

Cross-regional variations of Covid-19 mortality in Italy: an ecological study.

Authors:

Cristina Oliva, M.D. ¹, Francesco Di Maddaloni, PhD ¹, Andrea Marcellusi, PhD ^{1,2} and Giampiero Favato, DBA ^{1*}

¹ Institute of Leadership and Management in Health (ILMH), Kingston University London.

² Università Tor Vergata, Rome

*Corresponding Author.

Corresponding Author

Prof. Giampiero Favato

Director, Institute of Leadership and Management in Health (ILMH), Kingston University London.

Web page: <https://www.kingston.ac.uk/faculties/faculty-of-business-and-social-sciences/research/institute-of-leadership-and-management-in-health/>

Mailing address:

Kingston University London
Kingston Hill Campus,
Kingston Hill,
Kingston upon Thames KT2 7LB,
England

E-mail:

Primary: g.favato@kingston.ac.uk

Alternative: favatog@gmail.com

Telephone: +44 (0) 7901 114824

Abstract

Background

Disparities in cross-regional Covid-19 mortality remain poorly understood. The association between pre-epidemic health and epidemic mortality can inform a policy response to future outbreaks.

Method:

We conducted an ecological study of the association between the cumulative deaths attributed to Covid-19 epidemic in the twenty Italian regions and nine determinants of population health derived from a systematic review of the literature. We used a multiple least square regression to predict the cross-regional variation in mortality observed from the onset of the epidemic to September 23rd, 2020.

Results:

Four independent variables best explained the cross-regional differences in the number of deaths attributed to Covid-19: the force of infection, population density, number of elderly living in assisted facilities and the standard rate of diabetes. The semi-partial correlation coefficients suggest that the force of infection and the number of elderly residents in nursing homes were the dominant predictors of the number of deaths attributed to Covid-19. Statistical controls and validation confirmed the generalisability of the predictive model.

Conclusions:

Our findings indicate that a significant reduction of social contacts in main metropolitan areas and the timely isolation of elderly and diabetic residents could significantly reduce the death toll of the next wave of Covid-19 infection in Italy.

Introduction

The uncertainty of the second wave of Covid-19 infections hangs over a society already physically and mentally exhausted by lockdowns and social distancing. Future scenarios for the next Covid-19 infection vary from localised outbreaks to a new, stronger epidemic [1]. Policy-makers seem to rely on targeted lockdowns as a trade-off between reducing the risk of further outbreaks and allowing for some degree of social life. There is still no consensus, though, on a public policy to respond to a new wave of Covid-19.

The Covid-19 outbreak found most Countries unprepared to face a fast-spreading threat to public health [2]. The first Italian case of Covid-19 secondary transmission was identified in Codogno, a town close to Milan, on February 18th 2020. Only two hundred days later, on September 23rd, 2020, 302,537 confirmed cases and 35,758 deaths were attributable to Covid-19 infection. The twenty Italian regions paid a different toll to Covid-19. Mortality rates (deaths over 100,000 residents) and fatality rates (deaths over infected patients) observed in the Italian regions vary widely around the national mean. Mortality rates ranged from 168 deaths over 100,000 residents in Lombardy to just about 5 in Sicily, Calabria and Basilicata. Fatality rates show the same variability, from 0.16 in Lombardy to 0.04 in Umbria, Molise, Campania, Basilicata, and Sardinia [3].

Disparities in cross-regional Covid-19 epidemic mortality remain poorly understood. This ecological study uses cumulative regional data on the observed number of deaths attributed to Covid-19 to explore the determinants of epidemic mortality. The variables of the model (the “predictors”) could inform a data-driven response to the next wave of Covid-19.

Figure 1 Cumulative mortality and fatality rates stratified by Italian regions from the onset of the epidemic to September 23rd, 2020 (data source: Italian MinSan).

Method

The data underlying this article are available in the article and its online supplementary material.

Study design

We performed an ecological study of associations between the cumulative number of deaths attributed to Covid-19 and the determinants of population health derived from a systematic review of the literature. The number of regionally-stratified deaths attributable to Covid-19 was derived from the civil protection database and referred to the epidemic period between January 1st, 2020 and September 23rd, 2020. We defined the Covid-19 deaths as the “mortality space” in which their values along its dimensions could locate each of the twenty Italian regions. We then fitted an equation whose form was adequate to predict the number of Covid-19 deaths by Region with the smallest margin of error. The “mortality space” will then be defined by the predictors, the coefficients of this equation [4].

This analysis is divided into four main methodological steps: literature review, variable selection, regression analysis and model validation. Considering the previously published models of Covid-19 mortality, the authors showed three critical risks of bias: the selection of predictors, the method of analysis and, most importantly, the lack of validation of the model. [5]. To mitigate “a priori” these known sources of bias, we firstly performed a systematic review of the literature to inform the selection of predictors and the method of analysis.

The systematic review was conducted in adherence to the PRISMA guidelines [6], and the search protocol was published in the International Prospective Register of Systematic Reviews (PROSPERO) [7]. We identified and critically appraised fifty-six studies reporting predictors of mortality attributable to Covid-19. We extracted twelve potential predictors of Covid-19 mortality from the models included in the review, listed below in descending order of frequency (in parenthesis): age (twenty-eight), population tested/swabs administered (twenty), pre-existing medical conditions (nineteen), the severity of Covid-19 outbreak (fifteen), gender (thirteen), exposure to air pollution (twelve), hospital resources/health spending (eleven), Gross Domestic Product (GDP) per capita/income inequality/deprivation (eleven), ethnicity (seven), population density (six), climate (four), number of elderly residents in assisted living facilities (one).

Our second methodological step was to transform the twelve predictors identified by the systematic review into inputs to inform the predictive model. This required the adaptation of the predictors to the granularity of data available at Italian regional level. The process generated seventeen independent variables grouped by five main domains: five demographic variables (age, gender, population density, ethnicity, and elderly living in assisting facilities), three economic variables (GDP per capita, income inequality, and public expenditures in healthcare), two variables related to Covid-19 infection (force of infection and number of swabs carried out), five variables describing the population’s pre-existing medical conditions (standard rates of mortality, diabetes, hypertension, obesity, and comorbidity), and two environmental variables (air pollution and climate). For each predictor considered, data for 20 Italian regions were extracted from primary national sources (see supplementary material Table 4).

Thirdly, a univariate correlation analysis was performed to evaluate its association with the cumulative deaths attributed to Covid-19. The variables showing a significant association ($P \leq 0.05$) with the cumulative deaths attributable to Covid-19 were included in the multiple regression. Predicted and actual values were then compared to validate the predictive accuracy of the model.

Table 1. Inputs to the predictive model.

Statistical analysis

A univariate correlation analysis was performed to test the statistical association between the number of deaths attributed to Covid-19 in each Region and the seventeen independent variables identified by the literature review. Nine independent variables, whose correlation coefficient showed a significance level of $P \leq 0.05$, were included in the final regression model (Table 1).

We used multiple least square regression to estimate the coefficients of the predictive model. Since the variables' values spanned nine orders of magnitude (from millions to decimals), we transformed all inputs to the model into their natural logarithms [8]. A step-forward selection was adopted to add to the model the relevant variables (Person correlation index with a p-value < 0.05). All selected variables were added one at a time beginning with the predictor with the highest correlation with the dependent variable. If the added variable did not contribute to improving the goodness-of-fit of the regression, it was then excluded from the regression model [9].

Finally, statistical controls and a validation process were used to test the generalisability of our model. We calculated the zero-order simple regression coefficients, and we then reported the statistical controls related to the least-squares multiple regression. The accuracy of the predictive model was measured as the mean absolute percentage error (MAPE), or mean error between the actual and the predicted values of the dependent variable [10]. We used the following scale for the comparison and interpretation of MAPE values: $MAPE < 10$, highly accurate forecasting; ≥ 10 $MAPE \leq 20$, good forecasting; ≥ 20 $MAPE \leq 50$, reasonable forecasting; > 50 , inaccurate forecasting [11].

The predictive performance was tested following a validation process, aimed to avoid the “ecological fallacy”, which occurs when associations which exist for groups are assumed to also be true for individuals [12]. We validated the selection of variables included in the model using randomly generated “training sets”. The accuracy of the model was assessed by comparing the mean MAPE error of the original regression and the one obtained from the training set.

Results

Univariate association of all the independent variables identified by the review of the literature.

The univariate correlation values confirmed a strong correlation between nine predictors and the number of deaths attributed to Covid-19, as reported in Table 1. The univariate analysis seemed to question the predictive validity of some variables frequently used to model Covid-19 mortality. Population ageing, gender and the rate of comorbidity resulted poorly or negatively correlated to the number of deaths attributable to Covid-19.

Multiple Least Square Regression.

The nine predictors included in the model were regressed against the cumulative number of deaths attributable to Covid-19 by each of the 20 Italian regions. Dependent and independent variables' values are reported in Table 2.

Four independent variables best predicted the number of deaths attributed to Covid-19: the force of infection (attack rate), the number of elderly living in assisted facilities, the population density, and the standard rate of diabetes.

The predictive model equation was:

$$\ln(y) = -14.9165 + 1.2950 \ln(\text{attack rate}) + 0.7841 \ln(\text{elderly in RSA}) + 0.5985 \ln(\text{population density}) + 2.0941 \ln(\text{standard rate of diabetes}) \quad [1]$$

The sign of the correlation was expected, with all predictors positively correlated to the dependent variable.

Predictors and regression results are reported in Table 3. Compared to the actual number of deaths attributed to Covid-19 in each of the twenty Italian regions, the values predicted by the model showed a Mean Average Percentage Error (MAPE) = 4.65 (standard deviation 0.33; standard error of the mean = 0.07). The Lewis scale [11] rated the predictive model as “highly accurate”. Only Molise, one of the smallest Italian Region, showed a MAPE value above 10.

Table 3. Model predictors and regression results by Italian Region

Statistical controls

The regression statistics confirmed that the model selected a parsimonious number of independent variables (n=4), significantly correlated to the dependent variable (coefficients P values < 0.05) and modestly auto-correlated (VIF values <4). The predictive model showed an elevated coefficient of determination (adjusted R² = 0.95) and a high level of significance (P< 0.0001). Residuals were approximately normally distributed (P=0.3967).

The semi-partial correlation coefficients provided an indication of the dependent variable's variance that each independent variable explained. The force on infection and number of elderly residents in hospices and retirement homes were the dominant predictor of deaths attributed to Covid-19 (r semi-partial = 0.31), followed by the population density (r semi-partial = 0.14) and the standard rate of diabetes (r semi-partial = 0.11).

Validation

We validated the selection of the predictive variables in the regression equation: attack rate, elderly living in assisted homes, population density and standard rate of diabetes. We randomly allocated 70% of the available dataset (sixteen regions), to "the training set", and use the remainder of the dataset (4 Regions) as a validation set. We repeated the random allocation of variables until each Region was included in the validation set at least once. The mean MAPE value obtained from the nine random validation tests (10.99; resampled C.I 8.5 -14.3) was higher than the error of the predictive model (MAPE = 4.65). This was expected since we reduced the regressions' degrees of freedom from 19 to 14. The mean MAPE from the nine random tests, though, confirmed a high level of accuracy with only four out of nine tests reporting a MAPE value exceeding the average.

In conclusion, statistical controls and validation confirmed the robustness, accuracy and generalisability of our predictive model.

Discussion

Main findings of the study

Our analysis focused on the role of pre-existing determinants of public health in explaining the wide variation in cross-regional mortality attributable to Covid-19 in Italy. The focus was motivated by the outcomes of a review of the recent literature on Covid-19 infection modelling. The regression model showed that four predictors (force of infection, number of elderly living in assisted facilities, population density and standard rate of diabetes) could explain over 95% of the differences in cross-regional mortality observed in Italy from the onset of the epidemic to September 23rd, 2020.

What is already known on the determinants of Covid-19 mortality

Earlier reports suggest that elderly patients, patients with comorbidities (chronic obstructive pulmonary disease, cardiovascular disease, hypertension), and patients presenting with dyspnoea are vulnerable to more severe morbidity and mortality after Covid-19 infection [5]. In the systematic review of prediction models related to the Covid-19 mortality, we identified and critically appraised 56 studies to extract 12 candidate predictors. The majority of studies developed new models, but only a few reported information on the selection of the independent variables or proceeded to validate the predictive model with external data sample or training sets.

What this study adds.

We found that Covid-19 is an unequal killer: when its force increases, the frail living in highly populated areas are the most vulnerable to death. These results correspond with previously published studies on the association between Covid-19 mortality and pre-existing determinants of public health. The positive correlation between the force of Covid-19 infection and the number of deaths confirmed the findings of transmission models included in our systematic review [13]. Residents living in areas with high population density have a higher probability of coming into close contact with others and, consequently, any contagious disease is expected to spread more rapidly [14]. Italy's rural landscape can be classified into four types, according to the intensity of energy inputs used in the agricultural process, socioeconomic and environmental features. Italians living in underpopulated, rural areas are less exposed to Covid-19, despite a high number of elderly residents. The lowest energy-intensive landscapes have an average of 49 infected per square kilometre and 28 per 10.000 inhabitants, compared to 134 per square kilometre and 37 per 10.000 inhabitants in more energy-intensive zones [15]. Mortality data related to the first wave of Covid-19 infection (from February 1st to May 12th, 2020) confirm excess mortality of 61% in the main cities of Northern Italy. Excess mortality was calculated as the difference between the observed mortality in the period and the five-year mean. Covid-19, as the cause of death, explained 80% of the difference. Excess mortality was growing with age: from +37% in the 65-74 year range to +59% in the 75-84 range and up to +75% for the elderly over 85 [16].

The third variable of the regression analysis proposes a further insight into the association between age and Covid-19 mortality. The number of elderly living in hospices and retirement homes was one the dominant predictor of deaths attributed to Covid-19 (r semi-partial = 0.31). Anecdotal evidence confirmed an abnormal number of deaths among elderly living in assisted facilities during the peak of the epidemic outbreak. [17,18] A survey of a significant sample of Italian nursing homes (1,356 out of 3,417) was conducted from February 1st to May 5th, 2020 [19]. 97,521 elderly were living in nursing homes, 75,710 (78%) of which in the North of Italy, and 26,981 (28%) in Lombardy, the Region with the highest number of deaths attributable to Covid-19. During the observation period, 3,092 deaths were attributable to Covid-19 infection, 1,807 (48%) of which reported in Lombardy alone. During the observation period, 5,292 elderly residents had to be hospitalised: 965 (18%) of them were Covid-19 positive, while 2,021 (38%) reported symptoms consistent with a Covid-19 infection. If positive to Covid-19, only 48% of the residents could be isolated in a single room, 47% of them remained in rooms with multiple beds, while only 5% were transferred to a dedicated structure. Relatives and visitors accessed the premises of most of the nursing homes without any precaution until the end of February 2020. Most of the respondents complained about the lack of personal protection equipment (PPE) and of clear procedures to contain the Covid-19 infection. Consequently, a striking 21.1% of nursing home staff resulted positive to Covid-19. Infected patients in "post-acute" stage were discharged from hospitals to nursing homes for their rehabilitation, to make room for more severe patients [20]. From

March 8th, 2020, an undisclosed number of Covid-19 patients in Lombardy were transferred to local nursing home facilities to ease the pressure on the intensive care units of the “hubs”, the hospitals designated to treat Covid-19 severe patients in the Region.

Diabetes was not recognised at the onset of the Covid-19 epidemic as a determinant of mortality. Early observations from the countries most affected by the Covid-19 epidemic, including China, United States and Italy, seemed to indicate that prevalence of diabetes among patients affected by Covid-19 was not higher than that observed in the general population, thus suggesting that diabetes was not a risk factor for Covid-19 infection. However, a large body of evidence demonstrated that diabetes was a risk factor for disease progression towards critical illness, development of acute respiratory distress syndrome, need for mechanical ventilation or admission to intensive care unit, and ultimately death [21]. Diabetes patients should be regarded as a particularly vulnerable group for which specific strategies must be implemented, including an extensive serological screening and early containment measures [22].

Lastly, the aggregation of predicted values into two clusters (Northern and Southern Regions) raises a fundamental question about the effectiveness of the lockdown imposed on the population mostly affected by the first wave of Covid-19 infection in Italy. The model seems to accurately predict the cross-regional differences in mortality in the Northern Regions (MAPE= 2.87). Hence the four independent variables are highly associated with the number of deaths attributable to Covid-19. Table 3 shows the aggregate values of each predictor for the Northern and Southern Regions. The mean value of the Covid-19 attack rate is four times higher in the North (823 v. 228), while the total number of elderly residents in RSA is more than double in the North compared to the South (212,395 v. 88,654). The population density in the North is approximately 30% higher in the North, while the standard rate of diabetes tips the scale in favour of the South. Keeping in mind that association does not imply causality, was the successful isolation of the Southern Regions the outcome of the lockdown? Alternatively, did the lockdown create a deadly inequality, by failing to isolate vulnerable individuals in the highly populated “red zones”, where the force of Covid-19 infection continued to grow?

Limitations of the study

The main limitation of our study is related to the risk of bias of the unpublished publications included in the systematic review. Two factors mitigate the risk of bias: extraction methods and time of the review. Firstly, the objective of our systematic review was limited to create a comprehensive repository of variables potentially relevant to the number of deaths attributable to the Covid-19 epidemic. No data from these studies were used to inform our analysis. Secondly, since the first outbreak of Covid-19 was disclosed in January 2020, most of the models relevant to Covid-19 were still going through a peer-review process at the time we performed the systematic review.

Conclusions

Understanding the relationship between pre-epidemic health and epidemic mortality can provide data-driven inputs to inform the policy response aimed to contain the death toll imposed by the next outbreak of Covid-19 infection.

The enforcement of facial protection, social-distancing and targeted lockdown in highly populated areas, where the probability of contagion is highest, can significantly reduce the number of deaths attributable to Covid-19 infection. The adherence to lockdown can be extremely difficult for underprivileged individuals, consequently increasing the overall mortality of Covid-19 infection. Welfare support of six hundred euros for self-employed individuals was approved by the Italian Government in late March 2020 [23]. Still, the pay-out was delayed by red tape until the beginning of June for most of the entitled individuals. A furlough scheme for employed workers followed the same fate, and its pay-out was delayed by months. Many employers advanced their employees the pay-out, using the company and private financial resources to ease the economic hardship of employees and their families. An agile welfare scheme, promptly accessible to underprivileged residents, would significantly improve the effective implementation of total or partial social isolation.

Vulnerable individuals should be closely monitored and safely isolated to shield them from exposure to the Covid-19. Elderly living in assisted living facilities and diabetic patients should be continuously monitored by qualified medical and nursing staff, provided with adequate personal protection equipment.

Our findings indicate that a significant reduction of social contacts in the main metropolitan areas and the timely isolation of elderly and diabetic individuals could significantly reduce the death toll of the next wave of Covid-19 infection in Italy.

References

1. Osterholm M T, and the CIDRAP Viewpoint working group. COVID-19: The CIDRAP Viewpoint
Part 1: The Future of the COVID-19 Pandemic: Lessons Learned from Pandemic Influenza. 2020. Available at: https://www.cidrap.umn.edu/sites/default/files/public/downloads/cidrap-covid19-viewpoint-part1_0.pdf
2. World Health Organization (WHO). Thematic Paper on the Status of Country Preparedness Capacities Background report commissioned by the Global Preparedness Monitoring Board (GPMB) 2020. Available at: https://apps.who.int/gpmb/assets/thematic_papers/tr-2.pdf
3. Dipartimento della Protezione Civile. Aggiornamento casi COVID-19: Settembre 23rd, 2020. Available at: <http://opendatadpc.maps.arcgis.com/apps/opsdashboard/index.html#/b0c68bce2cce478eaac82fe38d4138b1>
4. Anson J. Model Mortality Patterns: A Parametric Evaluation. *Population Studies* 1991; **45**(1): 137-153.
5. Wynants L, Van Calster B, Collins G S, et al. Prediction models for diagnosis and prognosis of COVID-19: systematic review and critical appraisal. *BMJ* 2020; **369**:m1328.
6. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med* **6**(7): e1000097. doi:10.1371/journal.pmed1000097
7. Favato G, Oliva C, Di Maddaloni F. Factors associated with COVID-19 case-fatality and mortality rates: a systematic review. PROSPERO 2020 CRD42020188240. Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020188240
8. Liang D, Shi L, Zhao J, Liu P, Schwartz J, Gao S, et al. Urban Air Pollution May Enhance COVID-19 Case-Fatality and Mortality Rates in the United States. *medRxiv* [Internet]. 2020 January 1st;2020.05.04.20090746. Available from: <http://medrxiv.org/content/early/2020/05/07/2020.05.04.20090746.abstract>
9. Blanchet, F. Guillaume, Pierre Legendre, and Daniel Borcard. Forward Selection of Explanatory Variables. *Ecology* 2008; **89** (9): 2623-632.
10. Mayers, J H, and Forgy, E W. The Development of numerical credit evaluation systems. *Journal of the American Statistical Association* 1993; **58**(303): 799–806.
11. Lewis, C D. *Industrial and business forecasting methods : a practical guide to exponential smoothing and curve fitting*. London ; Boston : Butterworth Scientific, 1982
12. Steyerberg E W. *Clinical prediction models: a practical approach to development, validation, and updating*. Springer US, 2019. doi:10.1007/978-3-030-16399-0

13. Ciminelli G, Garcia-Mandicó S. Mitigation Policies and Emergency Care Management in Europe's Ground Zero for COVID-19. *medRxiv* [Internet]. 2020 January 1st;2020.05.19.20106575. Available at: <http://medrxiv.org/content/early/2020/05/26/2020.05.19.20106575.abstract>
14. Kadi N, Khelfaoui M. Population density, a factor in the spread of COVID-19 in Algeria: statistic study. *Bulletin of the National Research Centre* 2020; 44(1), 138.
15. Agnoletti M, Manganelli S, Piras F. Covid-19 and rural landscape: The case of Italy. *Landscape and Urban Planning* 2020; 204 Available online at: <https://doi.org/10.1016/j.landurbplan.2020.103955>
16. Italian Ministry of Health. Andamento della mortalità giornaliera (SiSMG) nelle città italiane in relazione all'epidemia di Covid-19. Final report: September 1st – May 12th, 2020. Available at: http://deplazio.net/images/stories/SISMG/SISMG_COVID19.pdf
17. Comas-Herrera A, and Zalakain J (2020) Mortality associated with COVID-19 outbreaks in care homes: early international evidence. 2020. Available at: <https://ltccovid.org/2020/04/12/mortality-associated-with-COVID-19-outbreaks-in-care-homes-early-international-evidence/>
18. Lorenz-Dant K (2020) *Report on the COVID-19 long-term care situation in Germany*. 2020. Available at: <https://ltccovid.org/2020/04/15/report-on-the-COVID-19-long-term-care-situation-in-germany/>
19. ISS. Survey nazionale sul contagio COVID-19 nelle strutture residenziali e socio-sanitarie, Istituto Superiore di Sanità. Epidemia COVID-19, Final report: May 5th, 2020. Available at: <https://www.epicentro.iss.it/coronavirus/pdf/sars-cov-2-survey-rsa-rapporto-finale.pdf>
20. Regione Lombardia. DELIBERAZIONE N° XI / 2906 Seduta del 08/03/2020. Available at: <https://www.regione.lombardia.it/wps/wcm/connect/5e0deec4-caca-409c-825b-25f781d8756c/DGR+2906+8+marzo+2020.pdf?MOD=AJPERES&CACHEID=ROOTWORKSPACE-5e0deec4-caca-409c-825b-25f781d8756c-n7b4IOB>
21. Pugliese G, Vitale M, Resi V, Orsi E. Is diabetes mellitus a risk factor for COroNaVIrus Disease 19 (COVID-19)? *Acta Diabetol.* 2020 Nov;57(11):1275-1285. doi: 10.1007/s00592-020-01586-6. Epub 2020 Aug 31. PMID: 32865671; PMCID: PMC7456750.
22. Apicella M, Campopiano MC, Mantuano M, Mazoni L, Del Prato S. Guida pratica alla prevenzione e gestione dell'infezione da COVID-19 nelle persone con diabete. *L'Endocrinologo.* 2020 Oct 23:1–5. Italian. doi: 10.1007/s40619-020-00767-3. Epub ahead of print. PMCID: PMC7582423.
23. DPCM Decree 17 marzo 2020, n. 18 Misure di potenziamento del Servizio sanitario nazionale e di sostegno economico per famiglie, lavoratori e imprese connesse all'emergenza epidemiologica da COVID-19. (20G00034) (GU Serie Generale n.70 del 17-03-2020). Available at: <https://www.gazzettaufficiale.it/eli/id/2020/03/17/20G00034/sg>

Table 1. Selection of predicting variables.

INDEPENDENT VARIABLES	DEPENDENT VARIABLE: DEATHS ATTRIBUTABLE TO COVID-19		
	Correlation coefficient r	Significance level*	95% Confidence Interval of r
Elderly in assisted living facilities	0.8993	P<0.0001	0.7588 to 0.9599
Total number of swabs administered	0.8381	P<0.0001	0.6288 to 0.9342
Population density	0.7084	P=0.0005	0.3873 to 0.8762
GINI index	-0.1259	P=0.5968	-0.5384 to 0.3353
Exposure to pollution	0.7407	P=0.0002	0.4436 to 0.8911
Attack rate	0.6815	P=0.0009	0.3421 to 0.8636
Healthcare public expenditure per capita	0.01704	P=0.9431	-0.4287 to 0.4561
Population > 75 as % of total population	0.0721	P=0.7626	-0.3826 to 0.4987
Males as % of population > 75	-0.3808	P=0.0976	-0.7046 to 0.07423
GDP per capita	0.6269	P=0.0031	0.2552 to 0.8372
African and Asian residents	0.8463	P<0.0001	0.6455 to 0.9377
Diff max temperature from mean C° Feb 2020	0.1131	P=0.6349	-0.3468 to 0.5291
Std mortality rate	-0.05728	P=0.8104	-0.4875 to 0.3953
Co-morbidity rate	-0.4189	P=0.0660	-0.7267 to 0.02898
Std rate diabetes	-0.4625	P=0.0401	-0.7512 to -0.02505
Std rate hyperthension	-0.1890	P=0.4249	-0.5828 to 0.2767
Std rate obesity	-0.5094	P=0.0218	-0.7768 to 0.08638

* Green boxes identify significance level acceptable to inclusion into the model ($P \leq 0.05$)

Table 2. Inputs to the predictive model (non-transformed values).

ITALIAN REGIONS	Dependent variable	Independent variables included in the predictive model								
	Deaths attributed to COVID-19 from onset to 23/09/2020	Population density: 100,000 residents per square kilometre	Number of African and Asian residents	Elderly in assisted living facilities	Gross Domestic Product (GDP) per capita (Euro)	Attack rate: COVID-19 confirmed cases per 100,000 residents	Cumulative number of swabs from onset to 06/06/2020	Standard rate		Exposure to air pollution (number of 2019 days over the limit times residents exposed)
								Diabetes	Obesity	
Piedmont	4157	172	665,780	36,279	30,300	793.7	682,282	4.9	39.3	158,856,813
Valle d'Aosta	146	39	3,665	992	35,200	1021.8	27,977	4.9	39.9	1,156,272
Liguria	1594	286	79,795	11,085	29,678	823.5	291,936	4.4	41.3	26,727,702
Lombardy	16925	422	711,779	78,306	38,200	1045.9	1,990,912	4.7	39.3	234,554,214
Trentino Alto Adige	697	79	34,174	8,326	39,200	853.2	383,677	3.5	38.3	5,441,248
Veneto	2167	267	180,449	37,073	33,100	533.1	1,820,101	4.4	43.2	107,437,139
Friuli Venezia Giulia	350	153	34,709	11,343	31,000	367.1	388,810	5.0	40.2	4,470,953
Emilia Romagna	4479	199	268,097	28,991	35,300	776.1	1,110,287	5.2	43.4	137,369,149
Tuscany	1153	162	192,466	17,864	30,500	377.0	694,204	5.2	43.2	17,941,648
Umbria	85	104	34,520	2,516	24,300	256.9	192,524	5.3	42.8	5,205,203
Marche	989	162	61,899	7,067	26,600	511.5	236,514	4.7	39.1	1,162,392
Lazio	902	341	303,854	15,442	32,900	254.7	810,809	6.0	44.7	80,247,263
Abruzzo	477	121	27,699	7,721	24,400	324.0	189,052	5.4	47.8	6,335,168
Molise	23	69	6,716	1,735	19,500	203.8	40,516	5.6	51.1	1,398,397
Campania	457	424	134,193	3,328	18,200	188.0	552,231	6.8	51.5	40,938,370
Pulia	583	206	56,066	8,052	18,000	179.5	385,490	6.3	50.2	7,868,080
Basilicata	28	56	8,755	1,197	20,800	118.9	68,081	6.6	48.9	3,338,450
Calabria	98	128	47,464	3,910	17,100	96.9	190,031	7.3	48.0	1,412,670
Sicily	303	194	108,285	14,856	17,400	124.7	448,412	5.9	47.4	1,350,200
Sardinia	145	68	27,742	4,966	20,300	207.7	175,829	5.3	42.1	37,069,268

Table 3. Model predictors and regression results by Italian Region

ITALIAN REGIONS	MODEL PREDICTIONS									PREDICTORS			
	NATURAL LOG VALUES		Mean Absolute Percent Error (MAPE)	ACTUAL, PREDICTED AND RELATIVE VALUES						ACTUAL VALUES			
	Deaths attributed to COVID-19	Deaths predicted by the model		Deaths attributed to COVID-19	Deaths predicted by the model	Deaths attributed to COVID-19 (per 100,000 residents)	Deaths predicted by the model (per 100,000 residents)	Mortality rate (deaths/Covid-19 cases)	Mortality rate predicted by the model (deaths/Covid-19 cases)	Attack rate (Covid-19 cases per 100,000 residents)	Totale anziani in RSA	Population density (residents per km ²)	Incidence of diabetes (standard rate)
Lombardy	9.74	9.78	0.46	16925	17694	168	176	0.16	0.17	1046	78306	422	4.70
Valle d'Aosta	4.98	4.99	0.07	146	147	116	117	0.11	0.11	1022	992	39	4.90
Emilia Romagna	8.41	8.38	0.35	4479	4348	100	97	0.13	0.13	776	28991	199	5.20
Piedmont	8.33	8.37	0.45	4157	4318	95	99	0.12	0.12	794	36279	172	4.90
Liguria	7.37	7.57	2.62	1594	1934	103	125	0.12	0.15	823	11085	286	4.40
Trentino alto Adige	6.55	6.14	6.22	697	464	65	43	0.08	0.05	853	8326	79	3.50
Veneto	7.68	7.91	2.98	2167	2724	44	56	0.08	0.10	533	37073	267	4.40
Friuli Venezia Giulia	5.86	6.43	9.81	350	622	29	51	0.08	0.14	367	11343	153	5.00
NORTHERN REGIONS			2.87	30515	32250	110	116	0.13	0.14	823	212395	231	4.71
Marche	6.90	6.40	7.26	989	599	65	39	0.13	0.08	511	7067	162	4.70
Abruzzo	6.17	5.99	2.88	477	399	36	30	0.11	0.09	324	7721	121	5.40
Tuscany	7.05	6.94	1.57	1153	1032	31	28	0.08	0.07	377	17864	162	5.20
Lazio	6.80	7.06	3.79	902	1168	15	20	0.06	0.08	255	15442	341	6.00
Pulia	6.37	5.90	7.36	583	365	14	9	0.08	0.05	179	8052	206	6.30
Campania	6.12	5.86	4.35	457	350	8	6	0.04	0.03	188	3328	424	6.80
Sardinia	4.98	4.68	5.88	145	108	9	7	0.04	0.03	208	4966	68	5.30
Umbria	4.44	4.68	5.36	85	108	10	12	0.04	0.05	257	2516	104	5.30
Sicily	5.71	5.73	0.36	303	309	6	6	0.05	0.05	125	14856	194	5.90
Molise	3.14	3.96	26.28	23	52	8	17	0.04	0.08	204	1735	69	5.60
Calabria	4.58	4.56	0.58	98	95	5	5	0.05	0.05	97	3910	128	7.30
Basilicata	3.33	3.19	4.30	28	24	5	4	0.04	0.04	119	1197	56	6.60
SOUTHERN REGIONS			5.83	5243	4611	16	14	0.07	0.06	228	88654	179	6.02
TOTAL ITALY			4.65	35758	36861	59	61	0.12	0.12	501	301049	200	5.42

