

News from the COVID-19 Epicenter^{*}

Paolo Buonanno^a, Sergio Galletta^{a,b}, Marcello Puca^{a,c}

^a*University of Bergamo*

^b*ETH Zürich*

^c*Webster University Geneva*

Abstract

We provide new results on the misreported level of mortality in the Italian region of Lombardy and in the province of Bergamo using official and original data sources. Since February 2020 Lombardy and in particular the province of Bergamo have been severely hit by the novel COVID-19 infectious disease. Combining official statistics, retrospective data and original data (i.e., obituaries and death notices) we provide a tentative estimate of the “real” number of deaths caused by COVID-19 as well as the total number of persons infected. Our findings suggest that the reported mortality rate attributable to COVID-19 accounts only for 26.6% of the observed mortality rate.

^{*} *This version: April 2, 2020 (First version: April 2, 2020)*

Email addresses: `paolo.buonanno@unibg.it` (Paolo Buonanno), `sergio.galletta@unibg.it` (Sergio Galletta), `marcello.puca@unibg.it` (Marcello Puca)

1. Introduction

The Lombardy region is at the epicenter of the novel coronavirus (COVID-19) Italian epidemic. At the end of March 2020 official reporting counts 7,593 deaths and 44,773 infected individuals, respectively the 57.7% and 40.5% of all the Italian cases.

Existing contributions have assessed the severity of COVID-19 providing either “crude” case fatality rate (CFR) or adjusted-biased measure (Ferguson et al., 2020; Jung et al., 2020; Mizumoto et al., 2020).¹ Yet, these measures have the limit of being representative only of the detected infected population, providing narrow guidance to policymakers as they do not clearly reveal the actual effect of COVID-19 on the general population.

This research aim is to improve on existing evidence by providing a first estimate of the mortality rate in the general population. Specifically, we compare the official number of deaths reported in March 2020 with those reported in the same month of the previous year in a sample of Lombard municipalities. With this result in hand, we recover the expected number of total persons infected for different levels of hypothetical CFR. Our results are confirmed by using an alternative method that we propose as a valid substitute for the use of unreliable (or unavailable) official data.

2. How severe is under-reporting in COVID-19 diffusion and fatality rate in Lombardy?

To assess whether the reported number of patients who died from COVID-19 is consistent with the actual number of deaths observed in the same period, we begin our investigation comparing the number of all deaths observed in March 2019 to those of March 2020 for a sample of Italian municipalities in the region of Lombardy. Mortality rates for this sample of municipalities have been released by the Italian National Statistical Institute (ISTAT) on April 1, 2020. The sample contains municipalities which, between January 1 2020 and March 21 2020, registered an increase of the mortality rate

¹One can expect “crude” estimate to be biased for several reasons (Ghani et al., 2005). On the one hand, there might be an underestimate during the first phases of an epidemic if several days pass between the moment an individual experiences the first symptoms and the day health authorities detect the infection. On the other hand, there might be an overestimate of the fatality rate if health authorities tend to concentrate only on severe cases, resulting in poor information on milder cases. Similar bias will be present if the health-care system has reached its capacity constraint and is not able to detect further cases regardless of the severity. Adjusted estimates might as well face measurement issues as they mainly rely on the underlying assumptions on how to attenuate the bias.

of at least 20% compared to the period 2015-2019.² We merge this information with data on COVID-19 deaths at municipality level obtained from the Lombardy Region website on March 25, 2020.³ The list contains, for each Lombardy’s municipality, the number of dead patients who tested positive during hospitalization or *post mortem*. The final dataset contains observations for 434 of the 1,506 total Lombardy’s municipalities, accounting for roughly 60% of Lombardy’s total population.

Table 1: Summary statistics

Variables	Obs. (1)	Mean (2)	Std. dev. (3)	Min (4)	Max (5)
<i>Municipality level</i>					
COVID Deaths	428	6.750	15.18	0	180
Total Deaths March 2019	434	8.111	43.48	0	885
Total Deaths March 2020	434	19.79	56.47	2	1,039
Excess Mortality Rate	434	0.165	0.188	-0.0394	1.418
Excess Unexpected Mortality Rate	428	0.0850	0.134	-0.107	1.233

Notes: This table reports summary statistics about the number of COVID-19 deaths, March 2019 deaths, March 2020 deaths, excess mortality rates and excess unexpected mortality rates for all municipalities in the sample.

Table 1 reports basic summary statistics for the municipalities in our sample. At the aggregate level, we report an increase going from 3,504 deaths in March 2019 to 8,560 in March 2020, consisting of a massive 266% growth of 2020 average mortality rate compared to 2019.⁴ As of March 21, 2020, however, only 1,866 deaths have been officially accounted as COVID-19 deaths in those municipalities. This implies that the official number of fatalities attributed to COVID-19 only account for 26.6% of the observed excess mortality. Narrowing down our focus, Figure 1 reports the distribution of mortality rates for Lombardy’s provinces. The difference between the official COVID-19 deaths and our measure of deaths is glaring and observed across provinces. There are several possible explanations for this form of under-reporting observed in the official statistics. A first may be attributable to the higher fatality ratio in older age groups, in line with Ferguson et al. (2020), a finding that we also point out in Section 4. Further, infected

²For more information see <https://www.istat.it/it/archivio/240401>.

³Data refer to the status reported by public authorities on the March 24, 2020. Data are available at https://www.arcgis.com/home/webmap/viewer.html?url=https://services1.arcgis.com/XannvQVnsM1hoZyv/ArcGIS/rest/services/COMUNI_COVID19/FeatureServer&source=sd

⁴For each municipality and year, we compute the mortality rate as the number of reported deaths over the total number of inhabitants.

Table 2: Number of individuals likely to be infected, by province

Province	Hypothetical COVID-19 fatality rate										
	1%	1.1%	1.2%	1.3%	1.4%	1.5%	1.6%	1.7%	1.8%	1.9%	2%
Bergamo	26.90%	24.45%	22.41%	20.69%	19.21%	17.93%	16.81%	15.82%	14.94%	14.16%	13.45%
Brescia	12.80%	11.64%	10.67%	9.85%	9.15%	8.54%	8.00%	7.53%	7.11%	6.74%	6.40%
Como	8.80%	8.00%	7.34%	6.77%	6.29%	5.87%	5.50%	5.18%	4.89%	4.63%	4.40%
Cremona	20.60%	18.73%	17.17%	15.85%	14.71%	13.73%	12.88%	12.12%	11.44%	10.84%	10.30%
Lecco	11.03%	10.03%	9.19%	8.49%	7.88%	7.36%	6.90%	6.49%	6.13%	5.81%	5.52%
Lodi	12.05%	10.96%	10.05%	9.27%	8.61%	8.04%	7.53%	7.09%	6.70%	6.34%	6.03%
Mantova	11.32%	10.29%	9.44%	8.71%	8.09%	7.55%	7.08%	6.66%	6.29%	5.96%	5.66%
Milano	7.81%	7.10%	6.51%	6.01%	5.58%	5.21%	4.88%	4.59%	4.34%	4.11%	3.91%
Monza	9.59%	8.72%	7.99%	7.38%	6.85%	6.40%	6.00%	5.64%	5.33%	5.05%	4.80%
Pavia	13.06%	11.88%	10.89%	10.05%	9.33%	8.71%	8.17%	7.68%	7.26%	6.88%	6.53%
Sondrio	12.33%	11.21%	10.28%	9.49%	8.81%	8.22%	7.71%	7.25%	6.85%	6.49%	6.17%
Varese	8.33%	7.57%	6.94%	6.40%	5.95%	5.55%	5.20%	4.90%	4.63%	4.38%	4.16%
Lombardy	12.37%	11.24%	10.31%	9.51%	8.84%	8.25%	7.73%	7.28%	6.87%	6.51%	6.18%

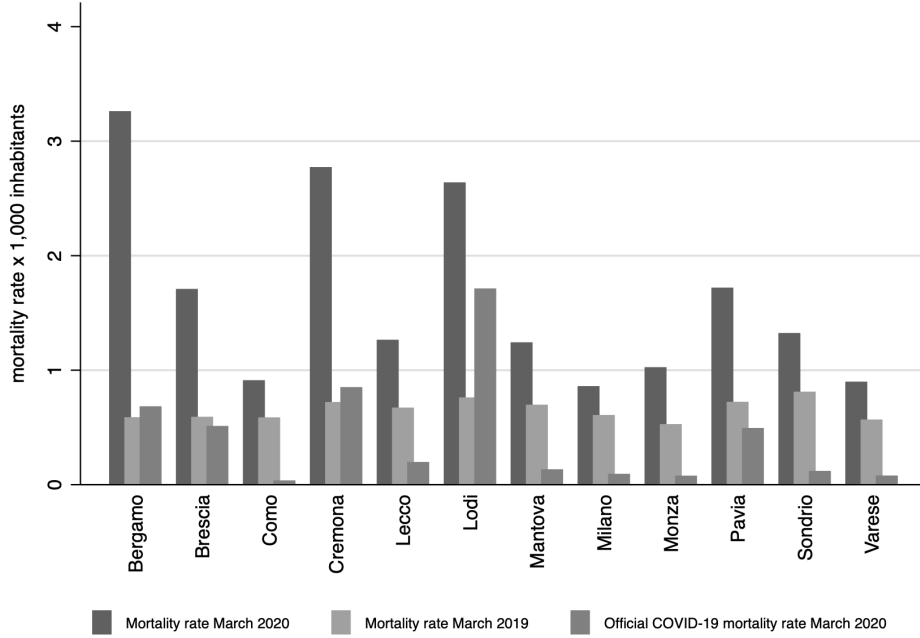
Notes: This table reports, for each province, the share of population that is likely to be infected for a range of different hypothetical COVID-19 fatality rate. These estimates are based on our sample of 428 municipalities covering 56% of the regional population.

patients may die at home without having the ability to be hospitalized and therefore be tested. Presumably, this issue exacerbates for older patients, who are more likely to have a bad pre-existent clinical condition. Connected to this issue, the mere capacity constraint on the number of tests that laboratories can run per day may compromise the correct assessment of the total number of infected patients. This evidence suggests that the actual number of COVID-19 deaths and, therefore, the exact number of infected individuals, is much higher than those reported in official records.

To assess the actual infection rate, we perform a back-of-the-envelope computation using conservative case fatality ratios reported in [Ferguson et al. \(2020\)](#). By assuming a constant fatality ratio across municipalities, a conservative case fatality ratio should range between 1% and 2%. Using these numbers, the real share of infected individuals in Lombardy should range between 6% and 12% of the population.⁵ Table 2 reports a sensitivity analysis of these calculations for different fatality ratios and each Lombardy’s province.

⁵It is worth noting, that these estimates are based on our limited sample of municipalities which experienced a higher increase in mortality rates between March 2019 and March 2020, compared to the other not included municipalities. Yet, we are covering 56% of the regional population.

Figure 1: Distribution of mortality rates (by province)



Notes: This graph reports the actual mortality rate in March 2020 and March 2019 as well as the reported mortality rate of COVID-19 as detected by the health authorities, by provinces.

3. An alternative source of data: obituaries and death notice

An alternative source for measuring the number of deaths is represented by obituaries or death notices published in local newspapers. Obituaries always contain individual characteristics such as name, surname, age, gender, date of death and place of death (municipality) representing an important and additional source of information to understand the phenomenon better. We digitalized obituaries published by L'Eco di Bergamo⁶ that is the most read and circulated daily newspaper in the province of Bergamo.⁷ Our final dataset contains 2,896 unique individuals from March 1 to March 31, 2020, while we count 1286 unique records from March 1 to March 21, 2020. In March 2019 the number of total deaths registered in the entire province was 886 less than one-third of deaths reported this March in newspaper obituaries. Summary statistics show that the average age is 80.9 years, and males represent 59.8% of fatalities. There has been a significant increase in the daily number of published obituaries after the first week of

⁶www.ecodibergamo.it

⁷In 2017, the number of sold copies has been 35,248, while the number of readers 336,000 <http://www.opq.it/testata.php?ID=1>.

March being constantly over 100, while during the previous year the daily number was on average around 20. Restricting our analysis to the municipalities in the province of Bergamo for which ISTAT released the official data of mortality from 1 to 21 march, we observed that on average obituaries account for 60% of the total registered deaths in the sample. Thus, we may estimate that the true number of deaths at the provincial level is roughly twice the number of obituaries and death notice published in the newspaper. This means that the total number of deaths in March 2020 is 4,826, while the excess mortality with respect to March 2019 is 4,826 minus 886 equals to 3,940 probably due to COVID-19. Considering a fatality rate of 1.57% (Ferguson et al., 2020) the true number of cases should be 250,000 in the province of Bergamo that has 1.1 million inhabitants. Considering more conservative estimates of the fatality rate of 1.1% the number of cases raises to 358,000, roughly one-third of the population of the province. Estimates using obituaries and death notices are consistent with previous estimates reported in Section 2.

4. Suggesting evidence using individual-level data

Finally, we collected data for individuals who tested positive to COVID-19 in the Italian region of Lombardy in March 2020.⁸ The dataset contains a total of 41,833 cases. We show in this section some graphical results using the “crude” measure of CFR.

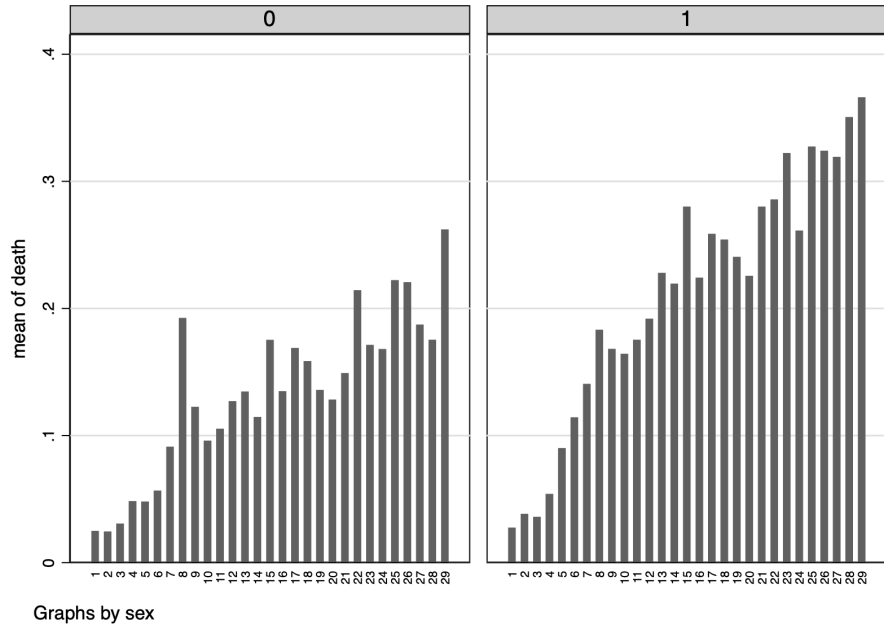
Despite the kind of biases we are facing using this measure, we can still provide informative elements to assess the relative risks between different demographic groups.⁹ In principle, all individuals that belong to the sample of detected positive had faced the same potential selection bias. To account for the possible delay between COVID-19 detection and patient’s recovery or death (i.e., data censoring), we report graphical evidence about the variation of individual death risks over time, starting from the day a patient tested positive, and for different demographic characteristics.

In Figure 2 we report such results by gender, while in Figure 2 we focus on different age groups. In line with previous studies – e.g. Ferguson et al. (2020) – we find a higher

⁸The available data do not identify the actual day of death. We only observe whether an individual in the sample was dead or alive at the moment we collected the data, i.e., on March 31, 2020. Data are available at https://www.arcgis.com/home/webmap/viewer.html?url=https://services1.arcgis.com/XannvQVnsM1hoZyv/ArcGIS/rest/services/TA_COVID19_RL/FeatureServer&source=sd

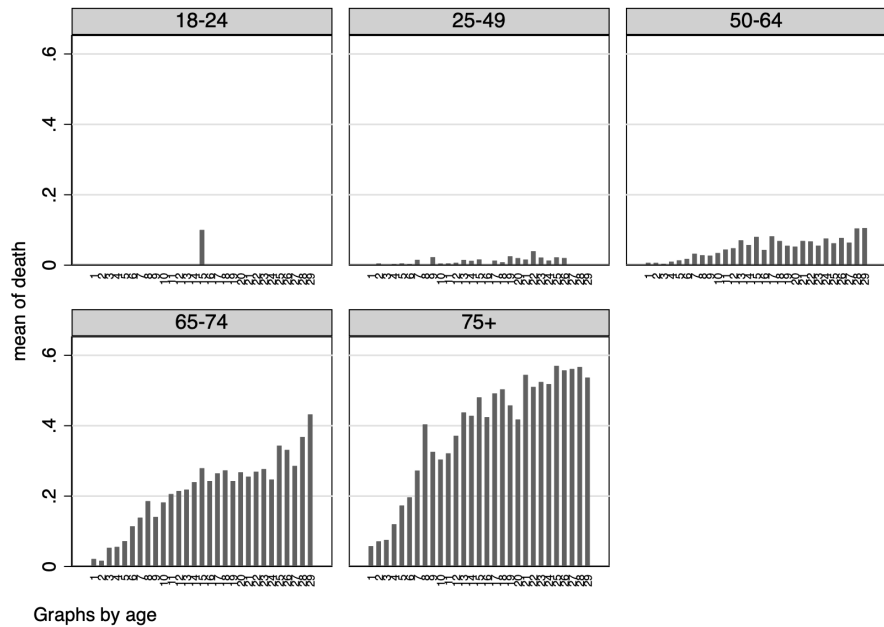
⁹Because of possible issues related with measurement errors and sample selection, we decided not to comment the absolute numbers.

Figure 2: Risk of death (by gender)



Notes: These graphs report the probability of being dead depending on the number of days since an individual has tested positive to COVID-19, by gender. Females on the left and male on the right.

Figure 3: Risk of death (by age group)



Notes: These graphs report the probability of being dead depending on the number of days since an individual has tested positive to COVID-19, by age group.

severity of the disease in older age groups as well as in male individuals. Also, individual death risk increase over time since the day of COVID-19 detection.

5. Concluding remarks

Since the spread of the novel COVID-19 emerged in Lombardy in February 2020, thousands of patients have died, especially in the Province of Bergamo. Understanding the diffusion and assessing the fatality rate of COVID-19 is crucial for implementing effective public and health policies in tackling the virus. Unfortunately, official reporting and statistics significantly underestimate the “true” numbers since there exists a vast proportion of asymptomatic infected patients among all infected individuals who are not detected. Anecdotal evidence suggests that only patients experiencing severe clinical symptoms are eventually tested with laboratory diagnostic methods. In this contribution, combining official statistics, retrospective data and original data we provide a tentative estimate of the “true” number of deaths caused by COVID-19 as well as the total number of persons infected. Our estimates, not free from limitations, suggest that the reported mortality rate attributable to COVID-19 accounts only for 26.6% of the observed mortality rate.

References

- Ferguson, N.M., Verity, R., Okell, L.C., *et al*, 2020. Estimates of the severity of coronavirus disease 2019: a model-based analysis. The Lancet Infectious Diseases URL: <http://www.sciencedirect.com/science/article/pii/S1473309920302437>, doi:[https://doi.org/10.1016/S1473-3099\(20\)30243-7](https://doi.org/10.1016/S1473-3099(20)30243-7).
- Ghani, A., Donnelly, C., Cox, D., *et al*, 2005. Methods for estimating the case fatality ratio for a novel, emerging infectious disease. American journal of epidemiology 162, 479–486.
- Jung, S.m., Akhmetzhanov, A.R., Hayashi, K., *et al*, 2020. Real-time estimation of the risk of death from novel coronavirus (covid-19) infection: Inference using exported cases. Journal of clinical medicine 9, 523.
- Mizumoto, K., Kagaya, K., Zarebski, A., Chowell, G., 2020. Estimating the asymptomatic proportion of coronavirus disease 2019 (covid-19) cases on board the diamond princess cruise ship, yokohama, japan, 2020. Eurosurveillance 25. URL: <https://eurosurveillance.org/content/10.2807/1560-7917.ES.2020.25.10.2000180>, doi:<https://doi.org/10.2807/1560-7917.ES.2020.25.10.2000180>.