



Explaining regional differences in mortality during the first wave of Covid-19 in Italy

Ugo Filippo Basellini, Carlo Giovanni Camarda

► To cite this version:

Ugo Filippo Basellini, Carlo Giovanni Camarda. Explaining regional differences in mortality during the first wave of Covid-19 in Italy. *Population Studies*, 2021, Online first, pp.1-20. 10.1080/00324728.2021.1984551 . hal-03436531

HAL Id: hal-03436531

<https://hal.science/hal-03436531>

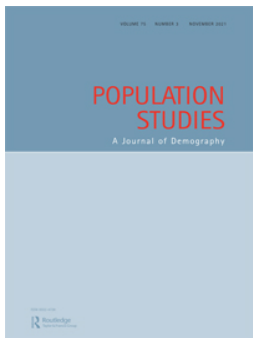
Submitted on 19 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



Population Studies
A Journal of Demography

ISSN: (Print) (Online) Journal homepage: <https://www.tandfonline.com/loi/rpst20>

Explaining regional differences in mortality during the first wave of Covid-19 in Italy

UgoFilippo Basellini & Carlo Giovanni Camarda

To cite this article: UgoFilippo Basellini & Carlo Giovanni Camarda (2021): Explaining regional differences in mortality during the first wave of Covid-19 in Italy, Population Studies, DOI: [10.1080/00324728.2021.1984551](https://doi.org/10.1080/00324728.2021.1984551)

To link to this article: <https://doi.org/10.1080/00324728.2021.1984551>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 09 Nov 2021.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

Explaining regional differences in mortality during the first wave of Covid-19 in Italy

Ugo Filippo Basellini ^{1,2} and Carlo Giovanni Camarda ²

¹Max Planck Institute for Demographic Research, ²Institut national d'études démographiques

Italy was hit harshly by the Covid-19 pandemic, registering more than 35,000 Covid-19 deaths between February and July 2020. During this first wave of the epidemic, the virus spread unequally across the country, with northern regions witnessing more cases and deaths. We investigate demographic and socio-economic factors contributing to the diverse regional impact of the virus during the first wave. Using generalized additive mixed models, we find that Covid-19 mortality at regional level is negatively associated with the degree of intergenerational co-residence, number of intensive care unit beds per capita, and delay in the outbreak of the epidemic. Conversely, we do not find strong associations for several variables highlighted in recent literature, such as population density or the share of the population who are older or have at least one chronic disease. Our results underscore the importance of context-specific analysis for the study of a pandemic.

Keywords: mortality modelling; SARS-CoV-2; Poisson regression; generalized additive mixed model; smoothing; socio-economic determinants; demographic factors; regional differences

[Submitted August 2020; Final version accepted August 2021]

Introduction

Coronavirus disease 2019 (Covid-19) is an infectious disease that has spread rapidly around the globe since the beginning of 2020. The disease is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), and it was first identified in the city of Wuhan in December 2019 (Du et al. 2020). Within a matter of weeks, the World Health Organization had declared the Covid-19 outbreak a public health emergency of international concern (on 30 January 2020) and then a pandemic (on 11 March 2020) (World Health Organization 2020a).

The global number of reported cases of Covid-19 rose at a very fast pace during 2020, increasing from about 10 thousand at the beginning of February to more than 17 million at the beginning of August. Similarly, the number of deaths attributed to the disease increased from around 250 to more than 675 thousand during the same period (Johns Hopkins University CSSE 2020; World Health Organization 2020b).

In Italy, the first case of Covid-19 was confirmed on 20 February 2020, although the virus had already been present in the country since January

(Cereda et al. 2020). Since the identification of 'patient one', the country has been hit harshly by the spread of the virus, with regard to both infections and deaths. The number of reported cases exceeded that of China on March 27, totalling almost 250 thousand cases by the end of July. In addition, Italy registered more Covid-19 deaths than any other country between 19 March and 7 April, surpassing 35 thousand deaths at the beginning of August (data from Johns Hopkins University CSSE 2020).

In response to the rapid spread of the virus, the Italian government adopted a series of measures to slow down its transmission. On 23 February 2020, 11 municipalities in the north of Italy were identified as the main cluster of the epidemic and put under quarantine. Simultaneously, out of the 21 NUTS-2 regions, six—all in northern Italy—implemented different restrictions, ranging from school closures to cancellations of public, religious, or sporting events. On 1 March, the Council of Ministers divided the country into three areas of risk: a red zone comprising the 11 municipalities in quarantine; a yellow zone comprising the regions of Lombardia, Veneto, and Emilia-Romagna, where public and sporting events were suspended and schools closed;

and the rest of the country, with less restrictive safety and preventive measures. Schools were then closed nationwide on 4 March, the entire country was put into lockdown on 9 March, and non-essential activities were suspended on 23 March. From 4 May onwards, restrictions were gradually eased, and freedom of movement across regions and other European countries was restored on 3 June (Ministero Della Giustizia 2021). Unfortunately, all these measures only partially mitigated the diffusion of the virus.

Since the early phases of the Italian outbreak, regional differences have emerged in terms of timing and magnitude of the virus's diffusion. Initially, the virus spread in the north of Italy: the regions in the yellow zone (Lombardia, Veneto, and Emilia-Romagna) were the first to confirm more than 200 cumulative infections at the start of March 2020. Southern regions, which were originally rather unaffected, witnessed increasing rates of contagion throughout the months of March and April. In terms of Covid-19 deaths, similar regional differences occurred, with northern regions generally more affected than southern ones.

Ongoing economic and demographic differences across Italian regions have been the subject of a vast body of literature (see e.g. Helliwell and Putnam 1995; Billari and Ongaro 1998; Di Giulio and Rosina 2007; Kertzer et al. 2009). However, little is known about the effects of such determinants in shaping Covid-19 mortality across Italy. Meanwhile, a fast-growing literature has highlighted the relationships between Covid-19 mortality and a large number of variables. The role of population age structure has been suggested as one explanation of the higher number of Covid-19 deaths in older vs younger populations (Dowd et al. 2020). Moreover, the prevalence of comorbidities can play an equally important role (Nepomuceno et al. 2020), but Boschi et al. (2021) found that Italian regions with greater prevalence of diabetes and allergies experienced lower Covid-19 mortality during the first wave of the pandemic. Higher population density may catalyse the spread of the disease (Rocklöv and Sjödin 2020; Wong and Li 2020; Sy et al. 2021), although other studies have shown no significant effect (Hamidi et al. 2020; Sun et al. 2020; Khavarian-Garmsir et al. 2021). Intergenerational relationships and co-residency structures have been found to be important with respect to the number of infections and deaths (Brandén et al. 2020; Esteve et al. 2020; Giorgi and Boertien 2020; Martin et al. 2020; Pengyu et al. 2021), but Arpino et al. (2020) and

Liotta et al. (2020) showed that a higher prevalence of intergenerational co-residence and greater numbers of contacts were negatively associated with Covid-19 case fatality rates and infection spread in Italian regions. Several studies have documented that a significant number of Covid-19 deaths in Italy occurred in nursing homes (Ciminelli and Garcia-Mandicó 2020; di Giacomo et al. 2020; Trabucchi and De Leo 2020) and that the sudden outbreak of Covid-19 saturated intensive care units (ICUs) in several regions, leading older patients to die at home because of the lack of available hospital beds (Favero 2020; Volpato et al. 2020). Furthermore, some workers are at higher risk of Covid-19 than others because they work in physical proximity to other people and/or they are more exposed to diseases and infections (Barbieri et al. 2020). This is the case, for example, for workers employed in the healthcare and manufacturing sectors (Chirico et al. 2020; Lapolla et al. 2021). Finally, differences between Italian regions in the start date of the epidemic have been shown to mask actual underlying heterogeneity in local dynamics (Scala et al. 2020); hence accounting for this diversity is crucial for disentangling the underlying factors behind regional differences in Covid-19 mortality.

In this paper, we investigate these demographic and socio-economic factors within the Italian framework. The aim is to reveal which of these factors contributed to regional differences in Covid-19 mortality during the first wave of the pandemic. Specifically, we analyse the period from the outbreak of the epidemic (late February 2020) until the first considerable reduction in Covid-19 deaths (mid-July 2020), a period during which over 35 thousand Covid-19 deaths were registered. We study the association between the daily reported number of Covid-19 deaths and a set of explanatory variables at the regional level. By using generalized additive mixed models within a Poisson framework, we identify the demographic and economic variables that had the strongest impact on the number of individuals dying. Our approach allows us to show how the epidemic unfolded in Italy and simultaneously to uncover the remaining regional heterogeneity that is not captured by the time trend and selected covariates.

This paper is organized as follows. The next section describes the mortality data that we use and the relevant covariates considered in our study, as well as the method that we use for our analysis. The following section illustrates the results of our analysis, and we then conclude with a discussion of our findings.

Data and methods

Mortality data

Since 25 February 2020, the Dipartimento Della Protezione Civile (2021) has been publishing the new (non-cumulative) number of reported Covid-19 deaths for each of the 21 NUTS-2 Italian regions on a daily basis. Regional mortality data are only available at the aggregate level (i.e. no information is provided on the age and sex breakdown of deceased individuals).

It is important to note that several regions did not register any Covid-19 infections or deaths for the first few days of the data set. As such, starting the analysis from 25 February for every region would not be appropriate, since regions with later epidemic outbreaks were not exposed to any Covid-19 mortality risk until the SARS-CoV-2 virus entered the region. To deal with different epidemic dynamics and to allow regions to share the same initial condition/state, we define region-specific start dates and shift the time scale of our regression analysis accordingly. A similar approach for studying the impact of geographic factors on Covid-19 in China has been adopted by Sun et al. (2020). In practice, for each region, we start the analysis from when cumulative cases surpassed 0.0001 per cent of the regional population. This allows us also to consider population size in the timing of the epidemic outbreak across regions.

The left-hand panel of Figure 1 shows a map of Italy with regions coloured according to their analysis start date. Earlier outbreaks are shown in darker shades, with later outbreaks in lighter shades. In the right-hand panels, Covid-19 log mortality rates over time are shown for each region, with colours corresponding to those in the map. Actual trends are smoothed here only for illustrative purposes. In the upper-right panel, rates are plotted over calendar time (dates); in the lower-right panel, start dates are shifted to begin from a common time point. An overall clustering of the curves is visible in the right-hand panels: regions with earlier epidemic outbreaks show on average higher levels of Covid-19 mortality than regions with later outbreaks.

In addition to this ‘relative’ definition of the analysis start date, we perform a sensitivity analysis and compute the start date using an ‘absolute’ approach: for each region, we start the analysis when cumulative cases surpassed five. Table A1 in the Appendix shows the start date of the epidemic by region in the main and sensitivity analyses. Specifically, we show that the outcomes of our analysis are robust

to the definition of the regional start date for the epidemic (see Table A2).

Explanatory variables

Here we describe the explanatory variables that we use in our regression setting and provide their sources and some descriptive statistics. For each variable, we also provide the expected effect on Covid-19 mortality as suggested by the relevant literature.

The total population for each region at the start of 2020 is retrieved from Istat (2021). Since deaths are reported on a daily basis, exposures (used as an offset in the regression setting) are approximated by dividing the regional population by the number of days in 2020, (i.e. 366). Consequently, in our analysis estimated mortality can be viewed as the daily risk of dying with Covid-19.

Istat (2021) also provides data by region on: (1) the share of population aged 65 years or over; (2) population density; and (3) the share of population with at least one chronic disease (comorbidity). Moreover, we retrieve regional data on the degree of intergenerational co-residence from Arpino et al. (2020), who computed the prevalence of older individuals (aged 60+) living in multigenerational households (of two or more generations) from the Family, Social Subjects and Life Cycle survey (Istat 2016).

Furthermore, Istat (2021) provides regional data on workers employed in the Italian economy. Specifically, we retrieve data on the number of physicians (general and specialized practitioners) per capita, and we compute the share of workers employed in the manufacturing sector.

Regional data on the number of ICU beds available in 2019 are retrieved from the Ministero della Salute (2020). Furthermore, we obtain the number of long-stay residential care homes (LSRCHs) in each region from the survey of the Istituto Superiore di Sanità (2020). This survey was carried out to monitor the spread of Covid-19 in LSRCHs, and it covered 3,417 out of the 4,629 total LSRCHs present in Italy. We also retrieve the cumulative number of Covid-19 swab tests performed from the first to the last day of the analysis for each region, from the Dipartimento Della Protezione Civile (2021). To account for the different sizes of regional populations, we use these three variables in per capita terms.

Finally, in addition to starting the analysis for each region at different time points, we also compute the number of days of delay in the start of the epidemic.

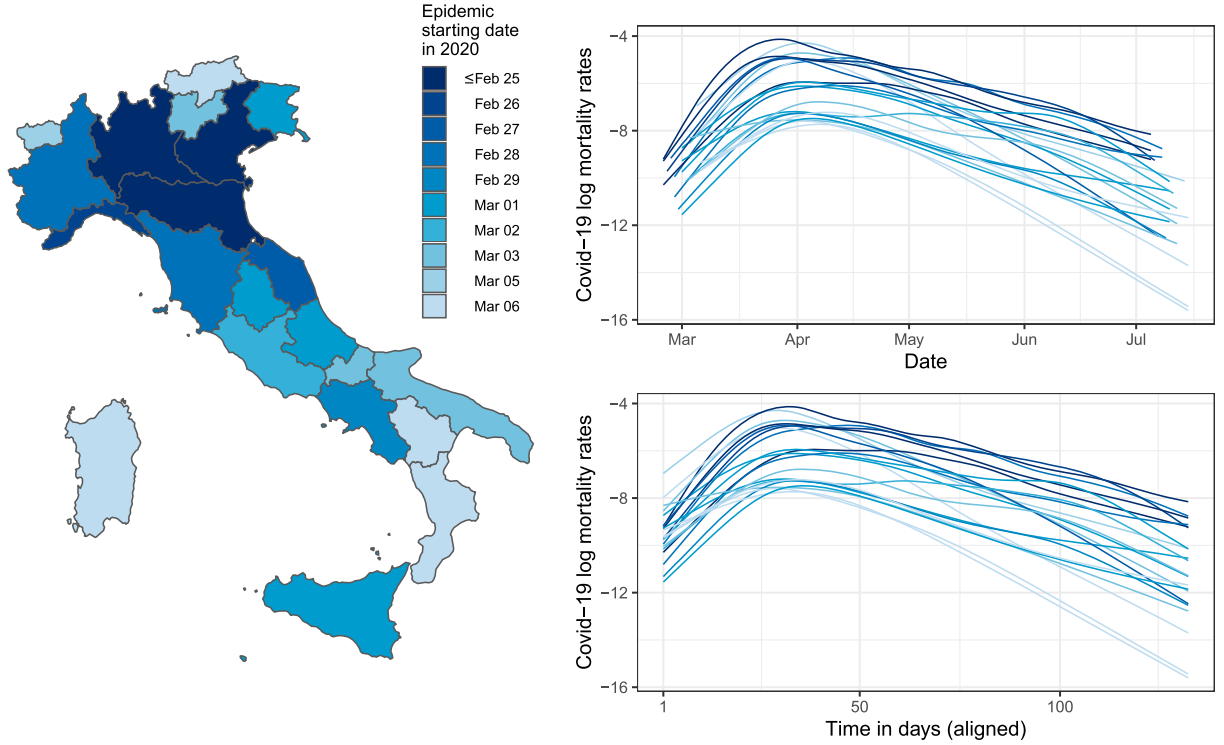


Figure 1 Start date of the epidemic by region (left-hand panel) and smoothed Covid-19 log mortality rates over time (right-hand panels): Italy, late February to mid-July 2020

Notes: On the right-hand side, calendar dates are shown in the upper panel and aligned days in the lower panel. Lines for the 21 regions are coloured according to their respective epidemic start date.

Source: Authors' own elaborations on data from Dipartimento Della Protezione Civile (2021).

This is derived as the number of days between 25 February (the first date in the data set) and the start of the epidemic in each region (see Table A1 in the Appendix).

As already discussed in the 'Introduction', we select this set of covariates since they have been suggested as relevant in recent literature on Covid-19 mortality. An overview of the explanatory variables is given in Table 1 along with their descriptive statistics and their expected effect on Covid-19 mortality based on published studies. In the Appendix (Figure A1), we report an exploratory analysis of these variables, including a graphical inspection of their linear relationship with Covid-19 mortality (Figure A2).

Modelling

Let $r = 1, \dots, n$ denote the regions and $t = 1, \dots, m$ denote the time points of the analysis. As discussed in the 'Mortality data' subsection, the first time point of the analysis corresponds to different calendar dates for each region (see Table A1 in the Appendix). Similarly, the last calendar date of

analysis differs by region, since we keep the same length of the time period for each region. In the main analysis, the length of the time period is 132 days ($m = 132$), whereas it is 128 days in the sensitivity analysis. The total number of regions, n , is always equal to 21.

Observed Covid-19 deaths, $d_{r,t}$, are assumed to be realizations from a Poisson distribution (Brillinger 1986) with mean $e_{r,t} \mu_{r,t}$, where $e_{r,t}$ denotes person-days of exposure to the risk of death for each region r . Exposures are assumed fixed over t for the period under study. The vector $\mu_{r,t}$ denotes the Covid-19 force of mortality for each region r at time t , and its estimation is the object of the proposed model.

We model the Poisson death counts via a log link function in a generalized additive mixed model (GAMM) framework. This type of model is particularly suitable in our setting since we clearly deal with non-independence in the data: the observed death counts in a given region over time are naturally correlated. Adding random effects at the regional level in the regression setting allows us to estimate correct standard errors associated to fixed effects and to avoid invalid relationships between Covid-19

Table 1 Explanatory variables considered in this study, together with descriptive statistics (mean, standard deviation, minimum, and maximum values for Italian regions) and their expected effect on Covid-19 mortality

Variable	Year	Mean	SD	Min	Max	Expected effect on Covid-19 mortality
Population ($\times 10,000$)	2020	284	252.9	12.5	1002.8	– (offset term)
Population aged 65+ years (%)	2020	23.8	2.2	19.3	28.8	↑ (Dowd et al. 2020)
Population density (per km ²)	2020	175.6	112.2	38.3	420.2	↑ (Rocklv and Sjdin 2020)
Population with one or more chronic diseases (%)	2019	41.4	3.4	30.4	46.6	– (Hamidi et al. 2020) ↑ (Nepomuceno et al. 2020)
Older people living in multigenerational households (%)	2016	40.6	6.1	30.9	52.2	↓ (Boschi et al. 2021) ↑ (Esteve et al. 2020)
Physicians per capita	2019	4.0	0.5	3.2	4.8	↓ (Arpino et al. 2020) ↑ (Chirico et al. 2020)
Employees in manufacturing sector (%)	2015	20.2	6.2	10.9	29.0	↓ (Volpato et al. 2020) ↑ (Barbieri et al. 2020)
ICU beds per capita	2019	8.6	1.4	5.7	12.0	↓ (Volpato et al. 2020)
LSRCHs per capita	2020	39.7	34.5	1.8	106.8	↑ (Trabucchi and De Leo 2020)
Cumulative Covid-19 swab tests per capita	2020	10.8	5.5	4.7	24.6	↑ / ↓
Delay in the epidemic (days)	2020	5.2	3.4	0.0	10.0	↓ (Scala et al. 2020)

Notes: Per capita variables are multiplied by 1,000 (physicians), 100,000 (ICU beds), 1,000,000 (LSRCHs), and 100 (cumulative Covid-19 swab tests) for illustrative purposes. Figures for the population variable are divided by 10,000 in this table only (and not in the analysis) for illustrative purposes.

Source: Istituto Superiore di Sanità (2020); Ministero della Salute (2020); Dipartimento Della Protezione Civile (2021); Istat (2021).

mortality and observed covariates. Moreover, GAMMs provide a powerful tool for including non-linear effects, making them particularly suitable in modelling the time dynamic of Covid-19 mortality.

Let d and e denote all observed deaths and associated exposures, respectively, arranged as a column mn vector. We model the expected values of the Poisson distribution as follows:

$$\ln[\mathbb{E}(d)] = \ln(e) + 1_n \otimes \eta + X\beta + Zu, \quad (1)$$

where $\ln(e)$ is the offset term and $\eta = (\eta_t)$ represents the common epidemic dynamic over time $t = 1, \dots, m$. This trend is repeated for all regions by means of a Kronecker product ($\otimes$) and 1_n , a column vector of '1's of length n . The $mn \times p$ design matrix X contains the values of the p explanatory variables, which vary across regions; the vector β with the associated p coefficients is common over regions and time, and it can be interpreted as in a classic regression setting. Finally, Z is the $mn \times n$ model matrix for the random effects for observed deaths in region r . Since we deal with a balanced design, Z has a block-diagonal structure and could be built as $Z = I_n \otimes 1_m$, where I_n is an identity matrix of size n and 1_m a column vector of '1's of length m . The n -vector u contains the region-specific random effects, added to capture average regional differences and assumed to be normally distributed with mean zero and constant variance across regions. Note that all data used in the model in equation (1) are observed (i.e. there are no missing data).

In addition to the GAMM, we also use a simpler generalized additive model (GAM) that only contains fixed effect terms. The model is identical to the one described in equation (1), with the exception of not containing the random effect terms Zu . In other words, this simplified approach assumes that mortality trends in each region are only the sum of a national smoothed trend η and regression term ($X\beta$) that uniquely depends on observed explanatory variables and a common coefficients vector. Comparisons between the GAM and GAMM approaches are performed throughout our analysis.

We assume smoothness for the common epidemic time trend η . Following a P -spline approach, we model this function as a linear combination of cubic B -splines and associated coefficients which are penalized by discrete penalties (Eilers and Marx 1996). Following the standard approach, we use an intentionally generous number of B -splines (29, corresponding to one knot at every fifth data point, sufficient to describe how the epidemic unfolded in Italy), and we assign to the penalty the

role of reducing the effective dimension, leading to a smoothed term η . A review of spline modelling, including penalized splines and their implementation in R, can be found in Perperoglou et al. (2019). Model selection is performed by minimizing the Bayesian Information Criterion (BIC) (Schwarz 1978), which provides a trade-off between the accuracy and parsimony of the model. BIC is given by:

$$\text{BIC} = \text{Dev} + \log(mn) \text{ED}, \quad (2)$$

where Dev denotes the Deviance (which captures the discrepancy between observed and fitted data) and ED denotes the effective dimension of the model. The latter term is the sum of the degrees of freedom used by smoothed, fixed, and random components in the model in equation (1).

The estimation procedure is implemented in R (R Development Core Team 2020), using the `mgcv` package (Wood 2019). The code can be obtained from a public OSF repository (Basellini and Camarda 2021).

We start by running five different models on the 10 covariates introduced in the 'Explanatory variables' subsection. First, we run two models that contain all covariates, one without random effects (denoted 'GAM') and one with random effects (denoted 'GAMM'). Second, we run a GAMM model that does not include any covariates (denoted 'GAMM w/o covars'). Third, for both the original models (with and without random effects), we look at all $2^{10} - 1 = 1,023$ possible combinations of the 10 covariates (different combinations), and we retain the model that minimizes the BIC. (This comprehensive approach leads to equivalent outcomes as using a forward selection or backward elimination variable selection approach.) These two optimal models are denoted 'GAM-opt' and 'GAMM-opt', respectively. All five models include a smoothed function of time and, as an offset, person-days of exposure.

Results

Table 2 and Figure 2 report the results for the five models introduced in the previous subsection. Note that results for the GAM and GAM-opt models are the same (in both Figure 2 and Table 2), and that the GAMM w/o covars model does not appear in Figure 2 since it does not contain explanatory variables. To aid the interpretation and comparison of the different coefficients, we normalize each variable by mean-centring and scaling to one standard deviation (1SD). As such, the estimated coefficients can be interpreted as the change in the Covid-19

Table 2 Estimated coefficients with 95 per cent confidence intervals for five Poisson regression models of Covid-19 mortality (dependent variable) and 10 explanatory variables: Italy's 21 regions, late February to mid-July 2020

	Model			
	GAM and GAM-opt	GAMM w/o covars	GAMM	GAMM-opt
Population aged 65+ years	0.49 (0.44, 0.54)	–	0.30 (–0.50, 1.10)	–
Population density	0.29 (0.27, 0.32)	–	0.17 (–0.57, 0.91)	–
Population with one or more chronic diseases	–0.48 (–0.55, –0.40)	–	–0.20 (–0.82, 0.41)	–
Older people living in multigenerational households	–0.54 (–0.57, –0.52)	–	–0.66 (–1.19, –0.13)	–0.82 (–1.11, –0.54)
Physicians per capita	0.19 (0.13, 0.25)	–	–0.06 (–0.67, 0.55)	–
Employees manufacturing sector	0.34 (0.29, 0.38)	–	–0.03 (–0.68, 0.62)	–
ICU beds per capita	–0.60 (–0.64, –0.55)	–	–0.39 (–0.87, 0.09)	–0.42 (–0.73, –0.10)
LSRCHs per capita	0.56 (0.50, 0.62)	–	–0.12 (–0.75, 0.51)	–
Cumulative tests per capita	–0.40 (–0.46, –0.35)	–	0.21 (–0.57, 0.99)	–
Delay in the epidemic	–0.25 (–0.30, –0.20)	–	–0.49 (–1.40, 0.43)	–0.61 (–0.92, –0.30)
Observations	2,772	2,772	2,772	2,772
Regions	21	21	21	21
Deviance	8,721.22	7,274.20	7,274.16	7,274.34
Effective dimension	24.87	35.04	35.01	34.86
BIC	8,918.35	7,552.00	7,551.71	7,550.67
Deviance explained (%)	87.9	89.9	89.9	89.9

Notes: All models include a smoothed function of time and person-days of exposure as an offset. Estimation is performed using (restricted) maximum likelihood. See ‘Modelling’ subsection for details of the five models.

Source: Authors' elaborations on data from Istituto Superiore di Sanità (2020); Ministero della Salute (2020); Dipartimento Della Protezione Civile (2021); Istat (2021).

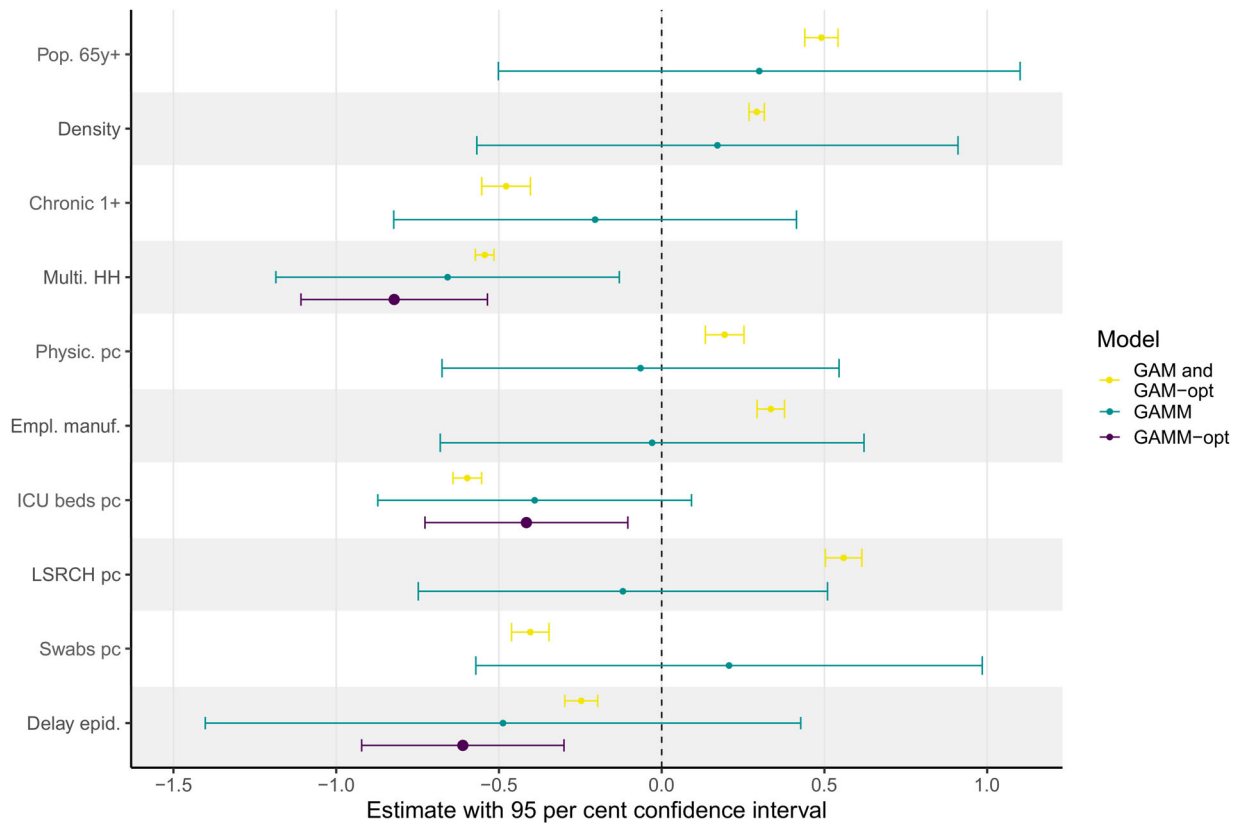


Figure 2 Estimated coefficients with 95 per cent confidence intervals from four models of Covid-19 mortality with 10 explanatory variables: Italy, late February to mid-July 2020

Note: Outcomes are also reported in Table 2. See ‘Explanatory variables’ subsection and Table 1 for details of the 10 explanatory variables.

Source: Authors’ elaborations on data from Istituto Superiore di Sanità (2020); Ministero della Salute (2020); Dipartimento Della Protezione Civile (2021); Istat (2021).

log mortality corresponding to a 1SD increase in the variable.

The first consideration that can be drawn from Table 2 is that the inclusion of random effects greatly improves the fit of the model, resulting in a considerable reduction of the deviance compared with models that do not consider them: 7,274 vs 8,721. This gain comes at a price of about 10 degrees of freedom (about 35 vs. 25). However, based on the BIC, any of the three models with random effects outperforms the models with only fixed effects.

As expected, random effects also increase the standard errors of the estimated fixed coefficients, thereby widening the 95 per cent confidence intervals associated with the covariates’ effects. Figure 2 clearly depicts these differences: error bars associated with point estimates increase considerably for the GAMM and GAMM-opt models compared with the GAM and GAM-opt ones. As such, it appears that the omission of random effects from the model specification likely underestimates the uncertainty associated with the coefficient estimates.

In all models containing covariates, the strongest effect on mortality is found for the prevalence of older individuals living in multigenerational households, with a 1SD increase (6.1 percentage points; see Table 1) associated with a 0.54–0.82 decrease in Covid-19 log mortality. The smallest effects are observed for the share of employees in the manufacturing sector and physicians per capita, whose coefficients are slightly negative in the GAMM model. Furthermore, most of the estimated coefficients are robust to the model specification, as the signs and values of the estimates are quite similar across the four models in Figure 2. For four variables, the estimates change sign (from positive to negative or vice versa) from the GAM to the GAMM models, but the confidence intervals of the GAMM model include zero for all four variables and include the point estimates of the GAM model (except for the LSRCHs variable). Nonetheless, this issue does not concern the variables selected for the optimal model (GAMM-opt). Finally, the BIC selection without and with random effects results in rather different

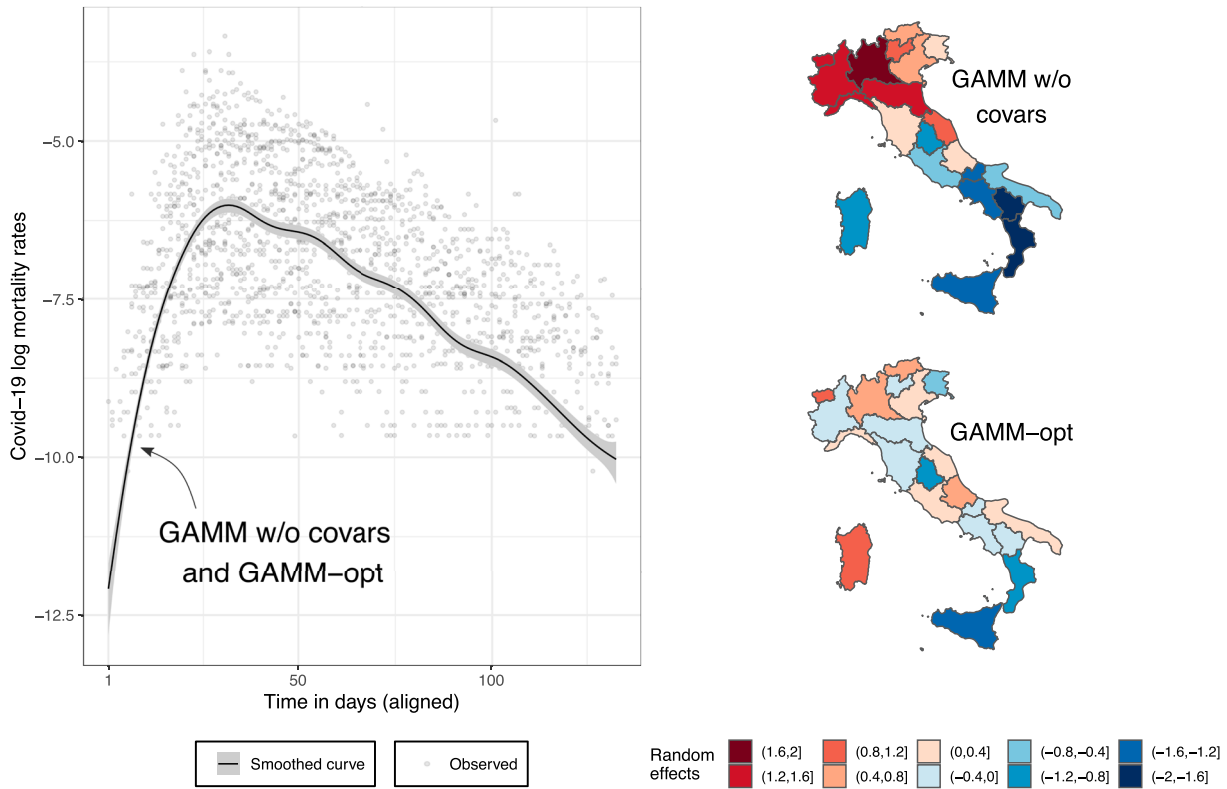


Figure 3 Observed Covid-19 log mortality rates and estimated smoothed reference curve for GAMM w/o covars and GAMM-opt models (left-hand panel), and estimated random effects for the two models (right-hand panel): Italy and its 21 regions, late February to mid-July 2020

Notes: In the left-hand panel, points show observed Covid-19 log mortality rates and lines with shaded areas show estimated smoothed reference curves with 95 per cent confidence intervals. The smoothed curves for the two models are visually indistinguishable from each other.

Source: As for Figure 2.

models: in the former, no covariates are excluded from the model, while in the latter only three covariates are retained.

The three variables retained in the optimal model (GAMM-opt) are: the prevalence of older individuals living in multigenerational households, the delay in the epidemic, and the number of ICU beds per capita. Regions with a greater prevalence of intergenerational co-residence experience lower Covid-19 mortality than regions with smaller intergenerational co-residence. Similarly, lower Covid-19 mortality is associated with greater delays in the arrival of the epidemic and with greater stock of ICU beds per capita. Very similar results are obtained in the sensitivity analysis using the absolute approach to define the start date of the analysis and the computation of the delay in the epidemic (see Table A2 in the Appendix).

Figure 3 shows additional results of our analysis by comparing the GAMM w/o covars model with the GAMM-opt model. The left-hand panel shows the observed Covid-19 log mortality rates alongside the

estimated smoothed curves over time for the two models, with 95 per cent confidence intervals. The smoothed curves for the two models are visually indistinguishable from each other, hence only one curve appears in the graph. The smoothed curve describes the average profile of the epidemic over time, with a rapid log mortality increase in the first 30 days, followed by a relatively stable but slower decline. In the right-hand panel, the random effects of the two models are plotted on corresponding maps of Italy. The inclusion of relevant covariates in the GAMM-opt model considerably reduces the size of the random effects, pointing towards the importance of the covariates in capturing the variability in the data. The random effects from the GAMM w/o covars model (upper-right panel) provide a direct visualization of the severity of the epidemic in different regions: higher (or lower) random effects translate into a vertical upward (or downward) shift of the estimated smoothed curve in the left-hand panel. The well-documented north-south divide emerges from the random

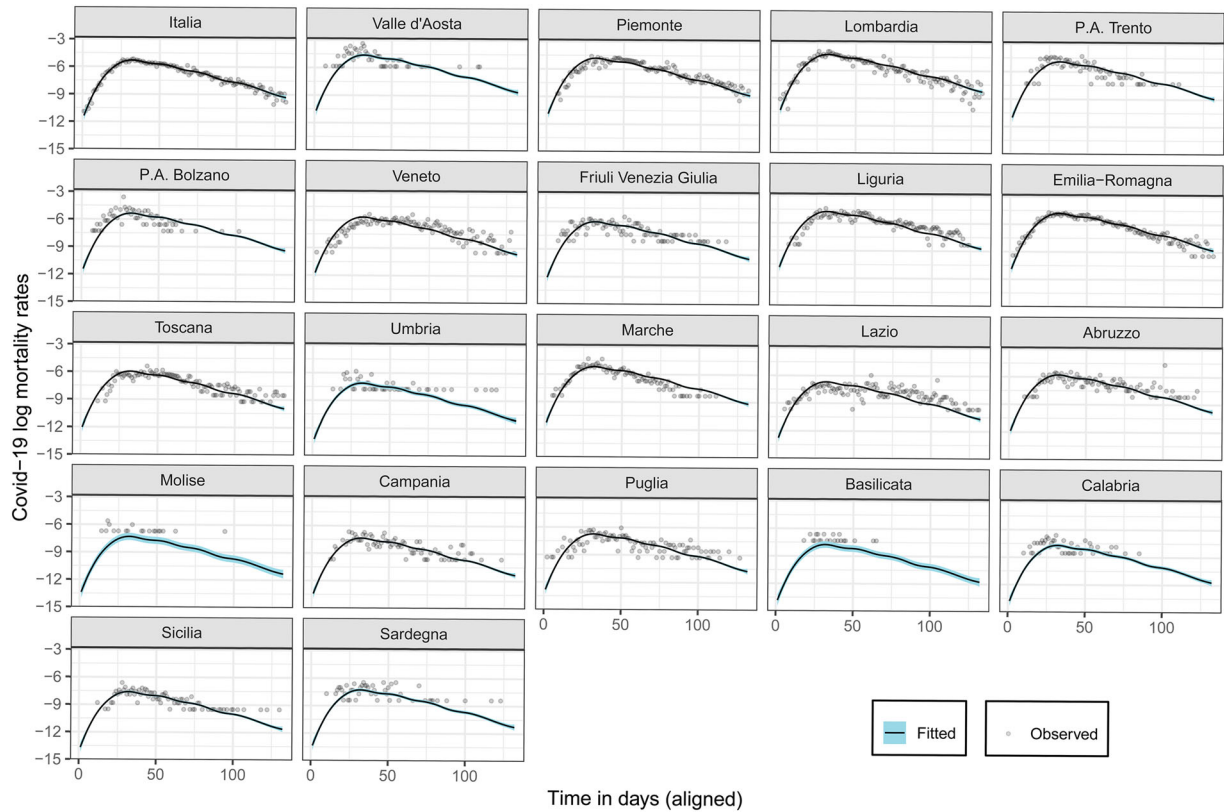


Figure 4 Observed and fitted (with 95 per cent confidence intervals) Covid-19 log mortality rates in Italy and its 21 regions for 132 aligned days between 25 February and 15 July 2020

Note: Fitted curves are from the GAMM-opt model.

Source: As for Figure 2.

effects of this model. Conversely, the random effects from the GAMM-opt model (lower-right panel) capture remaining unexplained heterogeneity in the data that is not captured by the time trend and the three covariates in this model.

Finally, we can inspect the goodness of fit of the GAMM-opt model by comparing the observed and fitted regional evolution of the epidemic. Figure 4 shows the observed and fitted Covid-19 log mortality with 95 per cent confidence intervals for each region, as well as for Italy overall (top-left panel). The graph shows that the model describes the data well, although the flatter epidemic shape in the Veneto and Lazio regions is not perfectly captured. For completeness, we report a similar plot for death counts, as well as the model's deviance residuals as a model diagnostic in Figures A3 and A4 in the Appendix.

Discussion

The research community has been very responsive in analysing the spread of Covid-19 across the globe, generating an extensive number of analyses.

Consequently, for the purpose of this paper, we focused mainly on the relevant literature for the Italian context. Early efforts were directed towards monitoring the virus's spread at the national or sub-national level using Poisson models. Chiogna and Gaetan (2020) proposed a dynamic generalized linear model for the Poisson distribution of new and total cases at the national, regional, and provincial levels. The analysis of the time-varying slope of the local linear trend allowed them to detect changes in the underlying process in terms of acceleration, deceleration, or stabilization of the disease's diffusion. Moreover, Bonetti and Basellini (2021) introduced a tool for visualizing the spread of Covid-19 in Italian provinces and regions by modelling the total number of cases with Poisson regression and using parametric and non-parametric hazard functions. Furthermore, Agosto et al. (2020) proposed a Poisson autoregression on the daily number of cases and compared the Italian context with that of China and other European countries.

All these works consider the spread of the virus in different territories *in isolation*, that is, each region or province is analysed independently without

taking into account their correlations. However, regional contexts have played a relevant role in the unfolding of the Italian epidemic. As such, in this paper we introduced a method that considers the country as the sum of different regional experiences. Our regression model allowed us to identify the most significant variables that contributed to the greater or lower burden of deaths across regional units during the first wave of the pandemic (i.e. from the end of February until mid-July 2020). In particular, we found that regions with three features—(1) a greater degree of intergenerational co-residence, (2) a longer delay in the start of the epidemic, and (3) higher number of ICU beds per capita—experienced lower levels of Covid-19 mortality. The other demographic and socio-economic variables that we analysed (share of the older population, population density, prevalence of one or more chronic conditions, number of physicians per capita, share of employees in the manufacturing sector, number of LSRCHs per capita, and Covid-19 swab tests per capita) were not retained in the optimal model selection.

Few attempts have been made to take into account cross-regional dependence in the spread of Covid-19 in Italy. Maltagliati (2020) was among the first to suggest analysing the Italian epidemic as the sum of region-specific outbreaks. This author described the cumulative number of deaths using a logistic model that considered the delay in the start of regional outbreaks. The proposed model produced large differences in the region-specific asymptotes, and the author argued that the regional perspective is fundamental to understanding the evolution of Covid-19 in Italy. Furthermore, Boschi et al. (2021) used functional data analysis techniques to investigate the association of Covid-19 mortality with mobility, positivity (i.e. the proportion of performed Covid-19 tests returning positive results), and other covariates at the regional level from mid-February to April 2020. The authors documented the outbreak of two starkly different epidemic types: an exponential one in the worst hit areas in the north of the country and a flat one in the remaining regions.

Our analysis shares commonalities with these two studies, but it differs substantially in a number of ways. First, the parametric model proposed by Maltagliati (2020) did not consider the role of explanatory variables in shaping the effects of the Covid-19 outbreak across the regions. Second, while the semi-parametric analysis by Boschi et al. (2021) was conceptually closer to our approach, the focus of the two studies was rather different. In their analysis, Boschi et al. (2021) concentrated on the role of

mobility and positivity as predictors of Covid-19 mortality, and other covariates were only considered, one at a time, as control variables in the regression model. In our work, we took a more comprehensive view and assesses the competing effect of several factors on mortality during the pandemic. Nonetheless, we acknowledge the contributions of these two studies; our work was intended to complement them and provide additional insights on the dynamics of the Italian epidemic.

An important contribution of our paper is the consideration of random effects in the analysis of the association between Covid-19 mortality and other factors. The mixed effects specification of the model indeed allowed us to isolate and estimate the effect of different covariates on Covid-19 mortality while controlling for: (1) the time trend of the epidemic; and (2) all other unobserved region-specific factors that are captured by the random effects of the model. This allowed us to control, for example, for regional differences in terms of testing and reporting procedures, which are not well documented and could have affected the number of reported Covid-19 deaths across regions.

Several of our findings are in line with those documented by recent research on the Covid-19 pandemic. The association between greater availability of ICU beds and lower Covid-19 mortality is reasonable, especially given that some regions (e.g. Lombardia) experienced a saturation in ICU bed and ventilator capacity during the Covid-19 outbreak and that could have resulted in a great number of Covid-19-related deaths (Favero 2020; Volpato et al. 2020). Similarly, the negative association between mortality and the delay in the outbreak is justifiable, since regions that experienced a later outbreak had relatively more time to prepare for it, for example, by increasing their stock of available ICU beds. Furthermore, we found a positive—albeit comparatively weaker, and with confidence intervals crossing zero in the model with random effects—association of mortality with the share of population aged 65+ and population density (in line with Dowd et al. 2020; Rocklöv and Sjödin 2020, respectively).

Interestingly, some of our findings do not align with other recent research on the pandemic. First, co-residence patterns (and, relatedly, household size) have been proposed as a key factor (together with population age structure) in determining countries' vulnerability to Covid-19 outbreaks (Esteve et al. 2020). In our analysis, we found that co-residence patterns were negatively associated with Covid-19 mortality in Italian regions during the period analysed. A similar negative relationship

for the Italian regions has been documented in other studies (Arpino et al. 2020; Belloc et al. 2020; Liotta et al. 2020). Given the aggregate level of our analysis, we are not able to infer whether this finding suggests that greater intergenerational support helped to reduce Covid-19 mortality or simply that regions with a higher prevalence of older individuals living in multigenerational households were not particularly affected during the first wave of Covid-19 in Italy. We also found some weak evidence of a negative relationship between prevalence of comorbidities in the population and Covid-19 mortality, possibly due to the fact that regions with a greater prevalence of individuals with comorbidities were hit less harshly by the epidemic during the first wave (Boschi et al. 2021). Moreover, nursing homes have been identified as hotspots of Covid-19-related deaths in Italy (di Giacomo et al. 2020; Trabucchi and De Leo 2020), and employees in healthcare and manufacturing sector have been found to be at greatest risk of Covid-19 (Barbieri et al. 2020; Chirico et al. 2020). In our analysis, these variables were not significantly associated with higher or lower Covid-19 mortality, after controlling for the time trend of the epidemic and random effects. On one hand, some of these discrepancies are related to the contexts of the analyses, as our subnational setting differed from those considered in some of the previous studies. On the other hand, the different methods used may also contribute to the different findings, as several of the previously cited works focused on only one or two covariates of interest and did not control for other factors. From our results, the importance of including random effects when assessing the relationship of different variables with Covid-19 mortality clearly emerges.

A growing body of literature is documenting the positive effects of non-pharmaceutical interventions (NPIs)—for example school closures, travel bans, and lockdowns—on reducing the spread and mortality of Covid-19 (see e.g. Flaxman et al. 2020; Davies et al. 2021; Brauner et al. 2021). In our study, we did not investigate the effect of such measures on Covid-19 mortality, due to the coincident introduction and relaxation of nationwide NPIs during the first wave of the epidemic. A very different region-based approach was implemented by the Italian government after the period that we analysed: from 3 November 2020, regions were divided into three different colours (yellow, orange, and red) according to 21 indicators related to the epidemic, with increasing levels of restrictions for darker colours (Ministero Della Giustizia 2021).

The effectiveness of NPIs in Italy could thus be better assessed by studying this second wave of Covid-19 mortality.

There are some limitations to our study that should be acknowledged. First, the geographical unit of analysis that we used (Italian NUTS-2 regions) may not be the most appropriate one, given that some covariates (e.g. population density, prevalence of older individuals living in multigenerational households) may vary greatly *within* regions. However, Covid-19 deaths and data on some covariates are not available for smaller geographical units (e.g. provinces or municipalities), and sparser data may not be powerful enough to investigate the relationship between Covid-19 mortality and several factors. Second, no information on the age and sex of individuals who have died with Covid-19 at the regional level is provided by the Dipartimento Della Protezione Civile (2021); this does not allow for standardization of mortality rates. Nonetheless, we believe that our study provides important insights into regional differences in Covid-19 mortality during the outbreak of the epidemic. Future work will be directed towards the analysis of excess mortality, as death counts by age group for the last five years have recently been released by Istat (2021).

In conclusion, our study has shed light on the most significant factors associated with Covid-19 mortality in Italian regions during the first wave of the pandemic. In addition to their scientific value, our findings highlight the importance of context-specific analysis, providing a warning regarding the generalizability of Covid-19-related hypotheses and results. Finally, the method that we have proposed in this paper is a novel contribution to the analysis of mortality during epidemics and can be fully replicated and applied to other countries and frameworks (even outside epidemic research) using the code we have made available (Basellini and Camarda 2021).

Notes

- 1 Ugofilippo Basellini is based at the Max Planck Institute for Demographic Research (MPIDR) in Rostock, Germany, and the Institut national d'études démographiques (INED) in Aubervilliers, France. Carlo Giovanni Camarda is based at INED. The two authors contributed equally to this work.
- 2 Please direct all correspondence to Carlo Giovanni Camarda, 9 cours des Humanités, 93322 Aubervilliers, France; or by Email: carlo-giovanni.camarda@ined.fr
- 3 We would like to thank Viviana Egidi, Laurent Toulemon, Marília Nepomuceno, Tiziana Torri, Stefano

Mazzucco, four anonymous reviewers, and the Editor for providing useful comments on a previous version of this manuscript. This paper was presented at the 2021 Annual Meeting of the Population Association of America, to the Laboratory of Digital and Computational Demography at MPIDR, and to the Mortality, Health and Epidemiology Unit at INED.

ORCID

Ugo Filippo Basellini  <http://orcid.org/0000-0003-0292-1404>

Carlo Giovanni Camarda  <http://orcid.org/0000-0003-0191-4093>

References

- Agosto, A., A. Campmas, P. Giudici, and A. Renda. 2020. Monitoring Covid-19 contagion growth in Europe, *CEPS working paper*. Available: <https://www.ceps.eu/ceps-publications/monitoring-covid-19-contagion-growth-in-europe/>
- Arpino, B., V. Bordone, and M. Pasqualini. 2020. No clear association emerges between intergenerational relationships and COVID-19 fatality rates from macro-level analyses, *Proceedings of the National Academy of Sciences* 117(32): 19116–19121. doi:10.1073/pnas.2008581117
- Barbieri, T., G. Basso, and S. Scicchitano. 2020. Italian workers at risk during the Covid-19 epidemic, *SSRN*. Available: <https://ssrn.com/abstract=3660014>
- Basellini, U. and C. G. Camarda. 2021. Modelling COVID-19 mortality at the regional level in Italy. OSF. Available: <https://doi.org/10.17605/OSF.IO/MU7HY>
- Belloc, M., P. Buonanno, F. Drago, R. Galbiati, and P. Pinotti. 2020. Cross-country correlation analysis for research on COVID-19, *Vox-CEPR Policy Portal*. Available: <https://voxeu.org/article/cross-country-correlation-analysis-research-covid-19>. (accessed: 16 April 2021).
- Billari, F. and F. Ongaro. 1998. The transition to adulthood in Italy. Evidence from cross-sectional surveys / Le passage à l'âge adulte en Italie, *Espace, populations, sociétés* 16(2): 165–179. doi: 10.3406/espos.1998.1833
- Bonetti, M. and U. Basellini. 2021. Epilocal: A real-time tool for local epidemic monitoring, *Demographic Research* 44(12): 307–332. doi: 10.4054/DemRes.2021.44.12
- Boschi, T., J. D. Iorio, L. Testa, M. A. Cremona, and F. Chiaromonte. 2021. Functional data analysis characterizes the shapes of the first COVID-19 epidemic wave in Italy, *Scientific Reports* 11: 17054. doi:10.1038/s41598-021-95866-y
- Brandén, M., S. Aradhya, M. Kolk, J. Härkönen, S. Drefahl, B. Malmberg, M. Rostila, et al. 2020. Residential context and COVID-19 mortality among adults aged 70 years and older in Stockholm: A population-based, observational study using individual-level data, *The Lancet Healthy Longevity* 1: e80–e88. doi: 10.1016/S2666-7568(20)30016-7
- Brauner, J. M., S. Mindermann, M. Sharma, D. Johnston, J. Salvatier, T. Gavenčiak, A. B. Stephenson, et al. 2021. Inferring the effectiveness of government interventions against COVID-19, *Science* 371(6531): eabd9338. doi: 10.1126/science.abd9338
- Brillinger, D. R. 1986. A biometrics invited paper with discussion: The natural variability of vital rates and associated statistics, *Biometrics* 42(4): 693–734. doi: 10.2307/2530689
- Cereda, D., M. Tirani, F. Rovida, V. Demicheli and al. 2020. The early phase of the COVID-19 outbreak in Lombardy, Italy, *arXiv*. Available: <https://arxiv.org/abs/2003.09320>
- Chiogna, M., and C. Gaetan. 2020. COVID-19 in Italy. Available: <https://github.com/cgaetan/COVID-19> (accessed 24 August 2020).
- Chirico, F., G. Nucera, and N. Magnavita. 2020. COVID-19: Protecting healthcare workers is a priority, *Infection Control & Hospital Epidemiology* 41(9): 1117–1117. doi: 10.1017/ice.2020.148
- Ciminelli, G., and S. Garcia-Mandicó. 2020. COVID-19 in Italy: An analysis of death registry data, *Journal of Public Health* 42(4): 723–730. doi: 10.1093/pubmed/fdaa165
- Davies, N. G., R. C. Barnard, C. I. Jarvis, T. W. Russell, M. G. Semple, M. Jit, and W. J. Edmunds. 2021. Association of tiered restrictions and a second lockdown with COVID-19 deaths and hospital admissions in England: A modelling study, *The Lancet Infectious Diseases* 21(4): 482–492. doi: 10.1016/S1473-3099(20)30984-1
- di Giacomo, E., G. Bellelli, G. Peschi, S. Scarpetta, F. Colmegna, G. de Girolamo, and M. Clerici. 2020. Management of older people during the COVID19 outbreak: Recommendations from an Italian experience, *International Journal of Geriatric Psychiatry* 35(7): 803–805. doi: 10.1002/gps.5318
- Di Giulio, P., and A. Rosina. 2007. Intergenerational family ties and the diffusion of cohabitation in Italy, *Demographic Research* 16: 441–468. doi: 10.4054/DemRes.2007.16.14
- Dipartimento Della Protezione Civile. 2021. Dataset of COVID-19 infected cases in Italy. Available: <https://github.com/pcm-dpc/COVID-19/tree/master/>. (accessed: 18 May 2021).
- Dowd, J. B., L. Andriano, D. M. Brazel, V. Rotondi, P. Block, X. Ding, Y. Liu, and M. C. Mills. 2020.

- Demographic science aids in understanding the spread and fatality rates of COVID-19, *Proceedings of the National Academy of Sciences* 117(18): 9696–9698. doi: [10.1073/pnas.2004911117](https://doi.org/10.1073/pnas.2004911117)
- Du, Z., L. Wang, S. Cauchemez, X. Xu, X. Wang, B. J. Cowling, and L. A. Meyers. 2020. Risk for transportation of coronavirus disease from Wuhan to other cities in China, *Emerging Infectious Diseases* 26(5): 1049–1052. doi: [10.3201/eid2605.200146](https://doi.org/10.3201/eid2605.200146)
- Eilers, P. H. C., and B. D. Marx. 1996. Flexible smoothing with *B*-splines and penalties (with discussion), *Statistical Science* 11: 89–102. doi:[10.1214/ss/1038425655](https://doi.org/10.1214/ss/1038425655)
- Esteve, A., I. Permanyer, D. Boertien, and J. W. Vaupel. 2020. National age and coresidence patterns shape COVID-19 vulnerability, *Proceedings of the National Academy of Sciences* 117(28): 16118–16120. doi: [10.1073/pnas.2008764117](https://doi.org/10.1073/pnas.2008764117)
- Favero, C. A. 2020. Why is COVID-19 mortality in Lombardy so high? Evidence from the simulation of a SEIHCRC Model, *Covid Economics. Vetted and Real-Time Papers*. Available: <https://iris.unibocconi.it/retrieve/handle/11565/4026151/122470/CovidEconomics4.pdf> (accessed 20 August 2020).
- Flaxman, S., S. Mishra, A. Gandy, H. J. T. Unwin, T. A. Mellan, H. Coupland, C. Whittaker, et al., Imperial College COVID-19 Response Team. 2020. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe, *Nature* 584:257–261. doi: [10.1038/s41586-020-2405-7](https://doi.org/10.1038/s41586-020-2405-7)
- Giorgi, J., and D. Boertien. 2020. The potential impact of co-residence structures on socio-demographic inequalities in COVID-19 mortality, *SocArXiv*. <https://doi.org/10.31235/osf.io/84ygx>
- Hamidi, S., S. Sabouri, and R. Ewing. 2020. Does density aggravate the COVID-19 pandemic?, *Journal of the American Planning Association* 86(4): 495–509. doi: [10.1080/01944363.2020.1777891](https://doi.org/10.1080/01944363.2020.1777891)
- Helliwell, J. F., and R. D. Putnam. 1995. Economic growth and social capital in Italy, *Eastern Economic Journal* 21 (3): 295–307.
- Istat. 2016. Family, Social Subjects and Life Cycle (FSS) survey. <https://www.istat.it/it/archivio/236637>. (accessed 8 June 2020).
- Istat. 2021. Resident population on January 1st and demographic indicators for years 2019 and 2020. <http://dati.istat.it>. (accessed 22 March 2021).
- Istituto Superiore di Sanità. 2020. Survey nazionale sul contagio Covid-19 nelle strutture residenziali e sociosanitarie. Aggiornamento nazionale: 6 aprile 2020 [A survey on COVID-19 infection in long-stay residential care homes. National updated: 6 April 2020]. Available: <https://www.epicentro.iss.it/coronavirus/sars-cov-2-survey-rsa>
- Johns Hopkins University CSSE. 2020. Novel Coronavirus (COVID-19) Cases. <https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases>. (accessed 17 August 2020).
- Kertzer, D. I., M. J. White, L. Bernardi, and G. Gabrielli. 2009. Italy's path to very low fertility: The adequacy of economic and second demographic transition theories, *European Journal of Population / Revue Européenne de Démographie* 25(1): 89–115. doi: [10.1007/s10680-008-9159-5](https://doi.org/10.1007/s10680-008-9159-5)
- Khavarian-Garmsir, A. R., A. Sharifi, and N. Moradpour. 2021. Are high-density districts more vulnerable to the COVID-19 pandemic?, *Sustainable Cities and Society* 70: 102911. doi: [10.1016/j.scs.2021.102911](https://doi.org/10.1016/j.scs.2021.102911)
- Lapolla, P., A. Mingoli, and R. Lee. 2021. Deaths from COVID-19 in healthcare workers in Italy—What can we learn?, *Infection Control & Hospital Epidemiology* 42(3): 364–365. doi: [10.1017/ice.2020.241](https://doi.org/10.1017/ice.2020.241)
- Liotta, G., M. Marazzi, S. Orlando, and L. Palombi. 2020. Is social connectedness a risk factor for the spreading of COVID-19 among older adults? The Italian paradox, *PLoS ONE* 15(5): e0233329. doi: [10.1371/journal.pone.0233329](https://doi.org/10.1371/journal.pone.0233329)
- Maltagliati, M. 2020. COVID-19 in Italia: una o tante epidemie? [COVID-19 in Italy: One or many epidemics?], *Neodemos*. Available: https://www.neodemos.info/articoli/covid_19-in-italia-una-o-tante-epidemie/. (accessed 3 April 2020).
- Martin, C., D. Jenkins, J. Minhas, L. Gray, J. Tang, C. Williams, S. Sze, et al. 2020. Socio-demographic heterogeneity in the prevalence of COVID-19 during lockdown is associated with ethnicity and household size: Results from an observational cohort study, *EClinicalMedicine* 25: 100466. doi: [10.1016/j.eclinm.2020.100466](https://doi.org/10.1016/j.eclinm.2020.100466)
- Ministero Della Giustizia. 2021. Gazzetta Ufficiale della Repubblica Italiana. Volume 161 - 45. Ufficio Pubblicazione Leggi e Decreti [Official Gazette of the Italian Republic]. All COVID-related decrees are available: <https://www.gazzettaufficiale.it/dettaglioArea/12>.
- Ministero della Salute. 2020. Dataset of beds per hospital facility. Available: <http://www.dati.salute.gov.it/dati/dettaglioDataset.jsp?menu=dati&idPag=18>. (accessed 30 March 2020).
- Nepomuceno, M. R., E. Acosta, D. Alburez-Gutierrez, J. M. Aburto, A. Gagnon, and C. M. Turra. 2020. Besides population age structure, health and other demographic factors can contribute to understanding the COVID-19 burden, *Proceedings of the National Academy of Sciences* 117(25): 13881–13883. doi: [10.1073/pnas.2008760117](https://doi.org/10.1073/pnas.2008760117)
- Pengyu, L., L. McQuarrie, Y. Song and C. Colijn. 2021. Modelling the impact of household size distribution

- on the transmission dynamics of COVID-19, *Journal of The Royal Society Interface*. doi: [10.1101/2021.01.12.21249707](https://doi.org/10.1101/2021.01.12.21249707)
- Perperoglou, A., W. Sauerbrei, M. Abrahamowicz, and M. Schmid. 2019. A review of spline function procedures in R, *BMC Medical Research Methodology* 19: 46. doi: [10.1186/s12874-019-0666-3](https://doi.org/10.1186/s12874-019-0666-3)
- R Development Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rocklöv, J., and H. Sjödin. 2020. High population densities catalyse the spread of COVID-19, *Journal of Travel Medicine* 27(3): taaa038. doi: [10.1093/jtm/taaa038](https://doi.org/10.1093/jtm/taaa038)
- Scala, A., A. Flori, A. Spelta, E. Brugnoli, M. Cinelli, W. Quattrocchi, and F. Pammolli. 2020. Time, space and social interactions: Exit mechanisms for the Covid-19 epidemics, *Scientific Reports* 10(1): 13764. doi: [10.1038/s41598-020-70631-9](https://doi.org/10.1038/s41598-020-70631-9)
- Schwarz, G. 1978. Estimating the dimension of a model, *The Annals of Statistics* 6(2): 461–464. doi: [10.1214/aos/1176344136](https://doi.org/10.1214/aos/1176344136)
- Sun, Z., H. Zhang, Y. Yang, H. Wan, and Y. Wang. 2020. Impacts of geographic factors and population density on the COVID-19 spreading under the lockdown policies of China, *Science of the Total Environment* 746: 141347. doi: [10.1016/j.scitotenv.2020.141347](https://doi.org/10.1016/j.scitotenv.2020.141347)
- Sy, K. T. L., L. White, and B. E. Nichols. 2021. Population density and basic reproductive number of COVID-19 across United States counties, *PLoS ONE* 16(4): e0249271. doi: [10.1371/journal.pone.0249271](https://doi.org/10.1371/journal.pone.0249271)
- Trabucchi, M., and D. De Leo. 2020. Nursing homes or besieged castles: COVID-19 in northern Italy, *The Lancet Psychiatry* 7(5): 387–388. doi: [10.1016/S2215-0366\(20\)30149-8](https://doi.org/10.1016/S2215-0366(20)30149-8)
- Volpato, S., F. Landi and R. A. Incalzi, on behalf of the Italian Society of Gerontology and Geriatrics. 2020. A frail health care system for an old population: Lesson from the COVID-19 outbreak in Italy, *The Journals of Gerontology: Series A* 75(9):e126–e127. doi: [10.1093/gerona/glaa087](https://doi.org/10.1093/gerona/glaa087)
- Wong, D. W. S., and Y. Li. 2020. Spreading of COVID-19: Density matters, *PLoS ONE* 15(12): e0242398. doi: [10.1371/journal.pone.0242398](https://doi.org/10.1371/journal.pone.0242398)
- Wood, S. N. 2019. mgcv: Mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation, *R Package Version* 1: 8–31.
- World Health Organization. 2020a. Rolling updates on coronavirus disease (COVID-19). Available: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen>
- World Health Organization. 2020b. Coronavirus disease 2019 (COVID-19): Situation Report – 194. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200801-covid-19-sitrep-194.pdf?sfvrsn=401287f3_2. (accessed 10 August 2020).

Appendix. Additional results

In this Appendix, we present additional results from our analysis.

We start with the exploratory analysis of the data set that we use in our study. Figure A1 shows the distributions (diagonal), correlations (upper-right quadrant) and scatter plots (lower-left quadrant) for the 10 covariates described in the ‘Explanatory variables’ subsection. The straight lines in the lower-left quadrant correspond to linear regression lines between the two variables. We do not include population size in this analysis because the variable enters the regression model as an offset term. Some strong correlations emerge among the variables. For example, the share of population aged 65+ is highly correlated with the share of the population with one or more chronic diseases ($r = 0.685$). Moreover, the number of Covid-19 swab tests per capita is highly correlated with the number of LSRCHs ($r = 0.722$). In our analysis, we keep all these variables, since we investigate all possible combinations and retain the model that minimizes the BIC. In the optimal model (GAMM-opt), these four variables are not retained.

Moreover, Figure A2 shows the linear relationships between Covid-19 log mortality rates and the 10 covariates that we analyse. In this figure, the time-varying values of log mortality rates for each region (y-axis) are plotted against the time-invariant value of each covariate (x-axis). Colours indicate the start date of the epidemic. Some of these relationships display an expected sign: for example, the share of population aged 65+, the share of employees in the manufacturing sector, and the number of LSRCHs per capita (all with a positive linear relationship). Some covariates display an opposite sign from those posited by the literature: for example, population density and intergenerational co-residence. For two variables (prevalence of chronic conditions, number of ICU beds per capita) the relationship is almost null.

Next, Table A1 reports the start date of the analysis and the delay variable by region for both main and sensitivity analyses. In the main analysis, the start date is defined as the date when cumulative cases surpassed 0.0001 per cent of the regional population. In the sensitivity analysis, the start date is defined as the date when cumulative cases surpassed five.

Figures A3 and A4 provide additional details on the goodness of fit of the GAMM-opt model. Figure A3 shows the observed and the GAMM-opt fitted number of Covid-19 deaths, together with 95 per cent confidence intervals in each region, as well as for Italy overall (top-left panel). Moreover, Figure A4 presents the Poisson deviance residuals of this model.

Table A2 reports the results of the sensitivity analysis on the definition of the start date of the epidemic. Here, the start date is computed using the absolute approach: we consider the epidemic to have begun when cumulative cases surpassed five (rather than 0.0001 per cent of the regional population as in the main analysis). Point estimates and 95 per cent confidence intervals are extremely close to those of the main analysis shown in Table 2.

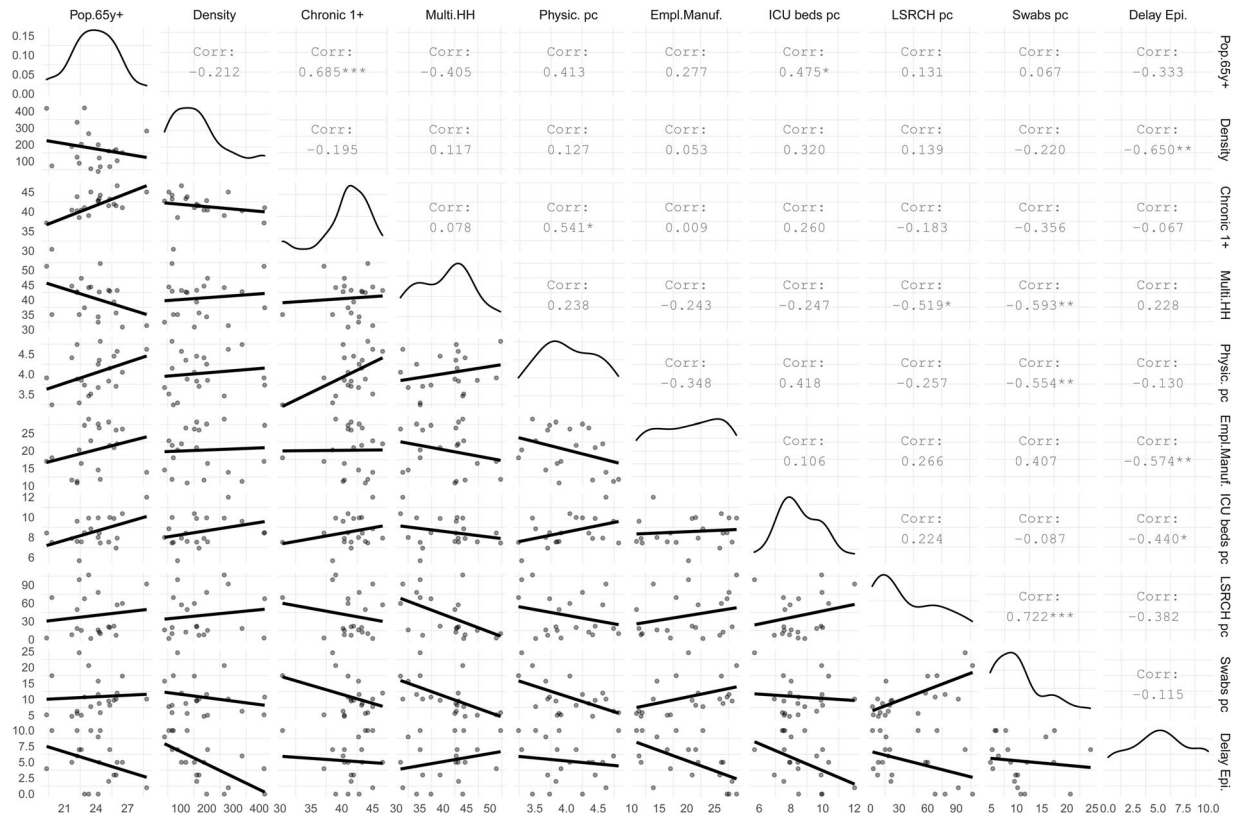


Figure A1 Exploratory analysis of the 10 explanatory variables used in the models: Italy's 21 regions, late February to mid-July 2020

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Notes: Distributions, correlations, and scatter plots are shown in the diagonal, upper-right, and lower-left quadrants, respectively. Labels on the axes correspond to the 10 explanatory variables (see the 'Explanatory variables' subsection for details).

Source: Authors' elaborations on data from Istituto Superiore di Sanità (2020); Ministero della Salute (2020); Dipartimento Della Protezione Civile (2021); Istat (2021).

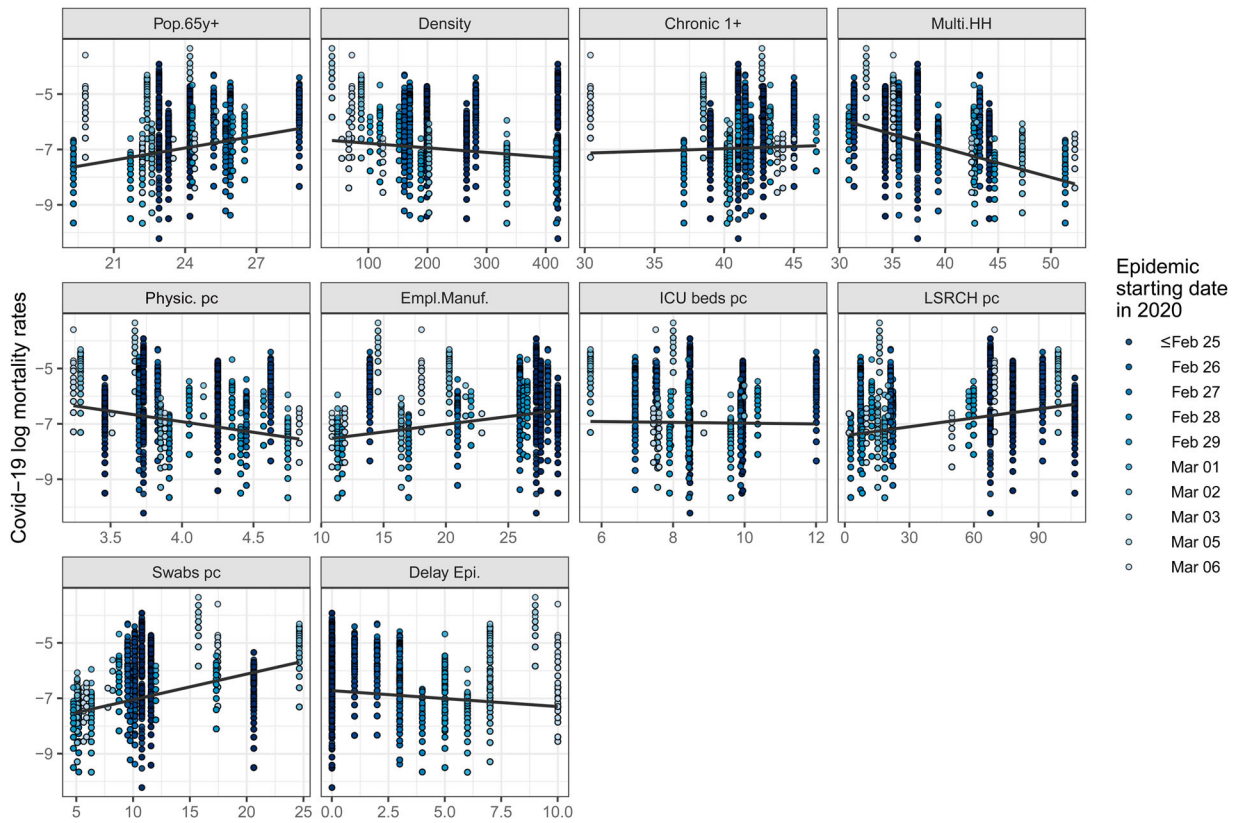


Figure A2 Linear relationship between Covid-19 log mortality rates and the 10 exploratory variables: Italy's 21 regions, late February to mid-July 2020

Notes: Colours correspond to the start date of the analysis. X-axis labels correspond to the panel titles (see 'Explanatory variables' subsection for details of the 10 explanatory variables). Covid-19 mortality rates are time varying, whereas explanatory variables are time invariant.

Source: As for Figure A1.

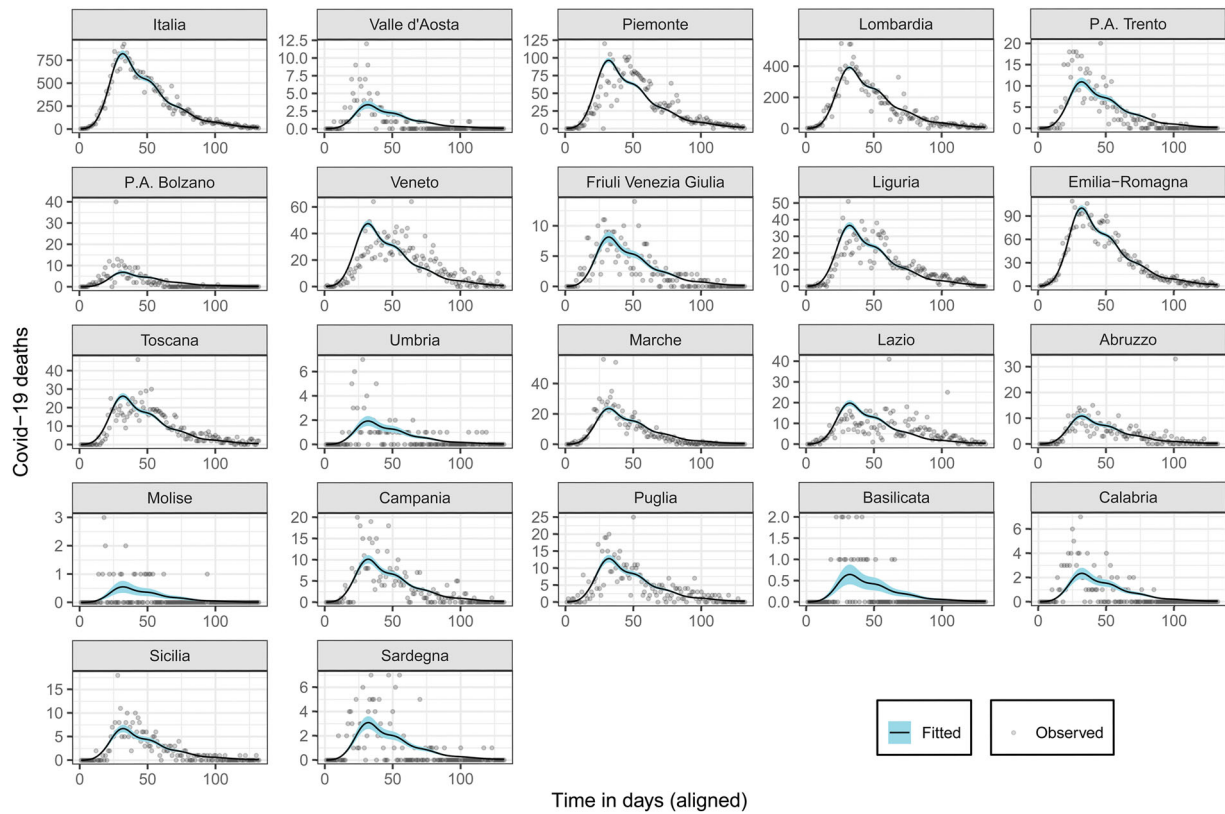


Figure A3 Observed and fitted (with 95 per cent confidence intervals) number of Covid-19 deaths in Italy and its 21 regions for 132 aligned days between 25 February and 15 July 2020

Note: Fitted curves are from the GAMM-opt model.

Source: As for Figure A1.

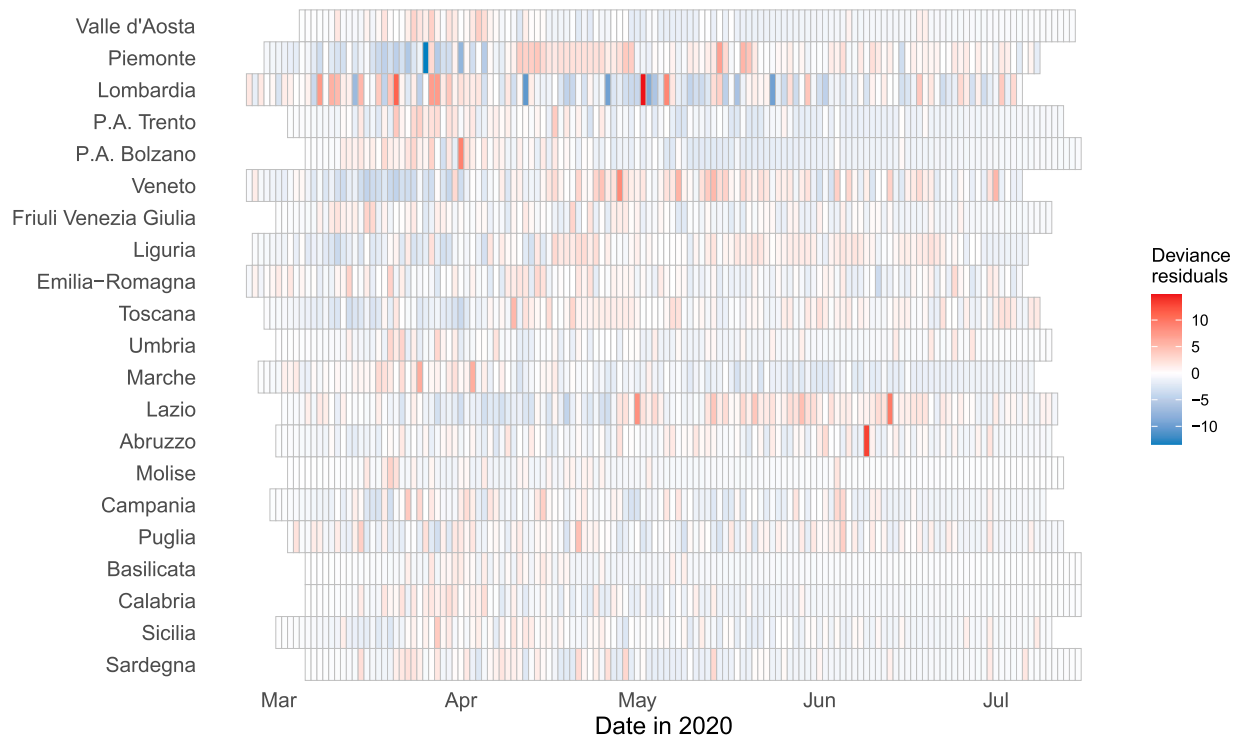


Figure A4 Poisson deviance residuals of the GAMM-opt model for Italy's 21 regions, 25 February to 15 July 2020

Source: As for Figure A1.

Table A1 Start date of the epidemic and number of days of delay in the start of the epidemic in the main analysis (start date defined as when cumulative cases surpassed 0.0001 per cent of the regional population) and sensitivity analysis (start date defined as when cumulative cases surpassed five), by Italian region

Region	Main analysis		Sensitivity analysis	
	Start date in 2020	Delay in start of epidemic (days)	Start date in 2020	Delay in start of epidemic (days)
Valle d'Aosta	5 March	9	6 March	10
Piemonte	28 February	3	28 February	3
Lombardia	25 February	0	25 February	0
P.A. Trento	3 March	7	5 March	9
P.A. Bolzano	6 March	10	7 March	11
Veneto	25 February	0	25 February	0
Friuli Venezia Giulia	1 March	5	1 March	5
Liguria	26 February	1	26 February	1
Emilia/Romagna	25 February	0	25 February	0
Toscana	28 February	3	28 February	3
Umbria	1 March	5	3 March	7
Marche	27 February	2	28 February	3
Lazio	2 March	6	29 February	4
Abruzzo	1 March	5	3 March	7
Molise	3 March	7	5 March	9
Campania	29 February	4	29 February	4
Puglia	3 March	7	3 March	7
Basilicata	6 March	10	10 March	14
Calabria	6 March	10	8 March	12
Sicilia	1 March	5	1 March	5
Sardegna	6 March	10	8 March	12

Source: Authors' own elaborations on data from Dipartimento Della Protezione Civile (2021).

Table A2 Sensitivity analysis on the definition of the start date of the epidemic (defined here as when cumulative cases surpassed five) for the five models used in the main analysis reported in Table 2: Italy's 21 regions, late February to mid-July 2020

	Model			
	GAM and GAM-opt	GAMM w/o covars	GAMM	GAMM-opt
Population aged 65+ years	0.53 (0.48, 0.58)	–	0.31 (–0.50, 1.13)	–
Population density	0.31 (0.28, 0.33)	–	0.18 (–0.63, 0.99)	–
Population with one or more chronic diseases	–0.48 (–0.55, –0.40)	–	–0.15 (–0.81, 0.51)	–
Older people living in multigenerational households	–0.52 (–0.55, –0.49)	–	–0.63 (–1.17, –0.08)	–0.82 (–1.11, –0.53)
Physicians per capita	0.21 (0.15, 0.27)	–	–0.08 (–0.73, 0.56)	–
Employees manufacturing sector	0.40 (0.36, 0.44)	–	0.06 (–0.52, 0.63)	–
ICU beds per capita	–0.62 (–0.66, –0.58)	–	–0.40 (–0.88, 0.09)	–0.42 (–0.74, –0.10)
LSRCHs per capita	0.62 (0.57, 0.68)	–	–0.06 (–0.68, 0.55)	–
Cumulative tests per capita	–0.45 (–0.50, –0.39)	–	0.19 (–0.59, 0.97)	–
Delay in the epidemic	–0.17 (–0.22, –0.12)	–	–0.42 (–1.33, 0.49)	–0.6 (–0.93, –0.28)
Observations	2,668	2,668	2,668	2,668
Regions	21	21	21	21
Deviance	8,915.60	7,391.15	7,391.10	7,391.23
Effective dimension	24.34	34.52	34.49	34.35
BIC	9,107.77	7,663.70	7,663.44	7,662.46
Deviance explained (%)	87.4	89.5	89.5	89.5

Notes: The dependent variable is Covid-19 deaths. All models include a smoothed function of time and person-days of exposure as an offset. In the main analysis the start date was defined as when cumulative cases surpassed 0.0001 per cent of the regional population.

Source: As for Table 2.