COVID-19 in Italy: An Analysis of Death Registry Data

Gabriele Ciminelli[1,2], Sílvia Garcia-Mandicó[3]

[1]Finance and Economics, Asia School of Business, 2 Jalan Dato Onn, 50480 Kuala Lumpur, Malaysia
[2]MIT Sloan School of Management, 100 Main St, Cambridge, MA 02142, USA
[3]Organisation for the Economic Cooperation and Development, 2 rue du Conseiller Collignon, Paris, France
Address correspondence to Sílvia Garcia-Mandicó, E-mail: Silvia.GARCIA-MANDICO@oecd.org

ABSTRACT

Background There are still many unknowns about COVID-19. We do not know its exact mortality rate nor the speed through which it spreads across communities. This lack of evidence complicates the design of appropriate response policies.

Methods We source daily death registry data for 4100 municipalities in Italy's north and match them to Census data. We augment the dataset with municipality-level data on a host of co-factors of COVID-19 mortality, which we exploit in a differences-in-differences regression model to analyze COVID-19-induced mortality.

Results We find that COVID-19 killed more than 0.15% of the local population during the first wave of the epidemic. We also show that official statistics vastly underreport this death toll, by about 60%. Next, we uncover the dramatic effects of the epidemic on nursing home residents in the outbreak epicenter: in municipalities with a high share of the elderly living in nursing homes, COVID-19 mortality was about twice as high as in those with no nursing home in town.

Conclusions A pro-active approach in managing the epidemic is key to reduce COVID-19 mortality. Authorities should ramp-up testing capacity and increase contact-tracing abilities. Adequate protective equipment should be provided to nursing home residents and staff.

Keywords COVID-19, Italy, mortality, mass testing, registry data, Veneto

JEL classification I10, I18, I30

Introduction

There are still many unknowns about COVID-19. We do not know its true mortality rate nor the speed through which it spreads across communities. This lack of evidence complicates the design of appropriate response policies. The case of the UK is illustrative. The government first opted for bare minimum mitigation. It then drastically reversed course after microsimulations by the Imperial College COVID-19 Response Team showed that that strategy could have resulted in hundreds of thousands of deaths.[1]

This paper provides first-hand evidence on some of these questions. We focus on Italy's north, for long the world's worst affected area by COVID-19. Using daily death registry data from the Italian statistical agency, we first quantify the death toll from and by COVID-19 and show that many deaths went unrecorded in official statistics. COVID-19 was responsible for the death of about 45 000 people—more than 0.15% of the local population—during the first wave of the epidemic (from mid-February to mid-May 2020). Its mortality is vastly undercounted in official statistics. A plausible estimate suggests that true COVID-19 deaths might have been 60% higher than what is officially reported. At the peak of the epidemic, one additional death went undetected for each officially recorded COVID-19 fatality.

We then uncover large regional differences in mortality, which can be partly explained through the different approaches taken to deal with the epidemic. In the Veneto region, which embraced mass testing, contact tracing, and at-home care provision, COVID-19-induced mortality was, respectively, three and six times smaller than in neighboring Emilia-Romagna and Lombardy.

We close the paper by discussing the management of nursing homes in Lombardy, Italy's COVID-19 outbreak epicenter. Nursing homes have been in the spotlight across Europe and beyond for being possible hotspots of contagion and deaths. Our results suggest that living in a nursing home may

Gabriele Ciminelli, Dr.
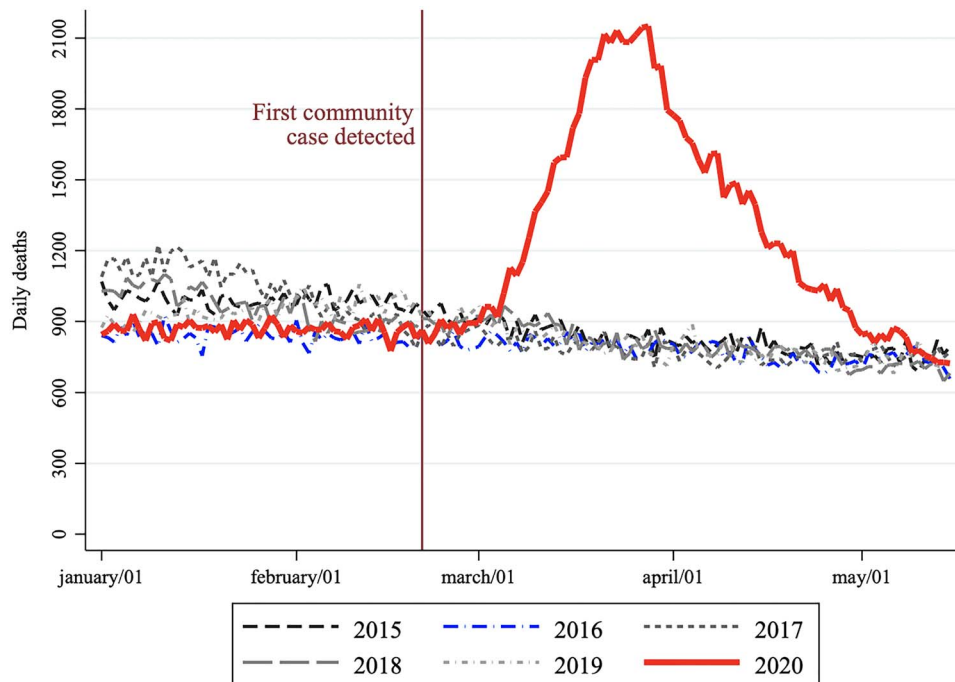
Sílvia Garcia-Mandicó, Dr.

Fig. 1 Daily deaths in 2020 compared to the five preceding years. Source: authors' own calculations from ISTAT daily death registry data.

have significantly increased the probability of dying during the COVID-19 epidemic. We find that in municipalities with a 10 percentage point higher share of people living in a nursing home, mortality among the elderly was about twice as high as in those with no nursing home in town. This result helps to rationalize the serious undercounting of COVID-19 fatalities in official statistics, which do not include deaths in nursing homes.

Methodology

Data

We source daily death registry data from ISTAT, the Italian statistical agency. (accessed 4th May 2020).[2] The data provide information on daily deaths by age and gender for the vast majority of Italian municipalities, for the period 1 January to 15 May, for the years 2015–2020. Our focus is on the eight regions of Italy's north (Emilia-Romagna, Friuli-Venezia Giulia, Liguria, Lombardy, Piedmont, Trentino-South Tyrol, Valle d'Aosta and Veneto), which together account for roughly half of Italy's population and about 85% of all COVID-19 fatalities (as of May 2020). Our data cover almost 94% of all municipalities in these eight regions and about 97% of their population. Table A1 in the Appendix reports descriptive statistics on the coverage of the dataset.

Figure 1 below takes a first look at the data by comparing deaths in 2020 (red line) against those in each of the

five preceding years. Soon after the detection of the first COVID-19 community case (vertical line), deaths increased rapidly. They jumped from about 900 at the end of February to over 2100 at the end of March. From the beginning of April, as the government social distancing policies began to bear an effect, deaths slowly started to decrease and returned to normal levels around mid-May. Our sample thus spans the entire first wave of the COVID-19 epidemic in Italy.

We then source Census population data from ISTAT and match them to death registry data to analyze COVID-19-induced mortality. Census data provide information on the resident population by age and gender, as of 1 January, for the years 2015–2019. (accessed 16th April 2020).[3] We impute 2020 values by using 2019 growth rates. We also source official COVID-19 fatality data from Protezione Civile, to compare officially reported mortality to the mortality calculated from death registry data.[2]

Lastly, we source municipality-level data on a host of co-factors of COVID-19 mortality, such as population density, commuting, digital employment, air pollution and several demographic characteristics. We also source data on health care management, in particular nursing homes, for the Lombardy region.

Empirical methodology

Deaths before the COVID-19 outbreak were following fairly closely deaths in 2016. Assuming that this trend would have
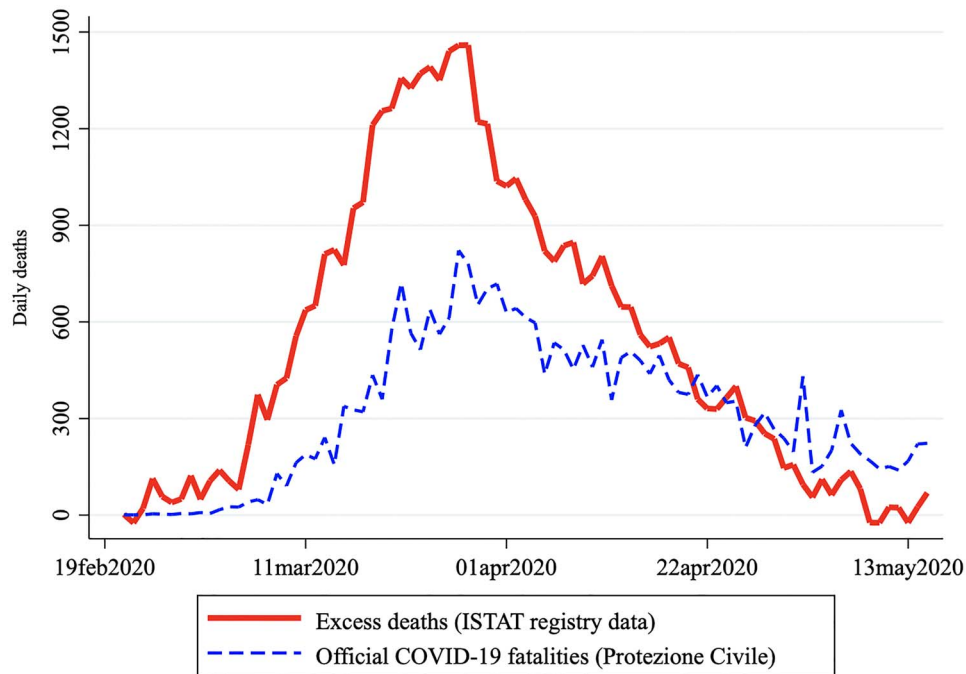
Fig. 2 Excess daily deaths in 2020 relative to 2016 vs. official COVID-19 fatalities. Source: authors' own calculations from ISTAT daily death registry data and Protezione Civile (2020).

continued in the absence of COVID-19, we calculate a measure of excess deaths by subtracting deaths in 2016 to deaths in 2020. To empirically evaluate the impact of COVID-19 on mortality, we embed the concept of excess deaths in a differences-in-differences regression model, which allows us to control for a host of potentially confounding factors. More detail on the methodology is provided in the Online Appendix.

Results

The scale of the epidemic is vastly underreported in official statistics

Using the concept of excess deaths, we calculate that the virus might have caused the death of about 45 000 people—more than 0.15% of the local population—during the first wave of the epidemic. Next, we compare excess deaths in 2020 to official COVID-19 fatality data. Figure 2 below plots the official daily number of COVID-19 fatalities (dashed blue line) and excess deaths in 2020 relative 2016, as recorded in municipal death registry data (solid red line). Strikingly, excess deaths were higher than official fatalities throughout the end of April, suggesting that COVID-19 deaths might have been vastly underreported in official statistics.

What is the scale of underreporting? Excess deaths include both indirect and direct COVID-19 deaths, while official data only cover direct deaths. To gauge the importance of indirect deaths, we consider the 12 regions of Italy, which are not included in our sample because they were relatively unaffected by COVID-19. These display ratios of excess deaths to official COVID-19 fatalities that are similar to the regions in our sample. Assuming that these regions correctly detected all COVID-19 deaths, we obtain a measure of indirect deaths as a share of the population, which we use to infer the number of indirect deaths in the eight regions in our sample. Using this approach, we estimate that indirect deaths might account for 3–5% of all deaths attributable to COVID-19. Discounting indirect deaths, we calculate that COVID-19 deaths may have been 60% higher than what was officially reported. At the peak of the epidemic (around end-March), deaths may have been underreported by about a factor of two. That is, for each officially recorded death, an additional one went undetected.

What can explain underreporting? In Italy, guidelines for the classification of COVID-19 fatalities vary by region, but in most cases, deaths outside hospitals are not counted in official statistics. Anecdotal evidence suggests that, as the health system struggled with a surge in demand, many old patients died of COVID-19 at home or in elderly care facilities, without being tested.[3,4] This hypothesis seems plausible since the extent of underreporting was higher at the peak of the epidemic. Limited testing capacity may be a complementary reason. While excess deaths from registry data decreased sharply after peaking at the end of March, official COVID-
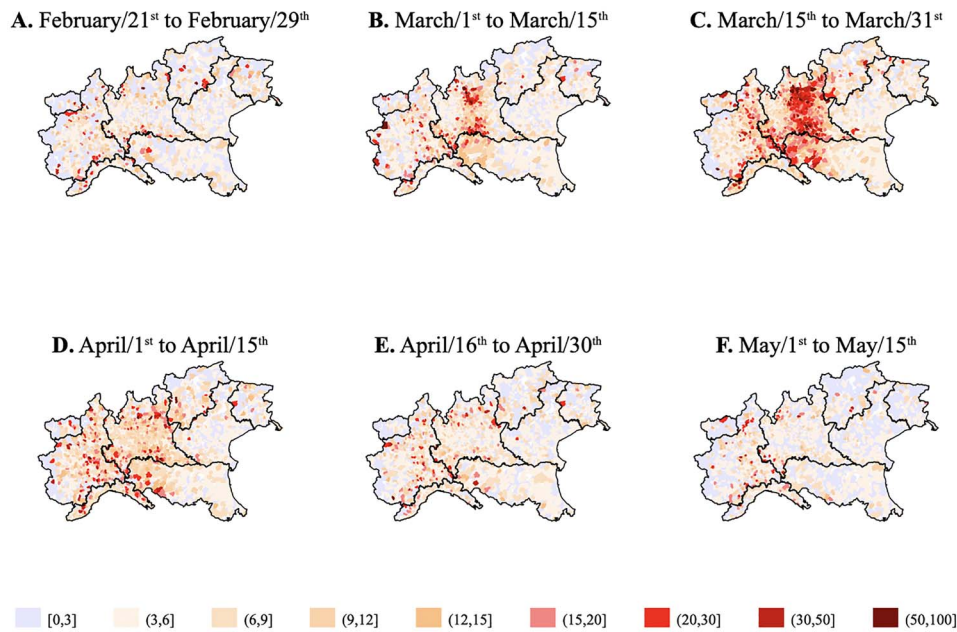
A. February/21st to February/29th   **B. March/1st to March/15th**   **C. March/15th to March/31st**

D. April/1st to April/15th   **E. April/16th to April/30th**   **F. May/1st to May/15th**

[0,3]   (3,6]   (6,9]   (9,12]   (12,15]   (15,20]   (20,30]   (30,50]   (50,100]

Fig. 3 Map of excess deaths in 2020 relative to 2016, by municipality. Notes: the figure shows municipal- and time-level variation in daily mortality rates per 100 000 inhabitants. Source: authors own calculations from ISTAT death registry and census data.

19 fatalities peaked later. They also did not fall as sharply, suggesting that the authorities may have progressively ramped up testing capability. In May, the number of official fatalities was even higher than that of excess deaths. This may be due to lags in the official reporting of fatalities relative to the fatality date.

Relatively few municipalities account for an outsized number of deaths

Data presented so far were aggregate for the entire sample of the 4100 municipalities covered in the dataset. In Fig. 3 below, we uncover variation in the extent of the epidemic both across different municipalities and over time. We assign different colors to municipalities, based on the mortality rate. Municipalities with a mortality rate within pre-epidemic 'normal' levels (0–3 daily deaths per 100 000 inhabitants) are denoted with blue. Those with a mortality rate between 3 and 15 daily deaths per 100 000 inhabitants are marked by different shades of orange, while red colors indicate those with very high rates (above 15 per 100 000 inhabitants).

At the start of the epidemic, the mortality rate was generally within normal levels, except for a few municipalities. By March, two large outbreaks in the Lombardy region had become apparent: one in the south, around the town of Codogno (the epicenter), and another in the north, around Alzano Lombardo (a town close to Bergamo). While deaths increased across most municipalities during the unfolding of the outbreak, the increase around these two clusters was much

higher. Along the border between Lombardy and neighboring regions, death rates were high in both Emilia-Romagna and Piedmont, while they stayed low in Veneto.

Next, we use a differences-in-differences method to confirm the patterns observed in Fig. 3 and quantify the effects of COVID-19 on the mortality rate by region. We estimate that the pandemic resulted in almost three deaths per 100 000 inhabitants per day in Lombardy, by far the region in which COVID-19 had the highest effect on mortality. Across the four regions bordering Lombardy, we estimate the lowest impact in Veneto, at around 0.5 daily death for every 100 000 inhabitants. That is about three times less than Trentino-South Tyrol and Piedmont and four times less than Emilia-Romagna—all sharing a long border with Lombardy—and suggests that the bold approach adopted by the Veneto region in managing the epidemic may have sensibly reduced COVID-19 mortality there. Details on the estimation and additional results are reported in the Online Appendix.

The Veneto region has been widely praised for its timely and proactive response to the pandemic. Much has been written about the 'Veneto Model',[5,6,7] and we only summarize its main characteristics: (i) mass testing, including testing of asymptomatic cases; (ii) at-home testing and care provision and (iii) tracing and quarantining contacts. Both ramping up the testing capacity and developing a comprehensive tracking system have been recognized as essential elements to coping with the epidemic.[8,9] Testing and providing care at home works towards preventing the spread of the virus.
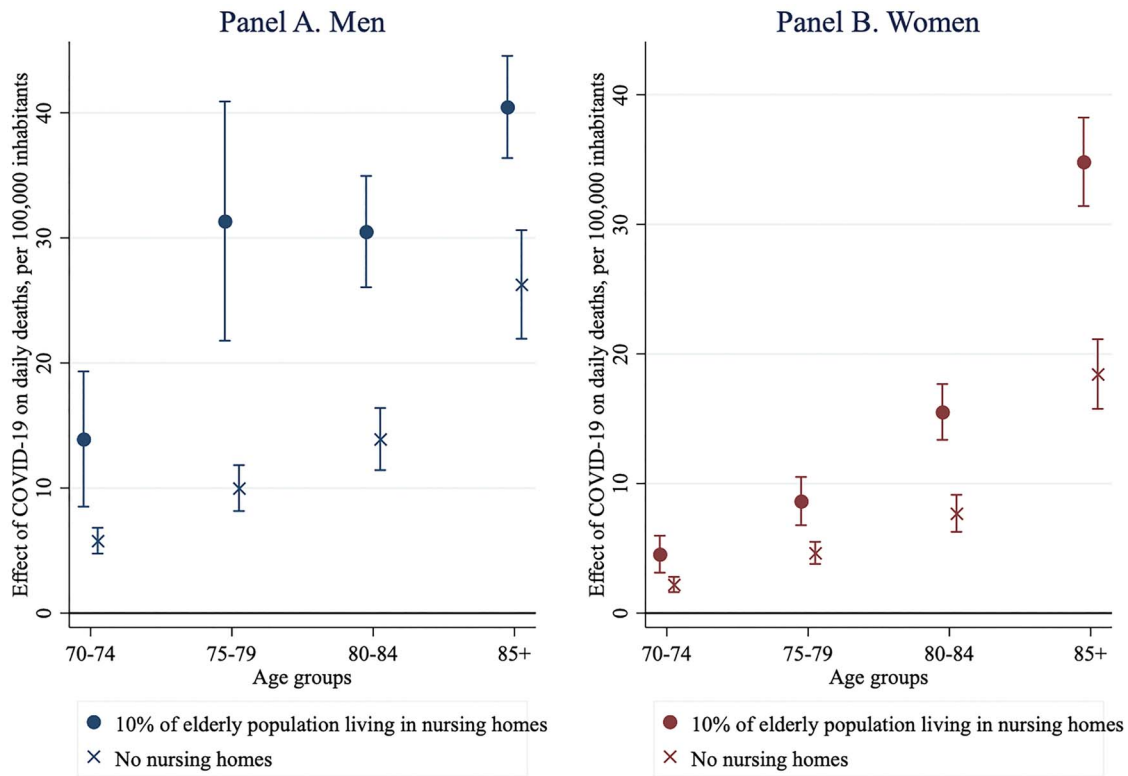
Fig. 4 Nursing homes and COVID-19 mortality, by gender and age. Notes: the figure shows the additional effect of COVID-19 on mortality for a 10% residency in nursing homes and age bins, for men (**A**) and women (**B**). Source: authors' own estimation based on ISTAT death registry and census data.

In a testament to the success of its strategy, on 21 May, the Veneto region registered zero new infections, while Lombardy had more than 300. The Veneto model provides useful lessons for policy makers around the world on how to manage new outbreaks of COVID-19 in the future.

Living in a nursing home may have increased the chances of dying for the elderly

Next, we explore gender and age differences in COVID-19 mortality. Details on the estimation are provided in the Online Appendix, while here we only summarize the main findings. We first show that mortality increases exponentially with age, but at much higher levels for men, confirming that gender differences play a crucial role in understanding the distribution of risk from the epidemic.[10] At the same time, underreporting greatly increases with age and is particularly high among older women.

Why do we observe so many undetected deaths among the elderly, and particularly so among women? All available anecdotal evidence points to nursing homes.[4,11] With a large number of residents sharing common spaces and having close contacts with multiple staff members, nursing homes may have acted as hotbeds of contagion. Moreover, as in Italy nursing homes do not qualify as medical centers, they

were heavily understaffed and unprepared to deal with the crisis, lacking protective equipment for staff and emergency care equipment for infected patients.[12] In Lombardy, these inherent characteristics may have been particularly aggravating, as the regional authority decided to relocate COVID-19 positive patients with mild symptoms from hospitals to nursing homes.[13] Since Italy does not include fatalities in nursing homes in its COVID-19 statistics, they likely account for a big chunk of the undetected deaths.

We zoom in on Lombardy, the worst affected region, to test whether COVID-19 had an additional effect on mortality in municipalities with a higher share of people living in nursing homes (see the Online Appendix for details on the estimation). In Fig. 4 below, we compare the mortality effect of COVID-19 in municipalities where 10% of the elderly population lives in nursing homes with no nursing home in town.

The results suggest that living in a nursing home may have significantly increased the chance of dying during the COVID-19 epidemic for both men and women. Strikingly, we find that COVID-19 mortality was twice as high in municipalities with a high proportion (10%) of the elderly living in a nursing home as in those municipalities with no nursing home in town.

Discussion

Main finding of this study

This paper provides first-hand evidence on the true death toll from and by COVID-19 using Italy's north—for long the world's worst affected area by COVID-19—as a case study. We show that COVID-19 caused the death of more than 0.15% of the local population during the first wave of the epidemic. We also show that many deaths go unrecorded in official statistics. A plausible estimate suggests that true deaths were about 60% higher than what was officially reported during the first wave of the COVID-19 epidemic in Italy. This undercounting is much more severe for the elderly, and in particular, for women. Since Italy does not include fatalities in nursing homes in its COVID-19 statistics, they likely account for a big chunk of the undetected deaths.

Our analysis quantifies the catastrophic effects of COVID-19 in nursing homes. In municipalities with a 10 percentage point higher share of the elderly population living in a nursing home, mortality was about twice as high as those with no nursing home in town. Our analysis shows that many of them could have been prevented through better preparedness. Providing adequate protective equipment is key to protecting nursing homes' residents and staff. Even more essential is the need to identify and isolate positive cases, and prevent staff from going to work if they are affected.

What is already known on this topic

The concept of excess deaths has been used by several statistical offices around the globe to quantify COVID-19 mortality. A notable example is the UK statistical office (ONS), which has produced relevant work comparing excess mortality across European countries.[14] Some media outlets have already relied on the concept of excess deaths to suggest that COVID-19 fatalities are undercounted in official statistics.[15,16] Others have provided some anecdotal evidence suggesting that nursing homes around the globe may have been hotbeds for contagion and mortality. For instance, The Economist has analyzed data from European countries that count deaths in nursing homes in official statistics and found that these make up for 30 percent of all official fatalities, on average. This ratio increases to 50 percent in Belgium.[11]

What this study adds

We use highly granular daily death registry data for thousands of municipalities in Italy's north to conduct a precise estimation of the true effect of COVID-19 on the mortality rate and compare the real death toll with what is reported in official statistics. In the next step, we extend the analysis by using detailed health care data as well as data on a host of socio-demographic, labor market and territorial characteristics to credibly estimate the 'nursing homes' effect on COVID-19 mortality.

Limitations of this study

This study relies on the definition of excess deaths to uncover the effect of COVID-19 and draw some policy lessons. This approach does not allow us to calculate the case fatality rate of COVID-19, which is essential to understand how prevalent the virus is in the population and thus inform social distancing policies.[17] Moreover, our definition of excess deaths attributable to COVID-19 includes both direct deaths—people dying of COVID-19—and indirect deaths—people dying for causes related to COVID-19, such as overcrowded hospitals.[18]

Supplementary data

Supplementary data are available at the *Journal of Public Health* online.

Acknowledgements

Disclaimer

The views expressed in this paper are those of the authors and do not represent those of the OECD or its member countries.

References

1 Ferguson N, Laydon D, Nedjati Gilani G *et al. Report 9: Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID19 Mortality and Healthcare Demand*, Imperial College London, 2020.

2 Civile P. *Bollettino Emergenza Coronavirus*, Roma: Dipartimento di Protezione Civile, 2020.

3 Reuters. *Death at Home: the Unseen Toll of Italy's Coronavirus Crisis*. United States: Reuters World News, 2020.

4 The Economist. *Many COVID Deaths in Care Homes are Unrecorded*, The Economist Group, London 2020.

5 Cedrone G. 'Scovare i positivi casa per casa: così abbiamo sconfitto il virus a Vo'. Il virologo Crisanti racconta il modello Veneto. In: *Sanità e Informazione*, Roma: Multichannel Media Production s.r.l., 2020.

6 Zanini L. *Managing the Pandemic: Lessons from Italy's Veneto Region*. Medium.com, Roma, 2020.

7 Zingales L. Why mass testing is crucial: the U.S. should study the Veneto model to fight COVID-19. In: *Pro Market Blog*, Chicago: Stigler Center, 2020.

8 Baldwin R. *COVID-19 Testing for Testing times: Fostering Economic Recovery and Preparing for the Second Wave*. VoxEU.org, London 2020.

9 Baldwin R, Weder di Mauro B. *Mitigating the COVID Economic Crisis: Act Fast and Do Whatever It Takes*. VoxEU.org eBook, London, 2020.

10 Wenham C, Smith J, Morgan R. COVID-19: the gendered impacts of the outbreak. *The Lancet* 2020;**395**(**10227**):846–8.

11 The Economist. *The Impact of COVID-19 On Care Homes*, . London: The Economist Group, 2020.

12 ATS Insubria. *Liste di Attesa nelle RSA*. Sistema Socio Sanitario Regione Lombardi, 2020.

13 La Stampa. *Fontana su anziani morti nelle Rsa lombarde: 'L'utilizzo di quelle case di riposo per i malati Covid è responsabilità dell'ATS*, Torino: GNN-GEDI Gruppo Editoriale S.p.A, 2020.

14 ONS. *Comparisons of All-cause Mortality Between European Countries and Regions: January to June 2020*, Office for National Statistics U.K., 2020.

15 Financial Times. *Global Coronavirus Death Toll could be 60% Higher than Reported*, London: The Financial Times Group, 2020.

16 The Economist. *Tracking COVID-19 Excess Deaths Across Countries*, London: The Economist Group, 2020.

17 Stock JH. *Data Gaps and the Policy Response to the Novel Coronavirus*. NBER Working Paper, USA: NBER, Cambridge, 2020.

18 The Economist. *Deaths from Cardiac Arrest Have Surged in New York City*. London: The Economist Graphic Detail, 2020.

A. Appendix

Table A1 Death registry data coverage

Region	Municipalities	% of all municipalities	Population (in 1000)	% of all population
Emilia-Romagna	304	92.68	4412	96.29
Friuli Venezia Giulia	200	93.46	1206	93.15
Liguria	218	93.56	1581	97.06
Lombardy	1455	97.00	10 499	98.43
Piedmont	1099	93.69	4486	95.33
Trentino-South Tyrol	256	88.58	1095	92.36
Valle d'Aosta	68	91.89	134	90.88
Veneto	500	87.87	4684	91.01
Full sample	4100	93.58	28 097	95.71

Notes: 'municipalities' report the number of municipalities covered in the daily death registry data. '% of all municipalities' reports the share of municipalities covered over the region's total. 'Population (in 1000s)' indicates the number of residents in the municipalities covered by the death registry data, imputed from Census population data, as of 1 January 2020. '% of all population' reports the share of the population covered over the region's total 'full sample' report statistics for the full sample of municipalities.

Sources: ISTAT.[2,3]