ISSN 2385-2755
DiSSE Working papers
[online]

WORKING PAPERS SERIES
DIPARTIMENTO DI
SCIENZE SOCIALI ED ECONOMICHE

Local mortality estimates during the COVID-19 pandemic in Italy

Augusto Cerqua, Roberta Di Stefano, Marco Letta, Sara Miccoli

N. 14/2020

SAPIENZA - UNIVERSITY OF ROME
P.le Aldo Moro n.5 – 00185 Roma T(+39) 0649910563
CF80209930587 – P.IVA 02133771002

Local mortality estimates
during the COVID-19 pandemic in Italy

Augusto Cerqua[a*], Roberta Di Stefano[b], Marco Letta[a], Sara Miccoli[b]

[a] Department of Social Sciences and Economics, Sapienza University of Rome

[b] Department of Methods and Models for Economics, Territory and Finance, Sapienza University of Rome

This version: July 5, 2020

First version: June 17, 2020

Abstract *Estimates of the real death toll of the COVID-19 pandemic have proven to be problematic in many countries, and Italy is no exception. Mortality estimates at the local level are even more uncertain as they require strict conditions, such as granularity and accuracy of the data at hand, which are rarely met. The 'official' approach adopted by public institutions to estimate the 'excess of mortality' during the pandemic is based on a comparison between observed all-cause mortality data for 2020 with an average of mortality figures in the past years for the same period. In this paper, we show that more sophisticated approaches such as counterfactual and machine learning techniques outperform the official method by improving prediction accuracy by up to 18%, thus providing a more realistic picture of local excess mortality. The predictive gain is larger for small- and medium-sized municipalities. After showing the superiority of data-driven statistical methods, we apply the best-performing algorithms to generate a municipality-level dataset of local excess mortality estimates during the COVID-19 pandemic. This dataset is publicly shared and will be periodically updated as new data become available.*

Keywords: COVID-19, coronavirus, mortality estimates, Italy, municipalities

JEL-Codes: C21, C52, I10, J11

Acknowledgments: We are grateful to Alessio D'Ignazio for his comments and suggestions which greatly improved the quality of this work.

Resource site (in Italian): https://www.stimecomunalicovid19.com

Dataset: Latest version available for download in the "Dataset" window of the resource site.

* Corresponding author.

E-mail addresses: augusto.cerqua@uniroma1.it (A. Cerqua), roberta.distefano@uniroma1.it (R. Di Stefano), marco.letta@uniroma1.it (M. Letta), sara.miccoli@uniroma1.it (S. Miccoli).

1. Introduction

The COVID-19 pandemic is a complex and constantly evolving phenomenon that is affecting the entire world heterogeneously. The disease caused by the spread of this new form of coronavirus rapidly propagated worldwide, affecting countries with different timing and intensity (Ceylan, 2020). At the time of this writing (2nd July 2020), the countries affected by the COVID-19 are 216, the confirmed cases 10,458,422, and the total deaths 511,082 (World Health Organization). Italy was the first country in Europe to be hit by COVID-19 and, to date, it ranks among the countries with the highest number of deaths. In most countries, including Italy, the growing availability of real-time data is driving the decision-making process to face the health, economic, and social emergencies. Monitoring the evolution of the pandemic allows policymakers to make decisions promptly and manage the economic recovery with differentiated place-based policies. In light of this, reliable data about infections and deaths are focal to identify correct policies aimed at loosening lockdown measures. Unfortunately, the issue of underreporting seems to be widespread with official data on coronavirus (Ghislandi et al., 2020). We refer, for example, to the asymptomatic or to the ill people who died at home without being tested for the presence of the virus. Besides, such measurement error is likely to be heterogeneous even within countries, and this inaccuracy may lead to wrong decisions (Leon et al., 2020). A valid alternative for estimating the number of deaths caused by the pandemic (directly or indirectly) consists in considering the number of daily certified all-cause deaths (deaths from any cause, not only related to coronavirus).[1] In particular, using historical data on the number of daily certified deaths to estimate the number of daily deaths in the absence of pandemic, could vastly reduce the uncertainty associated with official data, especially at a disaggregated geographical level.

This paper aims at estimating the excess of mortality in Italian municipalities (local administrative units, LAU) from 21st February (official date of the first coronavirus cluster in Italy),[2] to provide an accurate and geographically detailed measure on the impact of the pandemic. Given the diverse geographical spread of the pandemic, local estimates provide an important tool to mitigate or reinforce security measures where necessary. Conversely, official data on coronavirus reported cases are released only at the provincial level (the number of infected people) or at the regional level (the number of coronavirus deaths). In our research, we compare the official number of all-cause deaths for the period 21st February – 15th May 2020 with an estimate of the number of deaths in the same period in an ordinary situation, i.e., in the counterfactual situation without the pandemic. We then consider the difference between observed mortality in 2020 and our counterfactual predictions as the number of excess deaths, which are very likely to be due, either directly or indirectly, to COVID-19. As the pandemic has affected the entire country, it is not feasible to use the most common counterfactual

[1] The Italian National Institute of Statistics (Istat) released the number of daily certified deaths for the period 1st January – 15th May 2020 for 7,270 Italian municipalities, covering 93.5% of the total Italian population.

[2] However, there is some empirical evidence which suggests that COVID-19 was already present in Northern Italy at the end of January (see Cerqua and Di Stefano, 2020).

approach based on the comparison of treated and non-treated municipalities. For this reason, we make use of the data from the recent past (from 2015 to 2019) on all Italian municipalities to build the counterfactual scenario. This is possible because the outcome and predictive variables are not very volatile over time.

We report the estimates obtained using three estimation approaches: intuitive, counterfactual, and machine learning (ML). The intuitive approach, which is the one adopted by several public institutions, consists in comparing the actual trend of cumulative deaths in 2020 with the number observed in the past for the same municipality (in the previous year or for the average of previous years). We then use a recent counterfactual methodology, the trajectory balancing estimator (see Hazlett and Xu, 2018). It builds for each municipality a 'synthetic' one, as a linear combination of the mortality trends observed in the previous years for the same municipality and municipalities with similar characteristics, before 21st February. Finally, we employ three ML algorithms to estimate excess mortality for all municipalities. ML techniques allow predicting mortality trends by 'training data', i.e., by learning from past experience, and evaluating 'future' model performance on 'testing data' and comparing predicted and actual values. Specifically, we employ the Least Absolute Shrinkage and Selection Operator (LASSO), random forest, and stochastic gradient boosting. Although, as will be explained in Section 3.3, these methods differ in terms of complexity and inner workings, the underlying principle is common to all of them: 'learn' from (a part of) the data to produce a generalizable model able to accurately predict an outcome variable of interest (mortality in our case) in out-of-sample predictive tasks on unseen data. This makes them data-driven methods precisely targeted to ensure the best possible predictive performance given the data at hand.

We show how all these methods outperform, on average, the intuitive approach adopted by Italian institutions, with predictive gains up to 18% and especially sizable increases in smaller municipalities. Thanks to the superiority of data-driven methods, we improve the estimates of local mortality and provide a more reliable 'counterfactual' scenario. Credible counterfactuals are an indispensable tool to evaluate causes, effects, and policy responses rigorously. This evaluation is critical to improve our understanding of such an unprecedented phenomenon (at least in recent times) and becomes impellent as the gradual attenuation of the emergency finally allows us to look back.

It is important to notice that we estimate the gross excess of mortality due to COVID-19, i.e., the number of deaths due directly to COVID-19 infections as well as the deaths due to the collateral effects of lockdown. Such collateral effects have lowered the likelihood of dying from some causes such as road and workplace accidents, pollution-related diseases, or criminal activities, and increased the likelihood of dying for the excessive stress of the public health system (e.g., severe delays in the hospitalization process).

The rest of the paper proceeds as follows: Section 2 reviews the literature; Section 3 presents data and methods, while Section 4 reports the estimates. Section 5 discusses and concludes.

2. Literature review

Past and recent studies have dealt with the effects of the pandemics[3] on mortality trends by estimating the excess of mortality across several different countries. In particular, various research analyzed the excess of deaths due to the influenza pandemic occurred in the period 1918-1920, known as the Spanish flu. Despite the differences between the Spanish flu and the COVID-19 pandemics (living conditions of the populations, medical knowledge and technology), the existence of better data and more sophisticated methodologies nowadays, and the different stage of the two pandemics (one is in history books while the other is still in progress), it appears appropriate to focus the attention on how these studies were carried out.

The effects of the 1918-1920 pandemic on the mortality evolution have been studied both at a global as well as a country level. The majority of these studies applied intuitive methodologies to estimate the mortality excess. Murray et al. (2006), in order to assess the Spanish flu's potential effect on the global population of 2004, estimate the death toll due to the pandemic as difference between the annual death rates of the years 1918-1920 and the average annual death rates for the periods 1915-1917 and 1921-1923. The same methodology was recently used by Barro et al. (2020), who investigate the effects that the Spanish flu had on mortality and the economy in 48 countries and draw comparisons with the effects that the current COVID-19 pandemic may have. In 2009, Ansart et al. realized a study focused on the impact of the Spanish flu on 14 European countries' experience by using a regression model technique. The excess of mortality was calculated by computing the difference between the observed mortality and the expected baseline mortality for the years 1918-1922. This baseline was obtained through a periodic regression model for the period 1906-1922.

Concerning the effects of the Spanish flu on mortality observed in Italy, different studies have been carried out by applying intuitive methodologies. Mortara (1925) compared the monthly deaths verified during 1918-1919 with the average number of monthly deaths of the years 1911-1913 (Mortara, 1925). In the same way, by comparing deaths from influenza-related causes (such as pneumonia and bronchitis) for the same period, he estimated the role covered in the mortality excess by the diseases connected with the Spanish flu. In a study conducted on the mortality trends of the young Italian population during particular periods of crisis, Pinnelli and Mancini (1998) estimated the death excess due to the Spanish flu, by comparing mortality rates by age and causes of death (influenza, pneumonia and bronchitis) in the period 1918-1919 with the mortality rates in 1915-1917 and 1920-1922. To calculate deaths in the absence of a pandemic, they calculated moving averages, excluding the reference year and the two adjacent years. They estimated the excess deaths by applying the inverse mortality rate formula, finding that the mortality evolution for the groups of young people would have been different without the pandemics.

Furthermore, excess mortality has also been estimated to analyze the effect on mortality of

[3] The following influenza pandemics occurred in the XX and XXI centuries: the Spanish flu in 1918-1920, the Asian flu in 1957-1958, the Hong Kong flu in 1968, the Swine flu in 2009 (World Health Organization).

other pandemics in different countries of the world, using both intuitive and more advanced methodologies. To study the age-specific mortality during the Spanish flu and to compare it with the age-specific mortality during other pandemics, Luk et al. (2001) calculated the global excess mortality caused by influenza and pneumonia during the three pandemics of 1918, 1957 (Asian Flu), 1968 (Hong Kong Flu). The excess mortality is calculated by subtracting the mortality rate during the pandemic from the average mortality rates of the five years before the pandemic years. Viboud et al. (2016) realized a study to analyze the mortality related to the pandemic of 1957-1959 in 39 countries of the world. They estimate the excess of mortality as the difference between the observed mortality during the pandemic and the annual mortality time series from 1950 to 1956. Within the European Mortality Monitoring Project (EuroMOMO), aimed at monitoring the mortality excess due to influenza in Europe, Mazick et al. (2010) compared the all-cause deaths observed in 8 European countries during the 2009 A(H1N1) pandemic, known as the Swine Flu, with the deaths occurred in the three previous years in the same weeks. In particular, they model the expected weekly all-cause deaths for all age groups in the absence of pandemic or influenza, exploiting a time series Poisson regression model considering trend and seasonality. The same method was used by Gran et al. (2013) to study the mortality during the first wave of the A(H1N1) pandemic in Norway. A Poisson prediction model was applied by Yang et al. (2012) to estimate the excess mortality associated with A(H1N1) in 2009 and seasonal influenza in 1998–2008 in Hong Kong.

Not surprisingly, given the continuing increase of COVID-19-related deaths, interest in the death toll of this shock has grown worldwide. As soon as official data on deaths have become available, several researchers have begun to undertake studies similar to those conducted to assess the effect of other pandemics, by gauging the COVID-19 impact on mortality through intuitive or more sophisticated approaches.[4] Such an exercise appears useful for generating valuable insights for future pandemic policy. Given the economic, social, and health relevance of this issue, many newspapers showed excess mortality data between and within countries, using intuitive approaches, i.e., by comparing actual deaths with historical data over 2 or 5 years. Aron and Muellbauer (2020) quoted three prominent newspapers (The Financial Times, The Economist, and The New York Times) that showed the excess of mortality at the country-level. They also reported results tracked by EuroMOMO for 24 European states and sub-regions relative to the Z-score, a measure of excess mortality obtained by standardizing data on excess deaths. Several scholars have come up with country-level analyses that use more sophisticated approaches to estimate excess mortality due to COVID-19. For instance, Felix-Cardoso et al. (2020) used deviation from the expected value from homolog periods (DEV), and the remainder after seasonal time series decomposition (RSTS) considering total, age- and gender-specific excess mortality in five countries (England and Wales, France, Italy, Netherlands, and Portugal). Rivera et al. (2020) proposed a semiparametric method and a conventional time-series method and analyzed nine USA states. Pham (2020), likewise for the

[4] This paper provides a comprehensive, but not exhaustive, review of the existing literature on the COVID-19 impact on mortality.

USA situation, dealt with the cumulative number of deaths due to the ongoing COVID-19, based on the five-parameter logistic model. He also discusses a new criterion that enables one to choose the best model in the set of candidates. At a more detailed geographical level, there is research conducted by Hauser et al. (2020) by age group and according to symptom status in Hubei province, China, and Northern Italy.

Concerning the Italian situation, there are two important contributions by public institutions that used the intuitive approach for estimating the impact of the COVID-19 pandemic on total resident population mortality by age and gender. The Italian National Institute of Statistics (Istat) and the National Institute of Health (ISS)[5] presented a report at the provincial level[6] for the period 1st January-30th April 2020, gauging the excess of mortality as differences between observed mortality for all-cause and the average for the same period in 2015-2019. The National Institute of Social Security (INPS) has drawn a report for the first quarter of 2020, using the same approach but weighting mortality for the resident population. The analysis is conducted at the provincial level with a focus on Northern municipalities. A similar approach[7] is used by Del Re and Meridiani (2020) for analyzing the principal Italian cities and by Ghislandi et al. (2020) for the Lombardy region and five provinces strongly hit by the virus. Buonanno et al. (2020), by combining official statistics, retrospective data, and original data (i.e., obituaries and death notices), provided an estimate on Lombardy municipalities. They showed the reported mortality rate attributable to COVID- 19 accounts only for one half of the observed excess mortality rate between March 2020 and March 2019. Two researches used counterfactual analysis. In particular, Ciminelli and Garcia-Mandicó (2020) estimated daily death registry data for a sample of 1,161 Italian municipalities in seven regions (Emilia-Romagna, Liguria, Lombardia, Marche, Piemonte, Toscana, and Veneto) by running a differences-in-differences regression model using data on mortality from 2016 as a control to mortality in 2020. On the contrary, Modi et al. (2020) proposed a comparison of the weekly mortality rate for Italian regions in the first four months of 2020 with a model prediction obtained from historical mortality rates for the same time of the year. They use two different approaches, Conditional Mean with a Gaussian process (CGP) and Synthetic Control Method (SCM). The only study, to our knowledge, which considers a municipality level analysis was conducted by Blangiardo et al. (2020). They provided a measure of the weekly excess mortality by predicting the expected mortality with a Poisson distribution and specifying a Bayesian hierarchical model on the log mortality relative risk. They showed the estimates, for males and females, distinguishing five macro-areas and focusing on the hardest hit Italian municipalities.

---

[5] ISS (National Institute of Health) is the main institution for research, control, and technical-scientific advice on public health in Italy.

[6] The provincial-level analysis is conducted on 7,270 municipalities, representing 93.5% of the total population.

[7] Meridiani and Del Re obtained the excess of mortality by subtracting an average historical model, taking into account the seasonal mortality, i.e., the average of 2015-2019, normalized to the number of deaths observed in the first three weeks of February; Ghislandi et al. compared deaths from 1st of January to 15th of April with the average deaths of the same period for the years 2015-2019.

Although there are many papers using intuitive and counterfactual approaches, only two works, in our knowledge, employ ML techniques to estimate excess mortality. Deprez et al. (2017) investigated two classical models for estimating mortality rates and, by applying a regression tree boosting machine, detected the weaknesses of different mortality models. Levantesi and Pizzorusso (2019) extended the work of Deprez et al. (2017) by investigating the ability of ML to improve the accuracy of some standard stochastic mortality models, using not only decision tree, but also random forest, and gradient boosting. Moreover, they introduced an ML estimator to improve the forecasting quality provided by standard stochastic models. However, ours is the first paper to use ML algorithms to estimate excess mortality during the COVID-19 pandemic.[8]

3. Data and methodology

3.1 Data

On the 18th of June 2020, Istat released data on the daily number of all-cause deaths for the January/1st-May/15th period on 7,270 of the 7,904 Italian municipalities, covering the 93.5% of the Italian total resident population.[9] Besides, Istat released data on the daily number of all-cause deaths for all the Italian municipalities for the years 2015-2019. We exploit such historical data and other variables to estimate excess death during the coronavirus outbreak in Italy.

In particular, for the two data-driven approaches, i.e., the trajectory balancing estimator and the ML algorithms, we feed the models with 15 covariates covering aspects strongly related to deaths such as demographic, health system, economic and contamination (air pollution) variables. This set of variables allows us to estimate the mortality trend for 2020 in the counterfactual situation, i.e., without the outburst of coronavirus, in a more accurate way.

As COVID-19 is thought to be more lethal among men and the elderly (SARS-CoV-2 Surveillance Group, 2020; Dowd et al., 2020; Dudel et al., 2020), we control for the age structure, i.e., the share of men in the population, the share of those aged 65+ (overall as well as only men) and the share of those aged 80+ (overall as well as only men).

Besides, we control for the resident population, the overall number of deaths in the previous year, and the overall number of deaths in the period from the 1st of January to the 20th of

[8] ML algorithms applied to research questions related to COVID-19 are employed by Dandekar and Barbastathis (2020) and Magri and Doan (2020). The former made use of a neural network model to compare the role played by quarantine isolation measures in Wuhan, Italy, South Korea and the United States in controlling the infectious spread. The latter proposed a first-principles ML model to provide quantitative estimates on the infected, deaths rate and R0, i.e., the average number of infections generated by a single infected person within a susceptible population, equal to one on a few countries (the United Kingdom, Italy, Germany, France, Spain, Belgium, the USA, and China).

[9] Due to the creation of Mappano as a new administrative unit in 2017 and to the lack of mortality data for all years, we cannot analyze 8 municipalities: Balmuccia, Borgaro Torinese, Caselle Torinese, Leini, Malvicino, Mappano, Rassa, and Settimo Torinese. Besides, as 2020 is a leap year, we decided to ignore the deaths that occurred on February 29 for reasons of comparability with data from previous years.

February 2020, i.e., the 7 weeks before the coronavirus outbreak in Italy.

We also control for the number of employees as it is likely related to the heterogeneous spread of the contagion (see Ascani et al., 2020), for the share of employment in manufacturing and for PM-10 as a measure of air quality. The latter two variables take into account that the most vulnerable people are those affected by respiratory diseases, conditions associated with high mortality in COVID-19 infection, which are more widespread in industrialized areas.[10] For similar reasons, we also control for the degree of urbanization of the municipality.

As for health care characteristics, we control for a dummy variable equal to 1 if there is a hospital in the municipality and another dummy variable equal to 1 if there is a hospital in at least one of the neighboring municipalities. Lastly, as the lockdown imposed after the coronavirus outbreak has surely decreased the number of deaths due to road accidents, we control for the number of deaths due to road accidents in the previous year. This way, we compare municipalities with similar mortality rates due to road accidents.

3.2 Methodologies

We use three different sets of approaches to estimate excess-mortality: intuitive, counterfactual, and machine learning. As all municipalities are affected by COVID-19, even if heterogeneously, we cannot construct the counterfactual scenario by looking at the number of all-cause deaths in 2020 for non-affected municipalities. As described below, we exploit data on all-cause deaths for the period 2015-2019 to estimate the impact of coronavirus on mortality.

3.2.1 The intuitive approach

The intuitive approach consists in comparing the actual trend of cumulative deaths in 2020 with the trends of cumulative deaths observed in the past for the same municipality (in the previous year or for the average of previous years). In this case, we consider the average of the cumulative number of annual deaths in the period 2015-2019. This approach is easy to interpret, even if it does not allow to take into account unobserved factors such as flu epidemics or climatic conditions, which can vary over time. The intuitive approach has been recently used in several official reports, as mentioned in Section 2.[11]

---

[10] Employment data come from the Statistical Register of Active Enterprises (ASIA) archive, which covers the universe of firms and employees of industry and services. PM 10 data are taken from the European Environment Agency
http://aidef.apps.eea.europa.eu/?source=%7B%22query%22%3A%7B%22match_all%22%3A%7B%7D%7D%2C%22display_type%22%3A%22tabular%22%7D

PM-10 data is from 573 monitoring stations distributed across the Italian territory. We employ the kriging spatial interpolation to impute the PM10 average yearly value for each municipality.

[11] Specifically, it was adopted by:

-Istat-ISS report at the Italian provincial level
(https://www.epicentro.iss.it/coronavirus/pdf/Rapp_Istat_Iss_3Giugno.pdf);

3.2.2 The counterfactual approach

Counterfactual approaches are usually adopted to estimate the impact of a specific policy change (defined 'treatment') on an outcome of interest. In this paper, we consider as treatment the beginning of the coronavirus diffusion in Italy, i.e., the 21st of February, and as the outcome of interest the number of cumulative deaths. As a counterfactual evaluation method, we use trajectory balancing (TB), a recent estimator proposed by Hazlett and Xu (2018). Similar to the synthetic control method (see Abadie et al., 2010), TB builds a 'synthetic unit' for each municipality as a weighted average of units not affected by the intervention (defined 'control units') whose pre-treatment characteristics closely match that unit affected by the treatment (defined 'treated unit'). The 'synthetic unit' represents what would have happened to the treated unit in the absence of coronavirus spread, i.e., the counterfactual situation. Then, the difference between the outcome of interest of the treated unit and the outcome of interest of the 'synthetic unit' represents the treatment effect, in our case, the estimated excess mortality due (directly or indirectly) to coronavirus. This method enables us also to control for unobserved factors that can vary over time (see Abadie et al. 2015 for details).

In more detail, we build the synthetic unit for each municipality considering as control group the past values of all Italian municipalities. We take into account the mortality trends in the 2015-2019 period for the same municipality as well as for municipalities with similar demographic, health, and economic characteristics in the pre-treatment period, i.e., between 1st January and 20th February.[12] So, each municipality in each year represents a different unit, and the algorithm assigns a weight (between 0 and 1, whose weights sum have to be equal to 1) to each unit, considering the similarities to the municipality considered in the pre-treatment period. Then, the excess of mortality is equal to the difference between the number of cumulative deaths in the municipality observed from 21st of February to 15th of May 2020 and the weighted average of municipalities in the control group with positive weights, observed from 21st of February to 15th of May. Therefore, differently from the intuitive approach, we do not consider only the 'same' municipality to estimate the counterfactual scenario. On the contrary, the 'synthetic unit' is made up of municipalities (also the same

-INPS report at the Italian provincial level with a focus on Northern Italian municipalities (https://www.inps.it/docallegatiNP/Mig/Dati_analisi_bilanci/Nota_CGSA_mortal_Covid19_def.pdf)

-SISMG to monitor the situation of daily deaths for the elderly population (over 65 years) for Italian cities with more than 250,000 inhabitants (https://repo.epiprev.it/index.php/download/andamento-della-mortalita-giornaliera-sismg-nelle-citta-italiane-in-relazione-allepidemia-di-covid-19-report-1-febbraio-2-maggio-2020-settimo-rapporto/?wpdmdl=1626&refresh=5ee8e7ed5223a1592322029)

[12] To construct the synthetic unit we limit the set of potential control units following two criteria: the geographical area (the North-East, the North-West, the Centre, the South, and the Islands) and the population size (we split municipalities in four segments: 0-1,999 inhabitants, 2,000-4,999 inhabitants, 5,000-19,999 inhabitants, and 20.000+ inhabitants). As suggested in Abadie et al. (2015), by restricting the donor pool to municipalities with similar characteristics, we reduce the risk of interpolation bias. The municipalities in the geographical area considered have similar local economic structures and sector specialization, factors that can act as a vehicle of disease transmission (for details, see Ascani et al., 2020). In other words, we consider the municipalities in which the virus could spread equally. Moreover, the same geographical area means a similar impact of seasonal risk factors (climatic conditions and flu epidemics).

municipality is included) very similar to the treated one concerning the mortality trend and the other characteristics described in the Data section.

3.2.3 The ML approach

In the flourishing literature on the use of ML for policy purposes, one is typically concerned about two trade-offs: the one between bias and variance and the one between accuracy and interpretability. While the first is a common issue of ML techniques in every science domain, the latter is distinctive of fields in which ML is employed in the service of public policies that also require to take into consideration communication and accountability aspects. More complex, 'black-box' methods tend to be both more accurate but less, if at all, interpretable. So, the choice of the appropriate technique can often fall on simpler algorithms to get more intuitive, and hence explainable, outputs, at the expense of a loss in predictive accuracy. In our case, however, we are not interested in producing a transparent predictive model which clearly explains how the algorithm relates the features to the output. We just want to produce the most accurate estimates as possible. The aim, somewhat unusual in the ML literature, is to use these techniques to generate a 'counterfactual' scenario with predictions of what mortality figures would have been under 'ordinary' conditions, i.e., if the 'treatment' represented by COVID-19 would not have happened. This different framework is an advantage for our purpose because the trade-off between accuracy and interpretability does not impose a constraint on the selection of the techniques to employ. Thus, we opt for a mix of methods: a simpler algorithm and two black-box techniques and show that, for this particular task, the simpler method performs better than the most complex ones, at least at the aggregate level.

Specifically, we adopt the following ML algorithms: least absolute shrinkage and selection operator (LASSO); and two methods based on regression trees, random forest and stochastic gradient boosting. These algorithms are characterized by growing degrees of complexity and flexibility.[13] LASSO is a relatively simple technique that assumes an underlying linear relationship between the outcome and the predictors. In LASSO, the model is penalized for the sum of the absolute values of the weights. The implication of this regularization is that, depending on the value of the hyperparameter λ, LASSO forces the coefficients of uncorrelated or weakly correlated predictors exactly to zero, thus performing variable selection. This makes LASSO less flexible but more interpretable than standard OLS, as it produces a sparse model in which the outcome is related only to a smaller subset of the predictors. By contrast, random forest and boosting are fully non-linear methods, based on the aggregation of many decision trees.[14] Random forest build several different decision trees based on bootstrapped training samples and use at each split of the trees only a random subset of the predictors as split candidates, thus decorrelating the trees from one another. The key difference with boosting is that while random forest grows trees in parallel, boosting grows

[13] Here we only provide an overview of the main differences between the three methodologies. For an exhaustive description and more technical details on each of these methods, please refer to Hastie et al. (2009).

[14] Although boosting can be applied to other methods than the decision tree, the one based on decision trees is by far one of the most popular versions.

them sequentially. Similarly to random forest, boosting is based on the aggregation and growth of many decision trees. But unlike random forest, boosting does not involve bootstrap sampling, as each tree is based on the 'residual' of previously grown trees, i.e., each tree is fit on a modified version of the original dataset (Hastie et al., 2009). To make the results more comparable across the three methods and capture potential interactions between variables in all the selected models, we also include, for LASSO, all the pairwise interactions between the predictors as additional features.[15]

In the ML literature, the well-established routine is to randomly divide the sample in a training set, in which the model is built and tuned, and a testing set, in which its predictive power is tested through an evaluation of its out-of-sample predictive accuracy. In order to solve the other trade-off we mentioned above, the bias-variance trade-off, cross-validation on the training sample can be employed to select the best-performing values of key tuning parameters that regulate the complexity or flexibility of the algorithms and reduce the risk of overfitting.

Following this spirit, we proceed as follows: i) we split the 2015-2018 pooled dataset in a training sample, composed by 80% of the municipalities, and a testing sample, which consists of the remaining 20%; ii) we train and tune the three algorithms on the training sample, on which we perform 10-fold cross-validation to select the best-performing tuning hyperparameters of each algorithm[16]; iii) we test how well the algorithms perform in predicting observed mortality on unseen data, i.e., on the 2015-18 testing sample; iv) we test model performance on the entire 2019 sample, and show that algorithm performance is stable over time and that all the ML methods perform, on average, better than the commonly adopted intuitive method; v) we repeat this routine on the pooled 2015-2019 data, to train the models on most data as possible so as to maximize the accuracy gains; vi) we use the models built on the 2015-2019 dataset to predict, for the 2020 sample, estimates of local mortality in a 'no-COVID' scenario; vii) we derive municipality-level excess deaths for all the municipalities by subtracting the ML counterfactual estimates from the observed 2020 mortality data released by Istat.

[15] Unlike LASSO, random forest and boosting take into account by default all the possible non-linearities and interactions between the features.

[16] The hyperparameters we select via 10-fold cross-validation are the following: for LASSO, the parameter λ which controls the shrinkage penalty; for the random forest, the parameter m, i.e., the number of features randomly sampled as candidates at each split (for the number of trees to grow, instead, we use the default value of 1000); for boosting, the shrinkage parameter representing the learning rate, the number of trees to fit and the minimum number of observations in the terminal nodes of the tree. Cross-validation is used running several different models with several candidate values (or combinations of values, in the case of boosting) for all these parameters.

Table 1 – Descriptive statistics

	Year					
Variables	2015	2016	2017	2018	2019	2020
Number of deaths from 01/01 to 20/02 (per 10,000 inhabitants)	22.38	18.99	23.42	21.53	21.42	19.37
Number of deaths in the previous year (per 10,000 inhabitants)	76.98	84.45	80.65	84.77	82.44	82.92
Population	7815.32	7798.89	7789.81	7776.53	7761.00	7747.41
Share of those aged 65+ (%)	23.82	24.17	24.48	24.76	25.10	25.10
Share of those aged 80+ (%)	7.74	7.86	8.00	8.11	8.31	8.31
Share of men (%)	49.37	49.41	49.47	49.55	49.59	49.59
Share of men aged 65+ (%)	10.50	10.71	10.91	11.08	11.29	11.29
Share of men aged 80+ (%)	2.79	2.86	2.94	3.01	3.12	3.12
Number of employees	2118.02	2131.63	2183.24	2232.40	2232.40	2232.40
Share of employment in manufacturing (%)	24.94	24.86	24.87	24.52	24.52	24.52
PM 10	28.30	28.30	25.44	26.88	24.73	24.73
Hospital in the municipality (%)	7.96	7.89	7.89	7.56	7.48	7.48
Hospital in neighboring municipalities (%)	44.35	44.02	44.02	42.36	42.14	42.14
Number of deaths due to road accidents (per 10,000 inhabitants)	0.43	0.44	0.42	0.43	0.42	0.42
Number of deaths from 21/02 to 15/05 (per 10,000 inhabitants)	31.13	28.97	29.09	29.87	29.85	40.34

Notes: In case of no data available for 2019 and/or 2020, we use the latest year available. We also control for the degree of urbanization which is a constant across years (257 municipalities classified as large urban areas, 2,089 classified as small urban areas and 4,916 classified as rural areas). For comparability, we show the descriptive statistics of the 7,262 municipalities for which all-cause deaths data are available in 2020.

Although our dataset is a panel of Italian municipalities, we ignore the longitudinal component of the sample and consider the data as pooled. In fact, we treat each municipality-year pair as if it was a single observation. We do not deem this as a problematic aspect as we focus on a relatively short time period, whereas such a choice could entail issues if some of the variables we employ (either the outcome or the predictors) would exhibit drastic changes over time within the same municipality, and this is unlikely to happen in a 4-year time span. In any case, the descriptive statistics reported in Table 1 show that this is a minor concern in our case as the year-by-year variation in mortality data and key predictors is rather low. On top of this, please note that future model performance will, in any case, be evaluated on the same sample, i.e., Italian municipalities will stay the same. In other words, we do not have a sample but the full population of Italian municipalities.[17]

Importantly, we apply our random splitting on *municipalities*, not on *municipality-year* pairs. Going for the latter would make the same municipality appear in the training set in one year (e.g., 2016) but in the testing set in another year (say, 2018). To the extent that the predictors do not change or vary slowly over time, this would produce downward-biased estimates of the mean squared error (MSE), because the testing data would not truly be 'unseen' data, but data very similar to their counterparts for the corresponding municipality in the years that appear in the training sample.[18] By splitting on municipalities, instead, we are sure that the training and testing samples do not share in common any municipality, and the same municipality can only appear either in the training or in the testing set for the entire timespan.

4. Estimates

4.1 Predictive power of all methods employed

We begin the empirical analysis by examining the forecasting performance of all methods employed. Given the presence of approaches different from ML, we have opted for evaluating the performance of the various approaches in the estimation of the number of deaths per 10,000 inhabitants in an 'ordinary year'. We select 2019 as the 'ordinary year' and use data from 2015 to 2018 to test the predictive power of all methods. For forecast evaluation, we employ the Mean Squared Error (MSE), i.e., the mean of the squared value of the prediction errors, as the main model selection approach. However, as MSE is very sensitive to outliers, we also report the Mean Absolute Error (MAE), i.e., the mean of the absolute value of the errors. These two measures are our metrics to conduct a comparative analysis of predictive performances. Table 2 reports the results for all the sets of methods at the dates of March 31st and April 30th.

The first insight from Panel A of Table 2 is that counterfactual and especially ML techniques perform better than the intuitive approach. On average, MSE is reduced by up to 18% than if using the mainstream approach adopted by several Italian and international institutions. This is the key result of the paper. A closer look reveals that LASSO is the best-performing algorithm. Random forest and Boosting fare worse, but not by much, and still lead to an improvement in precision over both the trajectory balancing and the intuitive approaches. This ranking remains the same for both dates considered as well as when looking at the MAE. While the superior performance of the linear method is somewhat unusual compared to other works in the ML literature, it can be explained by the nature of the predictive problem and the limited number of observations and predictors. In such circumstances, it can be expected that

[17] This is true for the 2015-2019 dataset but not for 2020, for which, as explained above, we have data for 7,270 municipalities. However, considering that the model is built, trained, and tested on the 2015-2019 sample, this does not apply to the point we are making here.

[18] Tests run on the training and testing samples created via a random split on the municipality-year pairs confirmed this suspicion: the observed testing MSEs were slightly lower than those of our benchmark approach.

simpler algorithms can match the performance of more complex methods.

Table 2 - A comparison of predictive accuracy across the different methods

Panel A – Performance on all municipalities

Method	MSE – 31 March 2019	MAE – 31 March 2019	MSE – 30 April 2019	MAE – 30 April 2019
Intuitive (historical average)	216.64	8.70	340.80	11.27
Intuitive (past year)	352.04	10.38	585.56	13.98
Trajectory balancing	191.91	8.46	310.18	11.03
LASSO	178.01	8.09	280.03	10.50
Random forest	180.55	8.16	287.67	10.61
Boosting	179.63	8.21	282.83	10.71

Panel B – Performance by population size

		< 2,000 inhabitants		
Intuitive (historical average)	450.63	14.32	705.24	18.58
Intuitive (past year)	733.68	16.80	1222.33	23.00
Trajectory balancing	396.93	13.86	635.83	17.49
LASSO	369.63	13.29	575.41	16.76
Random forest	375.50	13.48	593.51	16.98
Boosting	371.46	13.35	577.79	16.71
		Between 2,000 and 5,000 inhabitants		
Intuitive (historical average)	56.91	5.89	93.94	7.56
Intuitive (past year)	91.13	7.38	148.62	9.48
Trajectory balancing	52.29	5.72	91.52	7.49
LASSO	46.40	5.36	79.97	5.50
Random forest	46.60	5.37	79.95	5.52
Boosting	47.51	5.44	82.52	5.56
		Between 5,000 and 20,000 inhabitants		
Intuitive (historical average)	18.90	3.37	31.37	4.36
Intuitive (past year)	29.88	4.19	48.98	5.44
Trajectory balancing	18.68	3.39	32.58	4.47
LASSO	16.86	3.25	29.44	3.35
Random forest	16.69	3.22	28.89	3.33
Boosting	18.61	3.46	32.85	3.45
		>= 20,000 inhabitants		
Intuitive (historical average)	4.76	1.70	8.56	2.25
Intuitive (past year)	6.57	1.95	12.05	2.69
Trajectory balancing	4.78	1.73	9.81	2.47
LASSO	4.81	1.77	9.87	1.92
Random forest	4.12	1.61	7.75	1.68

Boosting	6.51	2.09	14.16	2.16

Notes: There are 3,271 municipalities with less than 2,000 inhabitants, 1,817 municipalities with inhabitants between 2,000 and 5,000, 1,682 municipalities with inhabitants between 5,000 and 20,000 and 492 inhabitants with more than 20,000 inhabitants. MSE is from the testing sample for ML techniques and from the whole sample for the intuitive and counterfactual approaches.

Panel B of Table 2 reports the performance of all estimators by population size. First and foremost, the magnitude of the prediction error is inversely proportional to the municipality size. While such sharp inter-class heterogeneity may seem striking, this is actually not surprising when one takes into account that the dependent variable is defined as the number of deaths per 10,000 inhabitants and that the variability of growth rates of any variable in small municipalities is substantially higher. Second, concerning model performance, there is also some heterogeneity in performance, depending on population size.[19] For example, random forest is the best-performing algorithm for municipalities with a population above 5,000 inhabitants, while LASSO performs somewhat poorly for large municipalities. The intuitive approach does not perform well for small and medium-sized municipalities, but much better for large municipalities. Importantly, our key insight is confirmed in all subsamples: in no case, the intuitive method results as the best-performing method.

In the following section, we show the excess death mortality estimates for 2020 as computed by the best-performing algorithm for each of the four population cut-offs, i.e., LASSO for municipalities with less than 5,000 inhabitants and random forest for municipalities with 5,000 or more inhabitants.

4.2 Predicting excess deaths during the COVID-19 pandemic

The excess mortality estimates from all-causes deaths are not uniform throughout Italy, as can be seen in Figure 1. Significant differences emerge across and within geographical areas. The excess mortality estimates obtained via LASSO and random forest for the period from February 21st to May 15th are particularly high in many Northern municipalities. In this area, 3.61% of the municipalities record a percentage of excess mortality higher than 300% in comparison with the counterfactual scenario. In particular, in the first weeks after the outbreak, from February 21st to March 15th (Panel A of Figure 1), the excess mortality is over 300% in many municipalities of the Bergamo province (Lombardy). On March 15th, this province was the most affected byCOVID-19, featuring the highest number of infections recorded in Italy. At the end of March (Panel B of Figure 1), just over a month after the first outbreak and after three weeks of national lockdown, municipalities with excess mortality over 100% increase in Northern Italy. Municipalities with mortality excess higher than 100% are concentrated in the Lombardy region and the western part of the Emilia Romagna region. Many of these territories are located in the provinces that record the highest number of infections up to that date, namely the provinces of Milan, Bergamo, Brescia, Cremona, Lodi (Lombardy), and the provinces of Piacenza, Parma, Reggio Emilia (Emilia Romagna). On April 30th (Panel D of Figure 1), many municipalities of these provinces still record an estimated

19

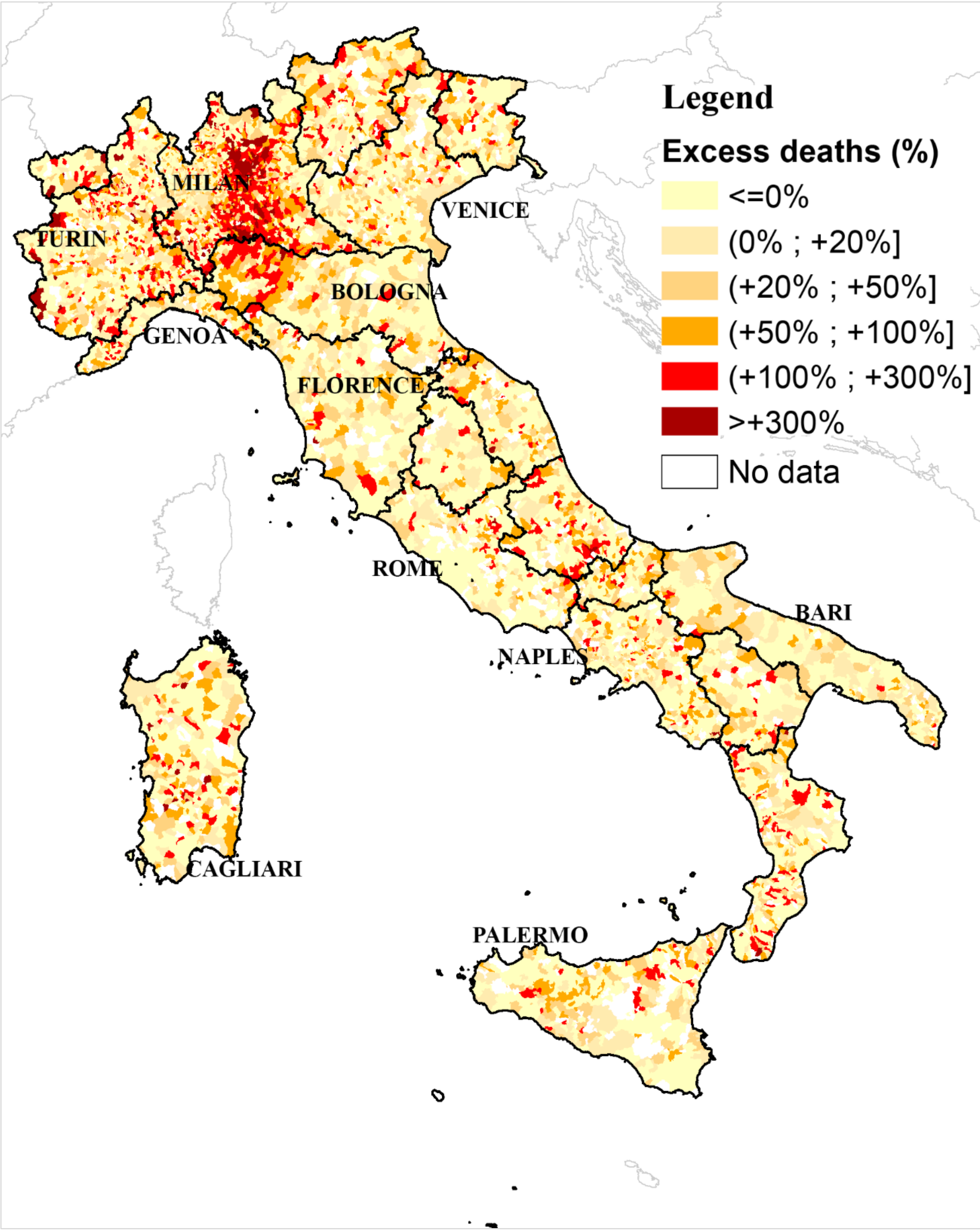
excess of deaths higher than 100%, and in some cases, over 300%, compared to the counterfactual scenario. On May 15th (Panel E of Figure 1) some of the latter municipalities show a lower estimated excess mortality in comparison to the snapshot on 30th of April. Some municipalities in Bergamo and Cremona provinces record values below the 300% threshold, while other municipalities, principally located in Milan and Reggio Emilia, record values below 100%. Clusters of municipalities with estimated excess mortality above 100% are observed in other Northern regions affected by the spread of the virus, such as Piedmont and Liguria. Both on the 15th and 30th of April (Panels C and D of Figure 1), several municipalities in the provinces of Cuneo, Alessandria (Piedmont) and Imperia (Liguria) constitute clusters with excess mortality higher than 100%. Similar clusters are in the provinces of Trento (Trentino-Alto Adige) and Pesaro-Urbino (Marche). On May 15th (Panel E of Figure 1), a decrease in the excess of deaths below 100% seems to occur for some municipalities in Cuneo and Imperia provinces. In the other geographical areas, the estimated excess mortality via LASSO is much lower. From February 21st to May 15th, the majority of municipalities record an estimated excess mortality under 50%, with few municipalities with an excess of deaths between 100% and 300%. On April 30th and May 15th (Panel D and E of Figure 1), in Central and Southern Italy, municipalities with higher excess mortality deaths seem to be slightly less widespread compared to the snapshots on March 30th and April 15th. Figure A.1 and Table A.1 in the Appendix clearly show that the excess death estimates in the Centre-South of Italy can be considered in line with what one would expect to happen in an 'ordinary year' (2019 in our case), while the observed trend in the North of Italy is extraordinarily abnormal. The consistency of the estimates with the geographical evolution of COVID-19 cases confirms that the estimated mortality excess from all-cause deaths is connected, both directly and indirectly, to the COVID-19 pandemic. A note of caution pertains to small municipalities where even sporadic deaths could determine large variations in percentage terms.

The estimated excess mortality for some municipalities between 21st of February and 30th of April is lower than the value estimated for the period 21st of February-15th of April. The decrease in the estimates of excess mortality in comparison to the estimates of the first periods emerges also by looking at the entire period from 21st of February to 15th of May. As pointed out by Istat, the worst-affected provinces in the first phase of the crisis start to experience a decrease of deaths in the second half of April. This may, therefore, explain the decrease in the excess mortality for those municipalities which had been more severely hit by the infection during the period of maximum spread of the virus. This decrease can be connected with the reduction of the most vulnerable population and also with the improved diagnostic capacity and a minor pressure on the national health system (Istat 2020). However, as already pointed out, the estimated excess mortality remains particularly high in many municipalities in northern Italy and on 15th of May clusters of particularly high excess mortality seems even more defined.

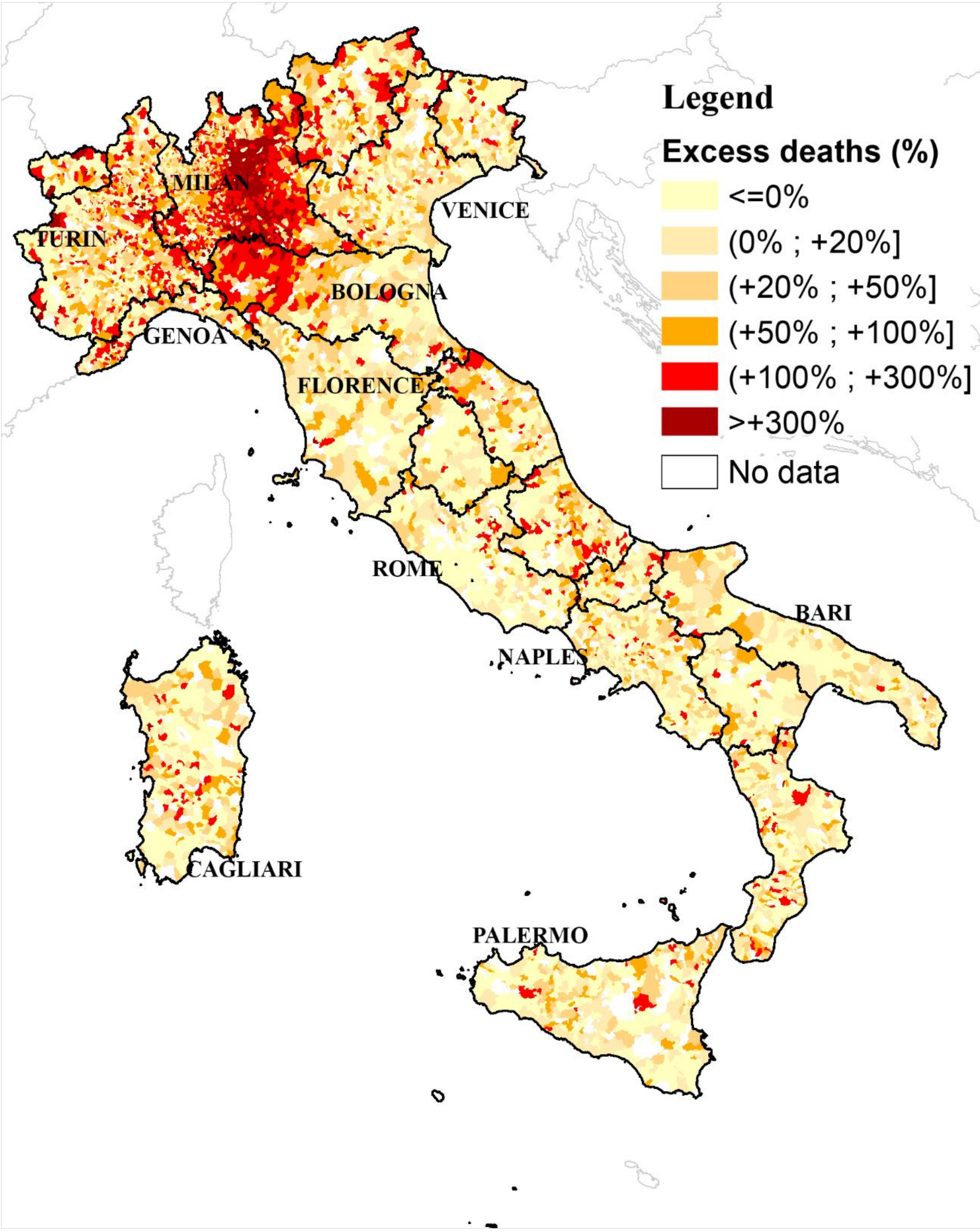
By aggregating the municipality excess mortality estimates at the country level, we find an increase of 41,841 deaths compared to the counterfactual scenario for the period from 21 February to 15 May 2020. This estimate is in line with the figures obtained by INPS and Istat by employing the intuitive approach and it suggests that the 'official' number of deaths directly due to COVID-19 (31,610) might be severely underestimated.

Figure 1 - Percentage of municipal excess deaths detected from 21.02.2020 to 15.05.2020 with respect to the counterfactual scenario estimated via ML techniques

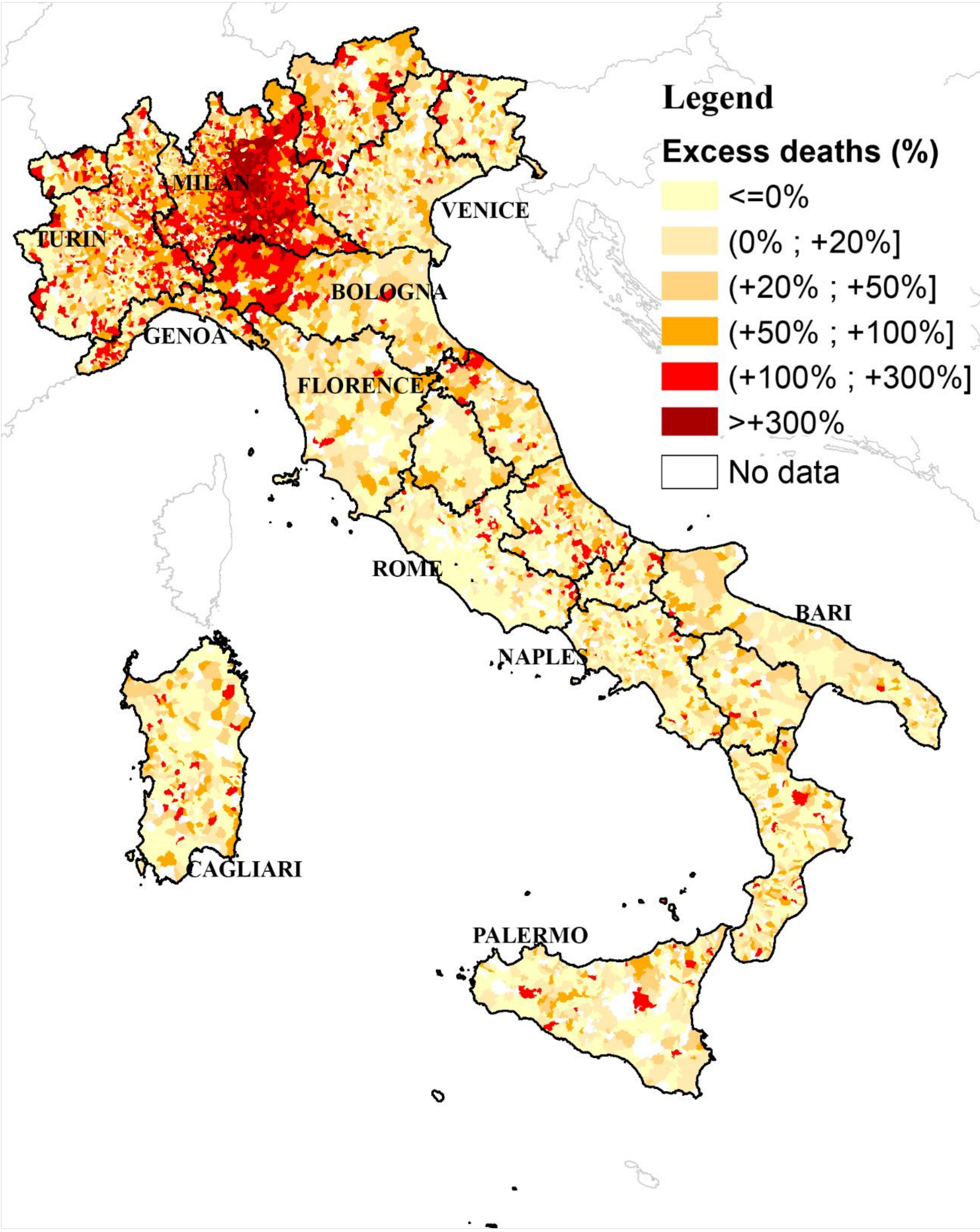
Panel A: from 21.02.2020 to 15.03.2020



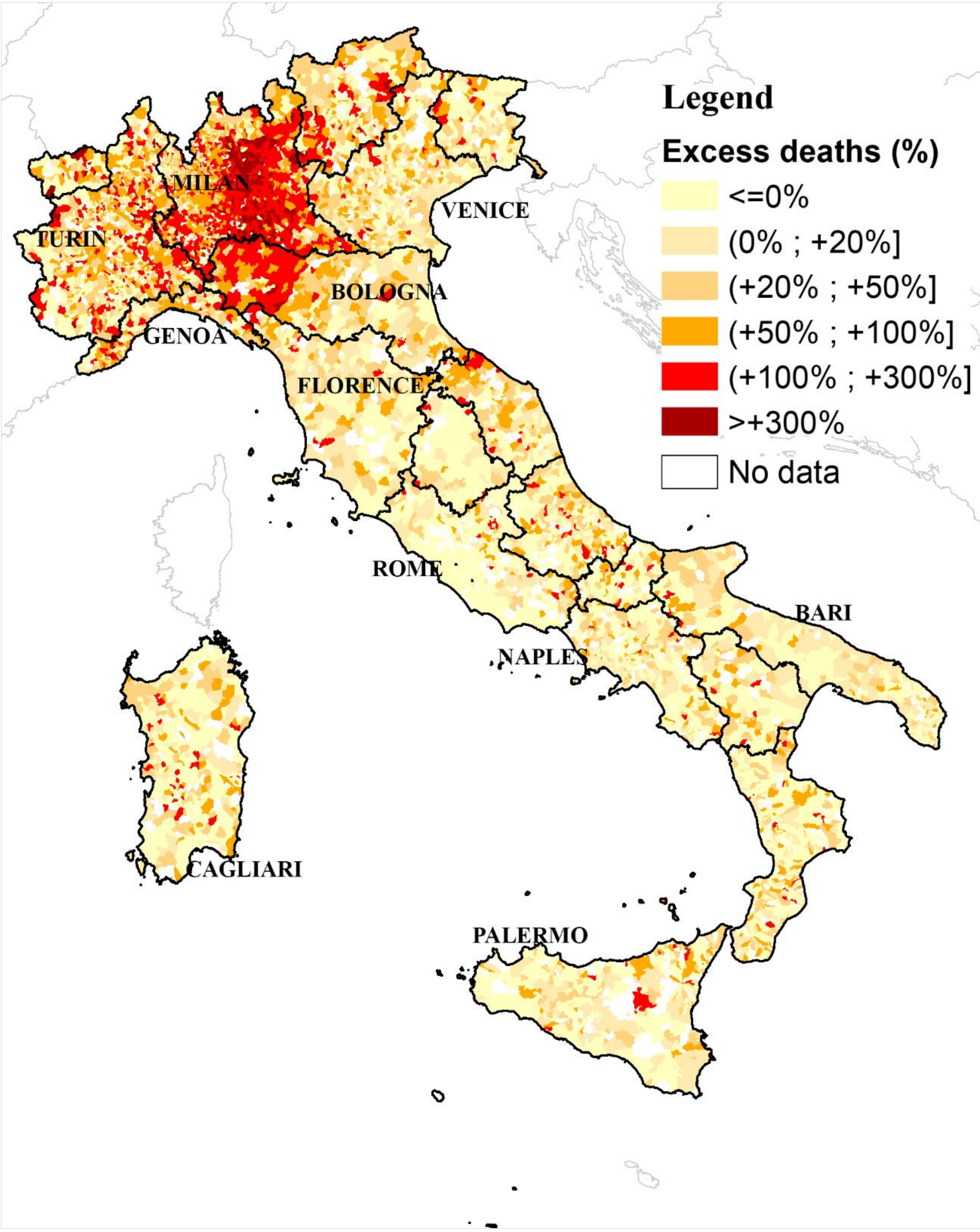
Panel B: from 21.02.2020 to 31.03.2020



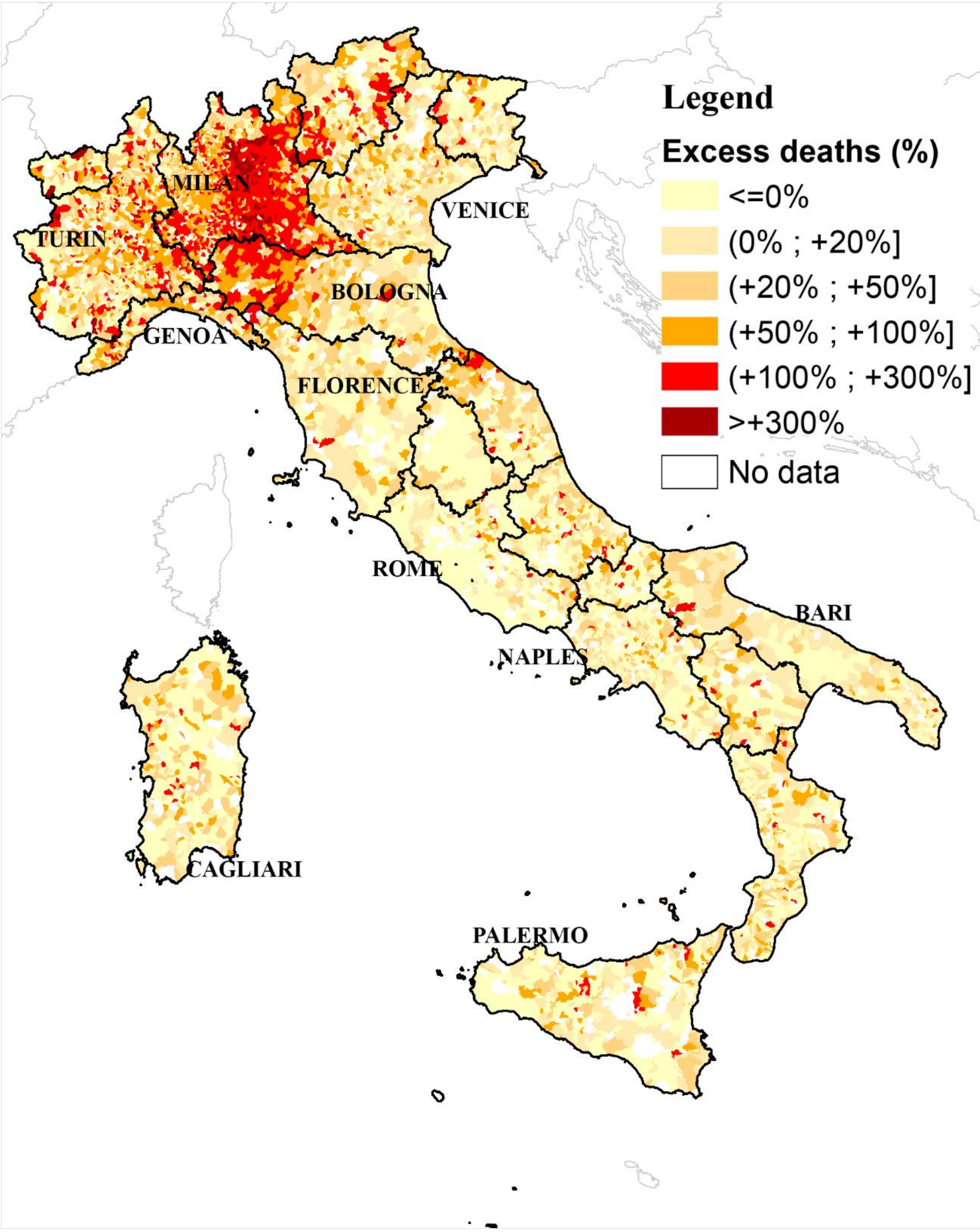
Panel C: from 21.02.2020 to 15.04.2020



Panel D: from 21.02.2020 to 30.04.2020



Panel E: from 21.02.2020 to 15.05.2020



5. Conclusions

At the time of this writing, the spread of the virus in Italy has either slowed down or stopped in all regions, thanks to an increased capacity to contain the virus and to the efficacy of social distancing policies. In any case, the severe COVID-19 death toll will leave deep scars in hundreds of municipalities especially in Northern Italy. As the pandemic slows down and the emergency attenuates, it thus becomes critical to look back and try to quantify human losses more precisely at the local level and identify municipalities that paid the highest price.

In this work, we propose a more sophisticated approach to produce local estimates of excess mortality during the COVID-19 pandemic in Italy. Specifically, we show that counterfactual and ML methods outperform the mainstream intuitive approach adopted by Italian institutions, with particularly sizable predictive gains at the more granular level of small- and medium-sized municipalities. After showing that these methodologies improve performance by up to 18%, we build a municipality-level dataset of the 2020 excess death mortality figures. This dataset, which is shared jointly with the paper, is intended to be available to the general public as well as to researchers interested in investigating local determinants and territorial factors that may have contributed to the rapid and heterogeneous spread of the pandemic across Italy, as well as to the evaluation of policy responses at the local level. We hope our methodological contribution will lead to further refinements of the current approaches targeted at estimating mortality during the pandemic and, in turn, to a broader understanding of the spread of the virus in Italy and the efficacy of the policies adopted to contain its impacts.

Finally, we emphasize that, in principle, other counterfactual or ML methods may perform even better than the ones we apply, thus leading to further increases in statistical accuracy compared to the intuitive approach. Besides, the methodological framework to estimate local mortality that we propose here could be extended to other countries and, possibly, to the entire European Union. In sum, there is indeed room for improvement. We defer these refinements to future research.

References

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American statistical Association*, 105(490), 493-505. https://doi.org/10.1198/jasa.2009.ap08746

Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2), 495-510. https://doi.org/10.1111/ajps.12116

Ansart, S., Pelat, C., Boelle, P. Y., Carrat, F., Flahault, A., & Valleron, A. J. (2009). Mortality burden of the 1918–1919 influenza pandemic in Europe. I*nfluenza and other respiratory viruses*, 3(3), 99-106. https://doi.org/10.1111/j.1750-2659.2009.00080.x

Aron, J., & Muellbauer, J. (2020). Measuring excess mortality: the case of England during the Covid-19 Pandemic. *INET Oxford Working Paper* 2020-11.

Ascani, A., Faggian, A., & Montresor, S. (2020). The geography of COVID-19 and the structure of local economies: The case of Italy. *GSSI Discussion Paper Series in Regional Science & Economic Geography* 2020-01.

Barro, R. J., Ursua, J. F., & Weng, J. (2020). The coronavirus and the great influenza epidemic. Lessons from the "Spanish Flu" for the coronavirus's potential effects on mortality and economic activity. *NBER Working Paper Series*, 26866. DOI: 10.3386/w26866

Blangiardo, M., Cameletti, M., Pirani, M., Corsetti, G., Battaglini, M. & Baio, G. (2020). Estimating weekly excess mortality at sub-national level in Italy during the COVID-19 pandemic. medRxiv. doi: https://doi.org/10.1101/2020.06.08.20125211

Buonanno, P., Galletta, S., & Puca, M. (2020). Estimating the severity of covid-19: evidence from the italian epicenter. *Center for Law & Economics Working Paper Series*, 3/2020. https://doi.org/10.3929/ethz-b-000414590

Cerqua, A., & Di Stefano, R. (2020). When did Coronavirus arrive in Europe? https://doi.org/10.13140/RG.2.2.14425.26723/4

Ceylan, Z. (2020). Estimation of COVID-19 prevalence in Italy, Spain, and France. *Science of The Total Environment*, 729, 138817. https://doi.org/10.1016/j.scitotenv.2020.138817

Ciminelli, G., & Garcia-Mandicó, S. (2020). COVID-19 in Italy: An Analysis of Death Registry Data. *VOXEU, Centre for Economic Policy Research*, 22.

Del Re, D., & Meridiani, P. (2020). Monitoring the Covid-19 epidemics in Italy from mortality data. *medRxiv*. doi: https://doi.org/10.1101/2020.05.07.20092775

Deprez, P., Shevchenko, P. V., & Wüthrich, M. V. (2017). Machine learning techniques for mortality modeling. *European Actuarial Journal*, 7(2), 337-352. https://doi.org/10.1007/s13385-017-0152-4

Dowd, J. B., Andriano, L., Brazel, D. M., Rotondi, V., Block, P., Ding, X., ... & Mills, M. C. (2020). Demographic science aids in understanding the spread and fatality rates of COVID-19. *Proceedings of the National Academy of Sciences of the United States of America*, 117(18), 9696-9698. https://doi.org/10.1073/pnas.2004911117

Dudel, C., Riffe, T., Acosta, E., van Raalte, A. A., & Myrskyla, M. (2020). Monitoring trends and differences in COVID-19 case fatality rates using decomposition methods: Contributions of age structure and age-specific fatality. *MPIDR Working Paper 2020-020*. DOI:10.4054/MPIDR-WP-2020-020

Felix-Cardoso, J., Vasconcelos, H., Rodrigues, P., & Cruz-Correia, R. (2020). Excess mortality during COVID-19 in five European countries and a critique of mortality analysis data. *medRxiv*, 2020.04.28.20083147. https://doi.org/10.1101/2020.04.28.20083147

Ghislandi, S., Muttarak, R., Sauerberg, M., & Scotti, B. (2020). News from the front: Estimation of excess mortality and life expectancy in the major epicenters of the COVID-19 pandemic in Italy. *medRxiv*. DOI:10.1101/2020.04.29.20084335.

Gran, J. M., Kacelnik, O., Grjibovski, A. M., Aavitsland, P., & Iversen, B. G. (2013). Counting pandemic deaths: comparing reported numbers of deaths from influenza A (H 1 N 1) pdm09 with estimated excess mortality. *Influenza and other respiratory viruses*, 7(6), 1370-1379. https://doi.org/10.1111/irv.12125

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

Hauser, A., Counotte, M. J., Margossian, C. C., Konstantinoudis, G., Low, N., Althaus, C. L., & Riou, J. (2020). Estimation of SARS-CoV-2 mortality during the early stages of an epidemic: a modelling study in Hubei, China and northern Italy. *medRxiv*. https://doi.org/10.1101/2020.03.04.20031104

Hazlett, C. & Y. Xu (2018): Trajectory balancing: A general reweighting approach to causal inference with time-series cross-sectional data. Available at SSRN 3214231.

Leon, D. A., Shkolnikov, V. M., Smeeth, L., Magnus, P., Pechholdová, M., & Jarvis, C. I. (2020). COVID-19: a need for real-time monitoring of weekly excess deaths. *The Lancet*, 395(10234), e81. https://doi.org/10.1016/S0140-6736(20)30933-8

Levantesi, S., & Pizzorusso, V. (2019). Application of machine learning to mortality modeling and forecasting. *Risks*, 7(1), 26. https://doi.org/10.3390/risks7010026

Luk, J., Gross, P., & Thompson, W. W. (2001). Observations on mortality during the 1918 influenza pandemic. *Clinical Infectious Diseases*, 33(8), 1375-1378. https://doi.org/10.1086/322662

Mazick, A., Gergonne, B., Wuillaume, F., Danis, K., Vantarakis, A., Uphoff, H., ... & Mølbak, K. (2010). Higher all-cause mortality in children during autumn 2009 compared with the three previous years: pooled results from eight European countries. *Eurosurveillance*, 15(5), 19480. https://doi.org/10.2807/ese.15.05.19480-en

Modi, C., Boehm, V., Ferraro, S., Stein, G., & Seljak, U. (2020). How deadly is COVID-19? A rigorous analysis of excess mortality and age-dependent fatality rates in Italy. *medRxiv*. https://doi.org/10.1101/2020.04.15.20067074

Mortara, G. (1925). *La salute pubblica in Italia durante e dopo la guerra*. Laterza & figli.

Murray, C. J., Lopez, A. D., Chin, B., Feehan, D., & Hill, K. H. (2006). Estimation of potential global pandemic influenza mortality on the basis of vital registry data from the 1918–20 pandemic: a quantitative analysis. *The Lancet*, 368(9554), 2211-2218. https://doi.org/10.1016/S0140-6736(06)69895-4

Pham, H. (2020). On estimating the number of deaths related to Covid-19. *Mathematics*, 8(5), 655. https://doi.org/10.3390/math8050655

Pinnelli, A., & Mancini, P. (1998). Mortality peaks in Italy in the late 19th and early 20th centuries: trends by age and sex. *European Journal of Population*, 14(4), 333-365. https://doi.org/10.1023/A:1006182029835

Rivera, R., Rosenbaum, J., & Quispe, W. (2020). Estimating Excess Deaths in the United States Early in the COVID-19 Pandemic. *medRxiv*. https://doi.org/10.1101/2020.05.04.20090324

Viboud, C., Simonsen, L., Fuentes, R., Flores, J., Miller, M. A., & Chowell, G. (2016). Global mortality impact of the 1957–1959 influenza pandemic. *The Journal of infectious diseases*, 213(5), 738-745. https://doi.org/10.1093/infdis/jiv534

Yang, L., Chan, K. P., Cowling, B. J., Chiu, S. S., Chan, K. H., Peiris, J. S. M., & Wong, C. M. (2012). Excess mortality associated with the 2009 pandemic of influenza A (H1N1) in Hong Kong. *Epidemiology & Infection*, 140(9), 1542-1550. https://doi.org/10.1017/S0950268811002238

Appendix

Figure A.1 - Percentage of municipal excess deaths detected from 21.02.2019 to 30.04.2019 with respect to the predicted deaths estimated via ML techniques

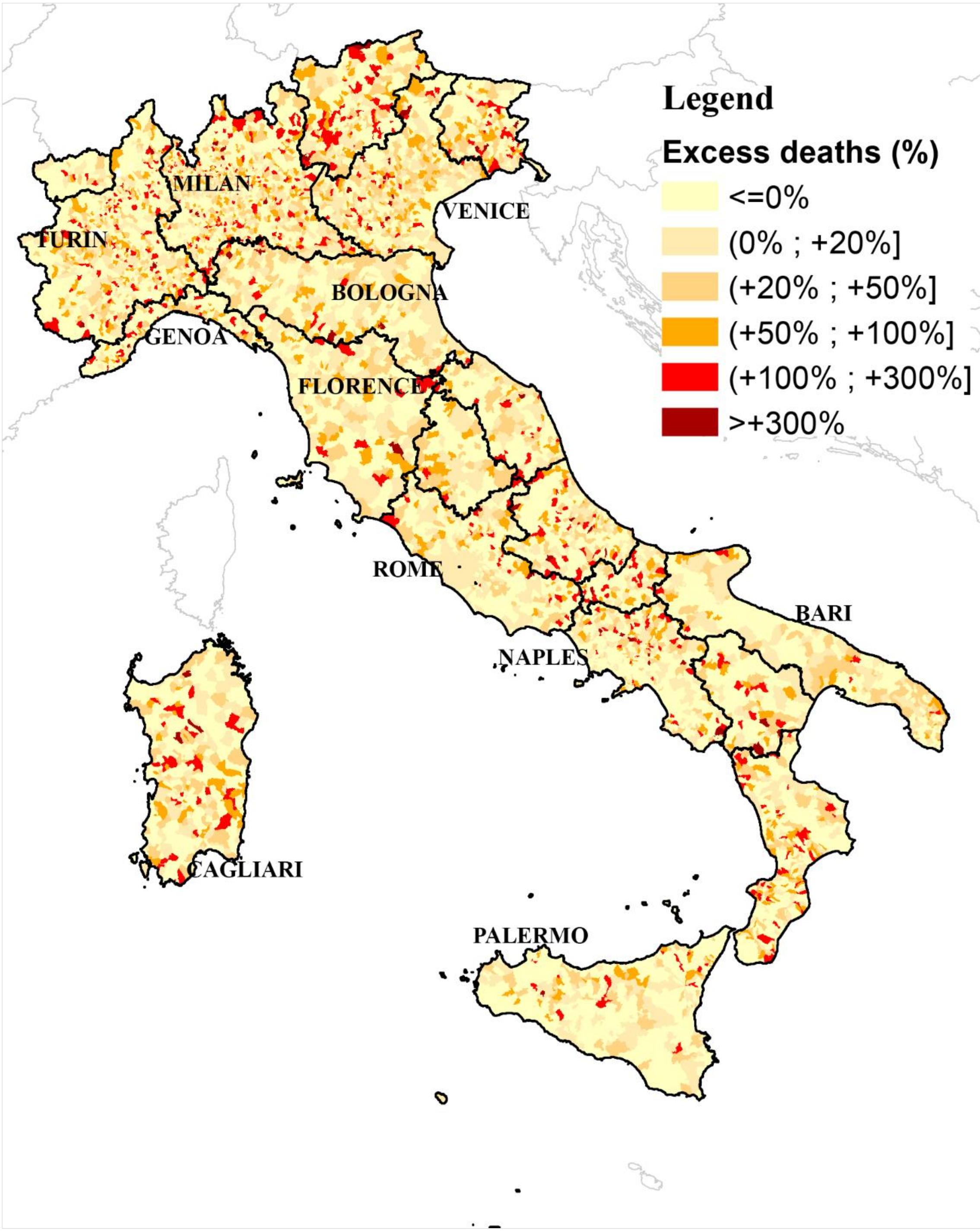


Table A.1 – Share of excess deaths 'observed' in 2019 and in 2020 by geographic and population size

Italy

	Share of municipalities with excess deaths above 50%		Share of municipalities with excess deaths above 100%		Share of municipalities with excess deaths above 300%	
	2019	**2020**	**2019**	**2020**	**2019**	**2020**
Overall	16.82%	32.35%	7.88%	17.86%	0.81%	3.30%
Less than 2,000 inhabitants	20.02%	35.22%	11.26%	20.51%	0.81%	3.82%
Inhabitants >= 2,000 and < 5,000	20.89%	33.74%	9.31%	18.05%	1.67%	3.19%
Inhabitants >= 5,000 and < 20,000	10.86%	29.73%	2.33%	15.93%	0.11%	3.27%
More than 20,000 inhabitants	1.34%	17.07%	0.00%	6.10%	0.00%	0.40%

North

	Share of municipalities with excess deaths above 50%		Share of municipalities with excess deaths above 100%		Share of municipalities with excess deaths above 300%	
	2019	**2020**	**2019**	**2020**	**2019**	**2020**
Overall	18.22%	46.50%	8.37%	28.41%	0.85%	5.74%
Less than 2,000 inhabitants	18.60%	45.25%	10.12%	29.44%	0.45%	6.30%
Inhabitants >= 2,000 and < 5,000	23.54%	49.13%	10.98%	30.19%	2.30%	5.58%
Inhabitants >= 5,000 and < 20,000	14.69%	47.84%	3.70%	27.23%	0.19%	5.78%
More than 20,000 inhabitants	2.49%	38.69%	0.00%	14.57%	0.00%	1.01%

Centre-South

	Share of municipalities with excess deaths above 50%		Share of municipalities with excess deaths above 100%		Share of municipalities with excess deaths above 300%	
	2019	**2020**	**2019**	**2020**	**2019**	**2020**
Overall	15.08%	13.71%	7.27%	3.99%	0.77%	0.10%
Less than 2,000 inhabitants	23.50%	20.67%	12.86%	7.57%	1.32%	0.22%
Inhabitants >= 2,000 and < 5,000	17.56%	13.13%	7.22%	1.80%	0.89%	0.00%
Inhabitants >= 5,000 and < 20,000	6.28%	6.16%	0.70%	1.23%	0.00%	0.00%
More than 20,000 inhabitants	0.62%	2.39%	0.00%	0.34%	0.00%	0.00%