-19 containment zones and joint living environment deprivation in the megacity of Kolkata. Through Geographically Weighted Principal Component Analysis (GWPCA), the authors highlighted that joint environmental and local socio-economic factors strongly influenced the spatial clustering of COVID-19 hotspots. A study by (Tian et al., 2021)—conducted on 3125 U.S. counties—applied spatial cluster analysis to COVID-19 mortality data and explored the ecologic factors associated with COVID-19 mortality within spatial clusters (high, medium, and low mortality spatial clusters). The drivers of COVID-19 mortality varied between county-level spatial clusters, with increasing air pollution associated with significantly elevated mortality in low COVID-19 prevalence counties and older population associated with significantly elevated mortality in the median and high COVID-19 prevalence counties. A study by (Antonietti et al., 2020) performed a country-level cluster analysis of ecologic factors, including multiple indicators of economic development indicators, temperature, CO2, and population size. Their study involved COVID-19 mortality data in 142 countries and explored the association between country-level PM$_{2.5}$ levels and COVID-19 mortality varied by exposure profile clusters. Notably, the data showed that PM$_{2.5}$ effects varied significantly by exposure profile clusters, with strong positive associations between PM$_{2.5}$ and COVID-19 deaths in clusters with higher-income economies of Western Europe and North America (including Italy). A nationwide mortality study in Italy in 2020 found that omitting ambient temperature from multivariable negative binomial regression models resulted in a downward (attenuation) bias of the PM$_{2.5}$ effect estimate on mortality in the first year of the pandemic (Aloisi et al., 2022).

Our analysis suggests similarly complex spatial relationships between multiple place-based environmental factors (air pollution, SES,
Environmental Research 216 (2023) 114484

Fig. 5. Boxplots of the posterior distribution of air pollutants and contextual covariate values (scaled) as derived from profile regression MCMC output. Clusters (x-axes) are ordered by adjusted risk of excess mortality, with each column containing a boxplot in a panel representing a different cluster (C1 to C13). Boxplots are displayed by convention, with a black line within a boxplot being the median and upper and lower segments showing the interquartile range, and the black dots being outliers. We note that colored dots (e.g. red, blue, or green dots) represent the 5th and 95th percentiles of the posterior distribution for the particular variable within a cluster. The black horizontal line within each panel represents the population baseline for the entire study area for the particular variable of interest. Therefore, red-colored boxplots indicate the credible intervals (CrIs) at the 10th percentile do not overlap with the overall population value (horizontal black line). In contrast, green-colored boxplots indicate that the 90th percentile or 10th percentile CrIs do overlap with the study population values. Blue-colored boxplots indicate that the 90th percentile CrIs do not overlap with the overall population value. All values for the spatial covariates were normalized for the cluster analysis. Special note: The displayed cluster numbers only have meaning based on the number assigned by the algorithm. There is no other inherent meaning beyond the assigned cluster number. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

E.S. Coker et al.
measures such as localized lockdowns (at regional level NUTS-2) which shutdown non-essential activities and eventual reduction in the circulation of the virus (Pelagatti & Maranzano, 2021). In a study by (Della Rossa et al., 2020), they tested the effectiveness of local lockdown measures imposed at the regional level by modeling the areas as an interconnected network in which the flows of people determine the weight of the connections. Their data show that regional lockdown measures were effective in Italy. However, in a paper by (Mingione et al., 2021), authors model the spatio-temporal variability (using a spatio-temporal CAR model) of the number of cases from COVID-19, finding substantial spatial and temporal dependence occurred during the first two waves of the pandemic, despite the restrictive measures implemented by the government. They attributed this spatial and temporal dependency to regional heterogeneity in environmental, social, and economic factors and inter-regional mobility (transportation).

These ecologic analyses confirm that while national policies and local restrictions have helped in mitigating spread of SARS-CoV-2 in Italy, their effectiveness may have been limited by regional heterogeneity (Mingione et al., 2021). Therefore, different sub-regions may benefit from national policies that have additional flexibility to suit the local context and needs for local public health actions, mainly when targeting one of the many ecologic factors that influence the severity of Covid-19 across populations. For instance, we found that not all areas with high levels of air pollution experienced significantly elevated excess mortality during the first year of Lombardy’s Covid-19 epidemic. Municipalities with high levels of multiple air pollutants combined with higher male-to-female ratio and lower incomes resulted in significantly elevated adjusted excess mortality (cluster 11). In contrast, municipalities with high levels of multiple air pollutants combined with significantly lower male-to-female ratio and higher incomes resulted significantly lower adjusted excess mortality (see exposure profiles of cluster 8 and cluster 2 in Fig. 4). This finding suggests that air pollution effects on Covid-19 mortality are heterogeneous between sub-populations with different underlying characteristics. Thus, if air pollution control is to be put forth as one of several public health approaches to mitigate Covid-19 mortality, underlying population characteristics may need to be understood to ensure there is a public health benefit. Although individual-level epidemiological research is recommended to help elucidate whether air pollution effects are differential between sub-groups.

Another notable finding in our study is that clusters with higher excess mortality also exhibited significantly elevated male-to-female ratios (Fig. 5). Poisson regression models presented in the supplemental materials confirmed an apparent linear relationship between male-to-female ratio and excess mortality during the pandemic year. Another recent study (Nielsen et al., 2021) analyzed sex-specific trends of pre-Covid-19 pandemic excess mortality for 27 European countries. Their analysis showed that excess mortality is significantly higher among males during periods of excess mortality. This finding led authors to conclude that higher excess mortality that has been seen in the Covid-19 pandemic among males in Europe is likely to be more of a general sex-specific phenomenon during periods of excess mortality, rather than specific to Covid-19.

4.1. Strengths and limitations

The major limitations of our study are the spatial ecological analysis and the possibility of uncontrolled spatial confounders. These two limitations preclude us from making causal statements regarding the observed exposure-response relationships between spatial covariate profiles and excess mortality. However, limitations related to uncontrolled spatial confounding in our study are balanced by using a Bayesian hierarchical model with a spatial random effect. The spatial random effect partially controls for unmeasured risk factors that vary locally. We view the use of spatial random effect as a strength of this study. Another limitation is the spatial misalignment between air pollution and temperature estimates compared with mortality counts recorded within municipal boundaries. We averaged raster air pollutant and temperature values to address spatial misalignment and derive municipality-level estimates. However, air pollution and temperature levels will vary within municipalities. Averaging these values over municipal boundaries will mask this variability and contribute to some amount of exposure misclassification. Since this exposure misclassification emerges for the entire study area, we expect attenuation bias of effect estimates as typically seen with non-differential misclassification of exposure. This limitation is offset somewhat by the fact that inference is not based on any single covariate (e.g., PM2.5). Instead, we focus inference in our study on clusters defined by municipalities that share similar levels of multiple covariates. Finally, while there is inherent uncertainty in cluster allocation of municipalities, this is mitigated by using an infinite mixture model for clustering that is also set in a flexible Bayesian framework. This Bayesian profile regression approach thus propagates and retains uncertainty in clustering allocation and the modeling output, which is another strength and innovation of our study.

5. Conclusion

Our results suggest a complex network of built, natural, and socio-demographic environmental factors that converge to influence spatial inequalities of Covid-19-related excess mortality in Italy’s Lombardy region. Studies must apply a multiple exposures framework to help guide policy decisions that address the complex and multi-dimensional nature of spatial inequalities of Covid-19 impacts on public health.

Funding sources

This research received grant funding from Fondazione Eni Enrico Mattei (FEEM).

Credit author statement

Eric Coker: Conceptualization; Data curation; Formal analysis; Funding acquisition; Methodology; Supervision; Visualization; original draft, John Molitor: Formal analysis; Methodology; review and editing, Silvia Liverani: Methodology; review and editing. James Martin III: Visualization; Data curation; review and editing, Paolo Maranzano: Methodology; review and editing, Nicola Pantarollo: Conceptualization; Review and editing; Data curation, Sergio Vergali: Conceptualization; review and editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Eric S Coker reports financial support was provided by Fondazione Eni Enrico Mattei (FEEM).

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envres.2022.114484.

References

Aloisi, V., Gatto, A., Accarino, G., Donato, F., Aloissi, G., 2022. The effect of known and unknown confounders on the relationship between air pollution and Covid-19 mortality in Italy: a sensitivity analysis of an ecological study based on the E-value. Environ. Res. 207, 112131 https://doi.org/10.1016/j.envres.2021.112131.
E.S. Coker et al.

Hoover, J.H., Coker, E., Barney, Y., Shuey, C., Lewis, J., 2018. Spatial clustering of metal and metalloid mixtures in unregulated water sources on the Navajo Nation - Arizona, New Mexico, and Utah, USA. Sci. Total Environ. 633, 1667–1678. https://doi.org/10.1016/j.scitotenv.2018.03.030.

Hopkins University, Johns Hopkins COVID-19 Map [WWW Document]. Johns Hopkins Coronavirus Resource Center. URL https://coronavirus.jhu.edu/map.html, accessed 5.6.22.

Islam, N., Shalikinov, V.M., Acosta, R.J., Klimkin, I., Kawachi, I., Irizarry, R.A., Alejandro, G., Khunti, K., Yates, T., Jeon, D.A., Whitehead, M., Lewington, S., Lacey, B., 2021. Excess deaths associated with covid-19 pandemic in 2020: age and sex disaggregated time series analysis in 29 high income countries. BMJ n1137. https://doi.org/10.1136/bmj.n1137.

Istat Statistics [WWW Document]. 2021. URL http://istat.it/lang-en, accessed 5.6.22.

Kato, P.O., Brandt, A.S., Bakan, B., Obada, P.M., Kuhangana, K., Kayembe-Kitenge, T., Kitenge, J.P., Nkuli, C.B.L., Vanoirbeek, J., Nasr, T.W., Hoet, P., Nemerly, B., 2021. Acute nickel exposure in children with incidence, prevalence, severity and mortality of COVID-19: a rapid systematic review. Environ. Health. 20, 41. https://doi.org/10.1186/s12940-021-00714-1.

Keijzer, L., Lapp, L., 2020. Classification of yearly provincial overall age and gender specific mortality patterns during the COVID-19 epidemics in Italy. Epidemiol. Prev. 44, 271–281. https://doi.org/10.19153/EP20.5-6.52.127.

Kogevinas, M., Castano-Vinyals, G., Karachalios, M., Espinosa, A., de Cid, R., Garcia-Aymerich, J., Carreras, A., Cortés, B., Pérezuelos, V., Jiménez, A., Vidal, M., O’Callaghan-Gordo, C., Crach, M., Santano, R., Barrios, D., Puyol, L., Rubio, B., Izquierdo, L., Nuenewijhuisen, M., Dadvand, A., Aguilar, R., Moncunill, G., Dobato, C., Tonne, C., 2021. Ambient air pollution in relation to SARS-CoV-2 infection, antibody responses, and COVID-19 disease: a cohort study in catalonia, Spain (COViCAT study). Environ. Health Perspect. 129, 117003. https://doi.org/10.1289/EHP9726.

Konstantinoudis, G., Cameletti, M., Gómez-Rubio, V., Gómez, L.I., Pirani, M., Bari, G., Lavigne, A., Riu, J., Eisner, M., 2021. Excess mortality during the COVID-2019 pandemic in five European countries. Nat. Commun. 13, 482. https://doi.org/10.1038/s41467-021-28157-3.

lavigne, A., Freni-Sterrattino, A., Fecht, D., Liverani, S., Blangiardo, M., de Hoogh, K., Molitor, J., Hanell, A.L., Eveland, K., 2021. A Bayesian hierarchical inference analysis of mortality differences in COVID-19 associated excess mortality and ambient particulate matter and mortality in England. Environ. Epidemiol. 4, e098. https://doi.org/10.1093/EE/0000000908.

Lavirani, S., Hastei, D., Aziri, A., Papathomas, M., Richardson, S., 2015. PReMiaM: An R Package for Profile Regression Mixture Models Using Dirichlet Processes, vol. 6. 2022. Modelling collinear and spatially correlated data. Spatial and Spatio-temporal Epidemiology 18, 63–73. https://doi.org/10.1016/j.sste.2016.01.002.

Lombardy, Region. 2020. Territory and population [WWW Document]. URL https://www.en.regione.lombardia.it/wws/pal formatting for Profile Regression Mixture Models Using Dirichlet Processes, vol. 6. 2022. Modelling collinear and spatially correlated data. Spatial and Spatio-temporal Epidemiology 18, 63–73. https://doi.org/10.1016/j.sste.2016.01.002.

Lombardy, Region. 2020. Territory and population [WWW Document]. URL https://www.en.regione.lombardia.it/wws/pal

Manucci, E., Nreu, B., Monaci, M., 2020. Factors associated with increased all-cause mortality during the COVID-19 pandemic in Italy. Int. J. Infect. Dis. 98, 121–124. https://doi.org/10.1016/j.ijid.2020.06.077.

Marotti, A., Jona-Lasinio, G., Divino, F., Lovison, G., Cicciocci, M., Farcomeni, A., 2021. Estimating COVID-19-induced excess mortality in Lombardy (preprint). Epidemiology. https://doi.org/10.1093/ije/dyaa169.

Micheolzi, P., de Donato, F., Scortichini, M., Pezzotti, P., Stafoggia, M., De Sario, M., Costa, G., Nociocci, F., Riccardo, F., Bella, A., Demaria, M., Rossi, P., Bruunserafo, S., Perez, G., Davoli, M., 2020. Temporal dynamics in total excess mortality and COVID-19 deaths in Italian cities. BMC Publ. Health 20, 1238. https://doi.org/10.1186/s12889-020-09335-8.

Middya, A.L., Roy, S., 2021. Geographically varying relationships of COVID-19 mortality with different factors in India. Sci. Rep. 11, 7890. https://doi.org/10.1038/s41598-021-12632-x.

Molitor, J., Papathomas, M., Jerrett, M., Richardson, S., 2010. Bayesian profile curve: an application to the Italian regions. Spatial Statistics. https://doi.org/10.1016/j.spysst.2017.09.002.

Morrissey, I., 2018. Theological reflections on the persecution of the Jews during the COVID-19 epidemic. Environ. Health 20, 41. https://doi.org/10.1186/s12940-021-00846-1.

New Mexico, and Utah, USA. Sci. Total Environ. 633, 1667–1678. https://doi.org/10.1016/j.scitotenv.2018.03.030.

Orso, M.I., Ciampichini, R., Fornari, C., Doba, C., Rouge, C., 2021. COVID-19-induced excess mortality in Lombardy – territory and population. Accessed 4.27.22.

Papathomas, M., Richardson, S., Molitor, J., Franssen, J., de Donato, F., Scortichini, M., Pezzotti, P., Stafoggia, M., De Sario, M., Costa, G., Nociocci, F., Riccardo, F., Bella, A., Demaria, M., Rossi, P., Bruunserafo, S., Perez, G., Davoli, M., 2020. Temporal dynamics in total excess mortality and COVID-19 deaths in Italian cities. BMC Publ. Health 20, 1238. https://doi.org/10.1186/s12889-020-09335-8.

Pebsma, E.J., 2004. Multivariable geostatistics in 5: the gstat package. Comput. Geosci. 30, 683–691. https://doi.org/10.1016/j.comgeo.2004.03.012.

Richie, H., Mathieu, E., Rodés-Guirao, L., Appel, C., Giattino, C., Ortiz-Ospina, E., Hasell, J., MacDonald, B., Belkhatir, D., Roper, M., 2020. Coronavirus Pandemic (COVID-19). Our World in Data. 2020. https://ourworldindata.org/coronavirus, accessed 7.7.2022.

Ritchie, H., Mathieu, E., Rodés-Guirao, L., Appel, C., Giattino, C., Ortiz-Ospina, E., Hasell, J., MacDonald, B., Belkhatir, D., Roper, M., 2020. Coronavirus Pandemic (COVID-19). Our World in Data. 2020. https://ourworldindata.org/coronavirus, accessed 7.7.2022.
Suñer, C., Ouchi, D., Mas, M.Á., Lopez Alarcon, R., Massot Mesquida, M., Prat, N., Bonet-Simó, J.M., Exposito Izquierdo, M., Garcia Sánchez, I., Rodoreda Noguerola, S., Teixidó Colet, M., Veddaguer Puigvedelló, J., Henríquez, N., Miralles, R., Negredo, E., Noguera-Julian, M., Marks, M., Estrada, O., Ara, J., Mitjá, O., 2021. A retrospective cohort study of risk factors for mortality among nursing homes exposed to COVID-19 in Spain. Nat Aging 1, 579–584. https://doi.org/10.1038/s43587-021-00079-7.

Tchicaya, A., Lorentz, N., Omrani, H., de Lanchy, G., Leduc, K., 2021. Impact of long-term exposure to PM2.5 and temperature on coronavirus disease mortality: observed trends in France. Environ. Health 20, 101. https://doi.org/10.1186/s12940-021-00984-1.

Tian, T., Zhang, J., Hu, L., Jiang, Y., Duan, C., Li, Z., Wang, X., Zhang, H., 2021. Risk factors associated with mortality of COVID-19 in 3125 counties of the United States. Infect Dis Poverty 10, 3. https://doi.org/10.1186/s40249-020-00786-0.

Ye, T., Xu, R., Yu, W., Chen, Z., Guo, Y., Li, S., 2021. Vulnerability and burden of all-cause mortality associated with particulate air pollution during COVID-19 pandemic: a nationwide observed study in Italy. Toxics 9, 56. https://doi.org/10.3390/toxics9030056.

Zoran, M.A., Savastru, R.S., Savastru, D.M., Tautan, M.N., 2020. Assessing the relationship between surface levels of PM2.5 and PM10 particulate matter impact on COVID-19 in Milan, Italy. Sci. Total Environ. 738, 139825 https://doi.org/10.1016/j.scitotenv.2020.139825.