medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070912; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

# An empirical estimate of the infection fatality rate of COVID-19 from the first Italian outbreak

Gianluca Rinaldi[☆a], Matteo Paradisi[☆b]

[a] *Harvard Economics Department*
[b] *NBER and Einaudi Institute for for Economics and Finance (EIEF)*

## Abstract

**Background:** The coronavirus 2019 (COVID-19) pandemic has been spreading globally for months, yet the infection fatality ratio of the disease is still uncertain. This is partly because of inconsistencies in testing and death reporting standards across countries. Our purpose is to provide accurate estimates which do not rely on testing and death count data directly but only use population level statistics.

**Methods:** We collected demographic and death records data from the Italian Institute of Statistics. We focus on the area in Italy that experienced the initial outbreak of COVID-19 and estimated a Bayesian model fitting age-stratified mortality data from 2020 and previous years. We also assessed the sensitivity of our estimates to alternative assumptions on the proportion of population infected.

**Findings:** We estimate an overall infection fatality rate of 1·29% (95% credible interval [CrI] $0{\cdot}89 - 2{\cdot}01$), as well as large differences by age, with a low infection fatality rate of 0·05% for under 60 year old (CrI $0 - 0{\cdot}19$) and a substantially higher 4·25% (CrI $3{\cdot}01 - 6{\cdot}39$) for people above 60 years of age. In our sensitivity analysis, we found that even under extreme assumptions, our method delivered useful information. For instance, even if only 10% of the population were infected, the infection fatality rate would not rise above 0·2% for people under 60.

**Interpretation:** Our empirical estimates based on population level data show a sharp difference in fatality rates between young and old people and firmly rule out overall fatality ratios below 0·5% in populations with more than 30% over 60 years old.

☆Contributed Equally

*April 18, 2020*

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070912; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

## 1. Introduction

Estimating the severity of the coronavirus disease 2019 (COVID-19), caused by the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) across demographic groups is of medical interest, but will also be crucial in the design of health and economic policies to deal with the next phases of the pandemic.

Several approaches to measure lethality have been proposed since the start of the outbreak[1,2,3,4], with most of the relevant population infected. Given the mounting evidence that a substantial proportion of infected people are either asymptomatic or display very mild symptoms[5,6], as well as the difficulties many countries are encountering in ramping up testing, it has been difficult to obtain precise estimates of the total number of infected. Testing mainly symptomatic cases on the basis of clinical studies on the symptoms of COVID-19[7,8,9,10,11], as announced for instance by the Italian government, might also have led to underestimating the total number of infected. While measurement of deaths is reliable, statistically significant deviations in total deaths relative to previous years have been observed in the most affected areas, leading to concerns that official COVID-19 death counts might also be underestimated in some cases[12]. Together, those hurdles make estimating the true infection fatality rate challenging.

To sidestep some of those issues, in this article we used an empirical approach employing publicly available aggregate deaths and demographic data to obtain infection fatality ratio estimates without relying on official data on COVID-19 positive cases and deaths. The key observation is that, assuming an accurate measurement of fatalities in a population, infection fatality ratio estimates are less strongly dependent on accurate measurement of total cases when the share of population infected is larger: keeping the number of deaths fixed, the estimate changes much less when the fraction of the population infected varies from 40% to 60% than when it changes from 2% to 3%, simply because ratios are nonlinear.

We were therefore able to obtain precise fatality estimates by age range focusing on one of the hardest hit areas in Lombardy, which was placed under lock-down order already on February 21st 2020. This area includes ten towns and has a population of around 50 thousands people. The first recorded patient infected through community spread in Italy was admitted

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070912; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

to intensive care on February 20th and the area already had 36 confirmed cases the day after[13]. While widespread randomized antibody testing hasn't yet been conducted in this area, 30% of a sample of blood donors from all ten municipalities tested positive to antibody[14], and a smaller sample of 60 asymptomatic blood donors in one of the towns under lock-down showed 40 (66%) positive cases[15]. Although these samples might not be fully representative of the population of these areas, these figures are highly suggestive of widespread contagion.

## 2. Methodology

### *2.1. Data*

We focused on ten Italian municipalities in Lombardy that experienced the initial outbreak of COVID-19. Data on deaths has been collected from the Italian Institute of Statistics (ISTAT)[16]. We built estimates of total death counts based on daily data on recorded deaths for 2020 and previous years for the period 2015-2019 that ISTAT collects from the *Anagrafe Nazionale della Popolazione Residente* (National Census of Resident Population). The data contains information about gender and age group (in 5-years bins) for each recorded death until April 4th 2020. ISTAT has released data for only seven of the ten municipalities that experienced the initial outbreak. Hence, we focused our analysis on this subsample. The excluded towns are much smaller than the others: the total population of those three towns in 2019 was 3,543, while the remaining seven municipalities had a total population of 47,020.

Death counts data has been complemented with information provided by ISTAT on the demographics of each municipality[17]. For every city we collected total population by age range in each year from 2015 to 2019. We used year 2019 information for year 2020 since data on 2020 has not been released yet.

### *2.2. Comparing 2020 deaths to previous years*

In order to motivate the use of administrative death counts for our infection fatality rate estimation, we begin by analysing patterns in overall mortality in 2020 and previous years. In particular, on every day between February 21st (the beginning of the outbreak) and April 4th the difference between the number of 2020 deaths and the 2015-2019 average has been computed for total population and for different age groups. We compared total deaths to

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070912; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

average number and fluctuations in deaths in past years to assess the signal to noise ratio of this measure.

### *2.3. Bayesian Estimation of COVID-19 infection fatality rate*

We employ a Bayesian framework to estimate the infection fatality rate of COVID-19 by adapting a standard binomial mortality model to our setting[18]. The likelihood function of the model is obtained by assuming deaths in the period between February 28th and April 4th of each year are binomially distributed according to:

$$D_{i,a,y} \sim \text{Binomial}\left(\delta_a \cdot \theta_i, N_{i,a,y}\right) \quad \text{for} \quad y \in \{2015, ..., 2019\} \tag{1}$$

$$D_{i,a,2020} \sim \text{Binomial}\left(\delta_a + \delta_a^{\text{Covid}} \cdot \theta_i, N_{i,a,2020}\right) \tag{2}$$

where $i$ denotes the municipality, $y$ the year, and $a$ the age range. We used seven age ranges: $0-20$, $21-40$, $41-50$, $51-60$, $61-70$, $71-80$, $81+$. $D_{i,a,y}$ and $N_{i,a,y}$ are the total deaths and population in town $i$, year $y$ and age range $a$, respectively. The baseline lethality rates $\delta_a$ are heterogeneous across age ranges, but were assumed to be constant across municipalities and years.
$\delta_a^{Covid}$ is the infection fatality rate for age range $a$ and we assumed $\delta_{a,t}^{Covid} = 0$ in every year before 2020 when COVID-19 was not present. We also assumed that infection rates $\theta_i$ are heterogeneous across municipalities but constant across age groups.
We assumed the following priors for the parameters of interest:

$$\delta_a \sim \text{Uniform}[0, 0{\cdot}1] \tag{3}$$

$$\delta_a^{\text{Covid}} \sim \text{Uniform}[0, 0{\cdot}3] \tag{4}$$

$$\theta_i \sim \text{Beta}(3, 2). \tag{5}$$

Priors on baseline and COVID-19 death rates were chosen to be uninformative, while we chose the prior on infection rates to reflect the results of the antibody testing in one of the municipalities[15], while at the same time maintaining a weakly informative prior.
We implemented a Bayesian procedure to derive point estimates and credible intervals for the infection fatality rates. The model was estimated using Markov Chain Monte Carlo (MCMC). We calculated the median and 95% credible interval using to the quantiles of the posterior distribution for all parameters. To check the sensitivity of our estimates we also calculated point

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070912; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

estimates as the mode of the posterior distribution and confidence intervals as 95% highest posterior density interval[19].

We fitted our model using R version 3.6.2. We drew 100,000 samples from the joint posterior distribution and used 10 independent chains, discarding the first 1000 samples for each chain. Trace plots of the Markov Chain Monte Carlo as well as posterior distributions for each variable are reported in Appendix A. All analyses are fully reproducible with the code available online.

Because our Bayesian procedures relies on modelling assumptions to derive the infected portion of the population, we also implemented a more agnostic approach, showing how infection fatality rates vary by contagion rate. In this exercise we computed the infection fatality rates from a simplified version of the model above in which we set a degenerate prior for each of the $\theta_i$ to be equal to a constant in the interval $[0{\cdot}1, 1]$ and estimated the model for each choice.

## Role of the Funding Source

The funders had no role in study design, data collection, data analysis, data interpretation, or writing of the report. All authors had full access to all data in the study and had final responsibility for the decision to submit for publication.

## 3. Results

### *3.1. Raw Deaths Counts*

We documented a substantial increase in total deaths at the beginning of the outbreak. Figure 1 shows the total daily deaths counts in the seven municipalities for 2020 and 2015-2019 (average) in the period between Jan 1st and April 4th. The year 2020 and previous years are very similar preceding the last week of February, when the first COVID-19 cases have been discovered. Starting that week, we observed a spike in the number of deaths. This spike is clearly related to COVID-19 as it starts at the beginning of the outbreak and it significantly overcomes average fluctuations in deaths observed in previous years. In total, deaths over the period from February 21st to April 4th in 2020 were almost five times the average in previous years over the same period of time (341 vs 70).

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070912; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

### *3.2. Infection Fatality Ratio Estimates*

Using our Bayesian model we estimated an overall infection fatality ratio of 1·29% (95% credible interval [CrI] 0·89−2·01). We also uncovered substantial heterogeneity by age. For under 60 years old the infection fatality rate was 0·05% (CrI 0 − 0·19, Table 2), while for over 60 years old it was 4·25% (CrI 3·01 − 6·39, Table 2).

Figure 2 shows the estimated infection fatality ratios by age group together with 95% credible intervals and interquartile range. As expected, infection fatality ratios were found much larger for older age groups. Point estimates are 4·66 and 9·04 for 71-80 and 81+, respectively. We cannot exclude, however, that the infection fatality ratio for over 80 years old is as high as 13·3% or as low as 6·61%. Interestingly, we found an infection fatality rate close to zero for under 50 years old, and around 0·1% for the 51-60 group. For robustness, we have also recalculated all point estimates using the mode and confidence intervals as the highest posterior density interval. Results remained virtually identical (data not shown).

Estimated infections rates were also heterogeneous by town, ranging between 21% and 79·5% (Table 2). Interestingly, Castiglione d'Adda, where antibody tests conducted on a sample of individuals detected a 66·6% infection rate, resulted as the municipality with the largest share of the population infected (79·51%). We estimate a population weighted overall infection rate for the seven towns of 40·5%, (CrI 25% − 58%). This is broadly consistent with a recent study on blood donors for the entire area[14] has found a 30% overall infection rate.

We finally performed an exercise to assess how sensitive our infection fatality rate estimates are to different levels of contagion (Figure 3). Focusing on a large range of potential infection rates, we found that even in the conservative assumption that only 15% of the population was infected, under 60 years old still experienced an infection fatality rate significantly below 1%, with under 40 being around 0·1%. Obviously, estimates spiked for older age groups as we set the infection rate very low. These results confirm the view that the infection has low lethality rates for younger individuals, but large rates for the elders. Moreover, this exercise showed that the overall infection fatality rate was significantly above zero and around 0·5% even in the most conservative assumption of 100% contagion.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070912; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

## 4. Discussion

In this paper we estimated the infection fatality rate of COVID-19 from administrative death counts on seven Italian municipalities that experienced the first outbreak of this desease in late February 2020. We found an overall infection fatality rate of 1·29% (CrI 0·89 – 2·01). We uncovered significant heterogeneity across age groups. Under 60 years old have infection fatality ratios around 0·05% (CrI 0·00 – 0·19). On the contrary, older people are at significantly larger risk: over-60 infection fatality ratio is 4·25% (CrI 3·01 – 6·4), and for over-80 it is estimated at 9·04% (CrI 6·61 – 13·30). Finally, we excluded very low fatality rates even under very conservative assumptions on the population infection rate: if the entire population had been infected, the overall infection fatality ratio would still be around 0·5%.

Our result for overall infection fatality ratio is larger than estimates in[20] based on travellers' data. It is, however, remarkably close to estimates for two case studies where the entire population was tested: Diamond Princess Cruise (1·3%)[21], and Vo' Euganeo (1%) - a 3,000 inhabitants municipality in the Italian Veneto region. Despite the similarity there are two main reasons to expect differences between our estimates and those in the two mentioned case studies. First, we estimate a large infection fatality ratio for over 80 years old who are likely underrepresented in the Diamond Princess Cruise. Second, total population in both Diamond Princess Cruise and Vo' Euganeo is limited and for this reason those estimates might be affected by under-sampling of the number of infected in the tails. These small differences notwithstanding, all these figures are substantially lower than estimated case fatality rates (CFRs) computed with official contagion data and that have previously been estimated for influenza pandemics[22]. The likely reason is that COVID-19 case counts are subject to downward biases due to limited testing capacity and testing strategies that prioritize symptomatic cases despite a large number of asymptomatic patients.

Our estimates could suffer from the fact that deaths data is missing for the period after April 4th 2020. This would be an issue if the contagion had not stopped by that date and therefore useful information on deaths could not be used in our model. However, as the trend in number of deaths suggests, the contagion likely stopped by April 4th in these seven municipalities. Indeed, the number of deaths in the last days of our sample went back to the average number in the previous five years. For this reason, there are reasons to believe that data for the following weeks would not be very informative on

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070912; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

the lethality of the first wave of contagion.

One limitation of our estimates is that they should not be taken at face value in the analysis of contexts where the hospital system is under stress or close to capacity. This is because our exercise was performed on a case study at the beginning of the Italian outbreak when the hospital system was still fully functioning. Moreover, the quarantine measures implemented in the seven municipalities on February 21st likely reduced contagion, potentially affecting infection fatality ratios and making them hard to extrapolate to context where similar measures were not undertaken.

Another limitation is that our model assumed a constant baseline lethality rate in absence of COVID-19. This implies that the COVID-19 outbreak did not change the baseline death rate in the population. Plausibly, the lockdown policy decreased deaths from, among others, violence and traffic, while at the same time the outbreak could have increased other fatalities due to lower availability of healthcare resources for other diseases. Fluctuations due to these causes are, however, likely to be quantitatively small compared to the large spike in deaths that we observed in 2020, where total deaths where almost five times the average in previous years over the same period of time (341 vs 70). For this reason, confounding factors in baseline deaths levels should not have substantially affected the signal to noise ratio of our death data.

In conclusion, our results support the need for isolating policies especially for the elder part of the population given the extremely high fatality rates from COVID-19 we estimated.

## Declaration of interests

GR and MP declare no competing interests.

## Data Sharing

All data is publicly available and can be found at the cited sources. Code for replication is available at `https://github.com/moritzplatz/covid_IFR_Lombardy`.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070912; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

## References

[1] Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. The Lancet Infectious Diseases. 2020;.

[2] Baud D, Qi X, Nielsen-Saines K, Musso D, Pomar L, Favre G. Real estimates of mortality following COVID-19 infection. The Lancet infectious diseases. 2020;.

[3] Kim DH, Choe YJ, Jeong JY. Understanding and Interpretation of Case Fatality Rate of Coronavirus Disease 2019. J Korean Med Sci. 2020 Mar;35(12). Publisher: The Korean Academy of Medical Sciences. Available from: http://synapse.koreamed.org/DOIx.php?id=10.3346%2Fjkms.2020.35.e137.

[4] Villa M, Myers JF, Turkheimer F. COVID-19: Recovering estimates of the infected fatality rate during an ongoing pandemic through partial data. medRxiv. 2020;Available from: https://www.medrxiv.org/content/early/2020/04/14/2020.04.10.20060764.

[5] Zhou X, Li Y, Li T, Zhang W. Follow-up of asymptomatic patients with SARS-CoV-2 infection. Clinical Microbiology and Infection. 2020;Available from: http://www.sciencedirect.com/science/article/pii/S1198743X20301695.

[6] Quilty BJ, Clifford S, group2 Cnw, Flasche S, Eggo RM. Effectiveness of airport screening at detecting travellers infected with novel coronavirus (2019-nCoV). Eurosurveillance. 2020;25(5). Available from: https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2020.25.5.2000080.

[7] Chan JFW, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet. 2020;395(10223):514–523.

[8] Chen N, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet. 2020;395(10223):507–513.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070912; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

[9] Guan WJ, et al. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med. 2020;.

[10] Huang C, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020;395(10223):497–506.

[11] Li Q, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med. 2020;.

[12] Galeotti A, Surico P. Deaths from Covid-19 - lessons from Italy and other nations. londonedu. 2020 04;Available from: https://www.london.edu/news/the-true-death-toll-of-covid-19-1771.

[13] Grasselli G, Pesenti A, Cecconi M. Critical Care Utilization for the COVID-19 Outbreak in Lombardy, Italy: Early Experience and Forecast During an Emergency Response. JAMA. 2020 03;Available from: https://doi.org/10.1001/jama.2020.4031.

[14] Il Giorno Pavia, editor. Coronavirus, Baldanti: "Test sierologico? Non parlerei di patente di immunità"; 2020. https://www.ilgiorno.it/ (accessed on 04/18/2020).

[15] La Stampa, editor. Castiglione d'Adda è un caso di studio: 'Il 70% dei donatori di sangue è positivo; 2020. https://www.lastampa.it/ (accessed on 04/15/2020).

[16] ISTAT. Dataset Analitico con i Decessi Giornalieri. https://wwwistatit/it/archivio/240401 (accessed on 04/14/2020). 2020 04;.

[17] ISTAT. Demografia in Cifre. http://demoistatit/ (accessed on 04/13/2020). 2020 04;.

[18] Manton KG, Woodbury MA, Stallard E, Riggan WB, Creason JP, Pellom AC. Empirical Bayes Procedures for Stabilizing Maps of U.S. Cancer Mortality Rates. Journal of the American Statistical Association. 1989;84(407):637–650. PMID: 12155378. Available from: https://doi.org/10.1080/01621459.1989.10478816.

[19] McElreath R. Statistical rethinking, texts in statistical science. Boca Raton: CRC Press; 2015.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070912; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

[20] Niehus R, De Salazar PM, Taylor AR, Lipsitch M. Using observational data to quantify bias of traveller-derived COVID-19 prevalence estimates in Wuhan, China. The Lancet Infectious Diseases. 2020;.

[21] Russell T, Hellewell J, Jarvis C, van Zandvoort K, Abbott S, Ratnayake R, et al. Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data from the outbreak on the Diamond Princess cruise ship, February 2020. Eurosurveillance. 2020;25:123–456.

[22] Wilson MB Nih. The emerging influenza pandemic: estimating the case fatality ratio. Eurosurveillance. 2009;25(5).

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070912; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

## List of Tables

**Table 1:** Total population and deaths by age range

| | Total Population | | | | | Total Deaths | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Age Range | 2015 | 2016 | 2017 | 2018 | 2019 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 |
| 0-20 | 8,046 | 8,039 | 8,012 | 8,035 | 8,039 | 0 | 0 | 0 | 0 | 0 | 1 |
| 21-40 | 10,202 | 10,058 | 9,851 | 9,729 | 9,711 | 0 | 0 | 1 | 0 | 0 | 0 |
| 41-50 | 7,984 | 7,832 | 7,629 | 7,461 | 7,225 | 1 | 2 | 0 | 0 | 2 | 2 |
| 51-60 | 6,922 | 7,133 | 7,284 | 7,484 | 7,599 | 3 | 2 | 5 | 2 | 5 | 5 |
| 61-70 | 5,723 | 5,704 | 5,707 | 5,680 | 5,681 | 4 | 5 | 4 | 7 | 6 | 29 |
| 71-80 | 4,733 | 4,760 | 4,843 | 4,857 | 4,903 | 13 | 10 | 18 | 17 | 10 | 107 |
| 81+ | 3,412 | 3,453 | 3,573 | 3,669 | 3,849 | 57 | 35 | 42 | 47 | 56 | 197 |
| Overall | 47,022 | 46,979 | 46,899 | 46,915 | 47,007 | 78 | 54 | 70 | 73 | 79 | 341 |

This table reports descriptive statistics by age range on total population for 2015-2019, and deaths for 2015-2020. Data refers to the period from January 1st to April 4th.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070912; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

**Table 2:** Model Estimates

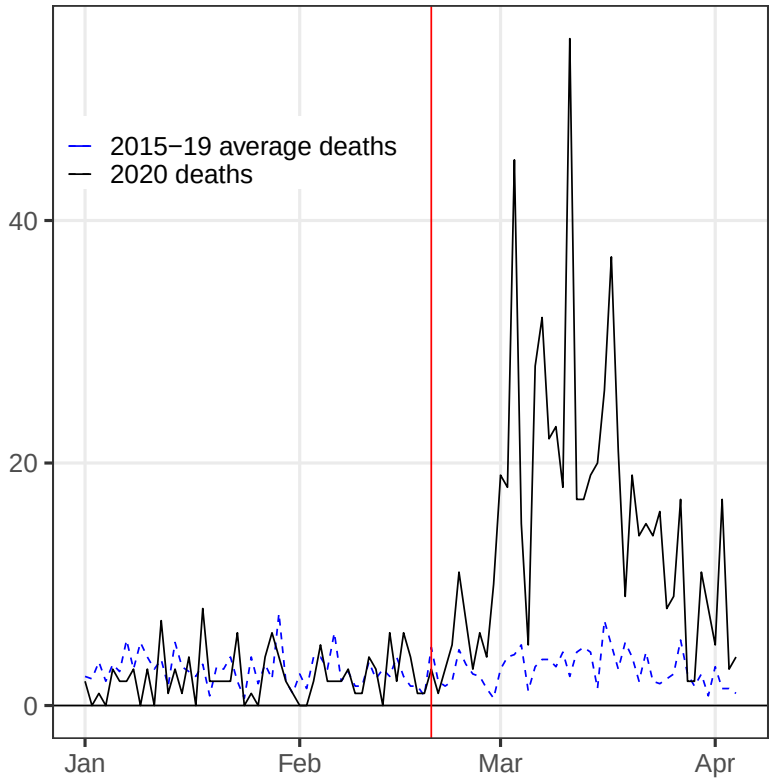
| | Estimate | CrI |
|---|---|---|
| Infection Fatality Ratios | | |
| $\delta^{Covid}_{\text{Overall}}$ | 1·2855 | [0·8878; 2·0110] |
| $\delta^{Covid}_{\leq 60}$ | 0·0524 | [0·0041; 0·1887] |
| $\delta^{Covid}_{61+}$ | 4·2509 | [3·0132; 6·3937] |
| $\delta^{Covid}_{0-20}$ | 0·0490 | [0·0048; 0·1775] |
| $\delta^{Covid}_{21-40}$ | 0·0176 | [0·0008; 0·0952] |
| $\delta^{Covid}_{41-50}$ | 0·0476 | [0·0020; 0·2007] |
| $\delta^{Covid}_{51-60}$ | 0·1076 | [0·0096; 0·3138] |
| $\delta^{Covid}_{61-70}$ | 1·0280 | [0·5876; 1·7380] |
| $\delta^{Covid}_{71-80}$ | 4·6620 | [3·3220; 6·9930] |
| $\delta^{Covid}_{81+}$ | 9·0400 | [6·6180; 13·3000] |
| Baseline Death Rates | | |
| $\delta_{0-20}$ | 0·0016 | [0·0000; 0·0085] |
| $\delta_{21-40}$ | 0·0028 | [0·0004; 0·0093] |
| $\delta_{41-50}$ | 0·0152 | [0·0060; 0·0299] |
| $\delta_{51-60}$ | 0·0452 | [0·0276; 0·0691] |
| $\delta_{61-70}$ | 0·0935 | [0·0624; 0·1331] |
| $\delta_{71-80}$ | 0·2830 | [0·2220; 0·3550] |
| $\delta_{81+}$ | 1·3290 | [1·1680; 1·5030] |
| Infection Rates | | |
| $\theta_{\text{Casalpusterlengo}}$ | 21·03 | [12·38; 32·48] |
| $\theta_{\text{Castiglione d'Adda}}$ | 79·51 | [54·23; 96·23] |
| $\theta_{\text{Codogno}}$ | 44·52 | [29·37; 62·76] |
| $\theta_{\text{Fombio}}$ | 57·54 | [31·12; 87·30] |
| $\theta_{\text{Maleo}}$ | 69·32 | [44·21; 92·53] |
| $\theta_{\text{San Fiorano}}$ | 43·64 | [20·38; 75·97] |
| $\theta_{\text{Somaglia}}$ | 21·86 | [9·02; 42·50] |

This table reports posterior median estimates from the model described in Section 2.3. For each parameter the last column reports the 95% credible interval.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070912; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

## List of Figures

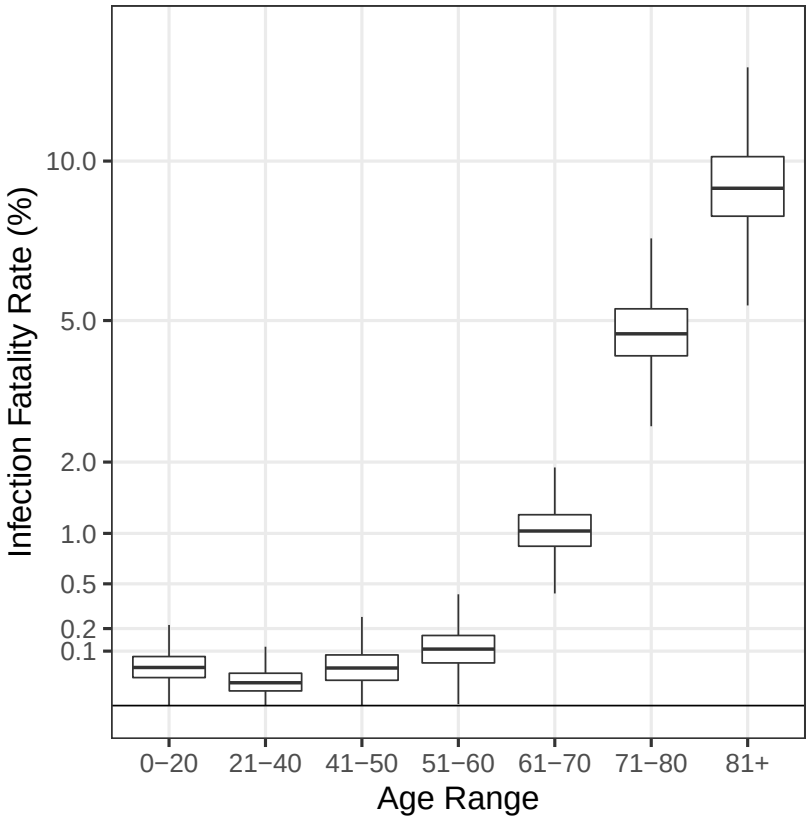
**Figure 1:** Daily deaths in 2020 and previous years average




The black solid line reports the total deaths by day in 2020 in the seven municipalities, while the blue dashed line reports the average fatalities on the same day of the previous five years. The horizontal red line marks February 20th 2020, when the first patient was admitted to the ICU in this area.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070912; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

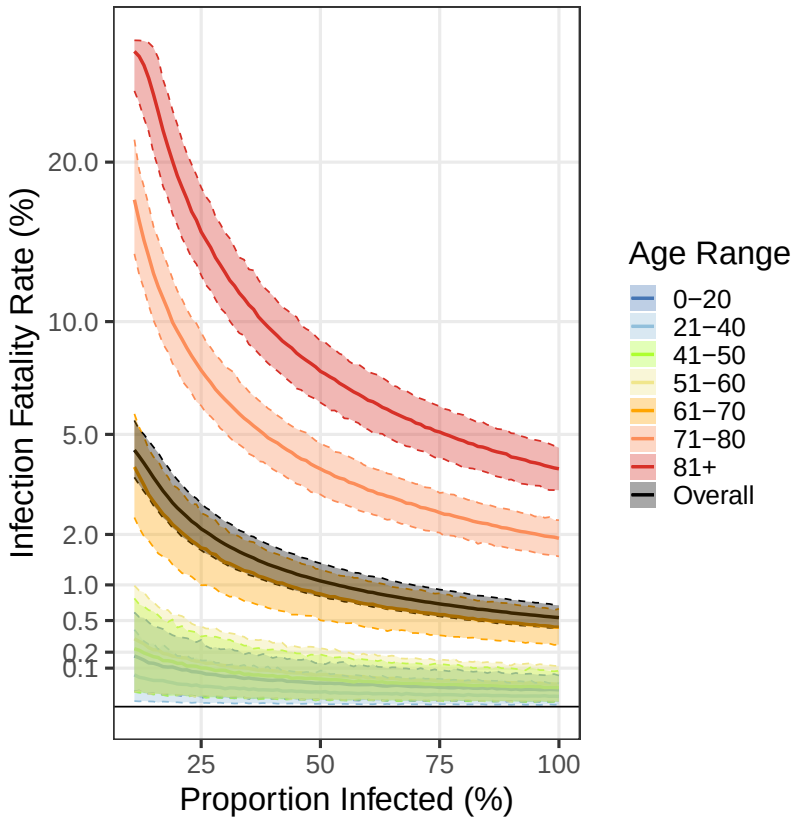
**Figure 2:** Estimates of the infection fatality rate by age range




For each age range $a$ on the horizontal axis, the figure reports information on the posterior distribution of the COVID-19 infection fatality rate $\delta_a^{Covid}$ from estimating the model described in Section 2.3. Boxes represent estimated interquartile ranges, with the central line reporting the median posterior estimate, vertical lines represent 1·5 times the interquartile range.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070912; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

**Figure 3:** Estimates of infection fatality rate by age and proportion of population infected

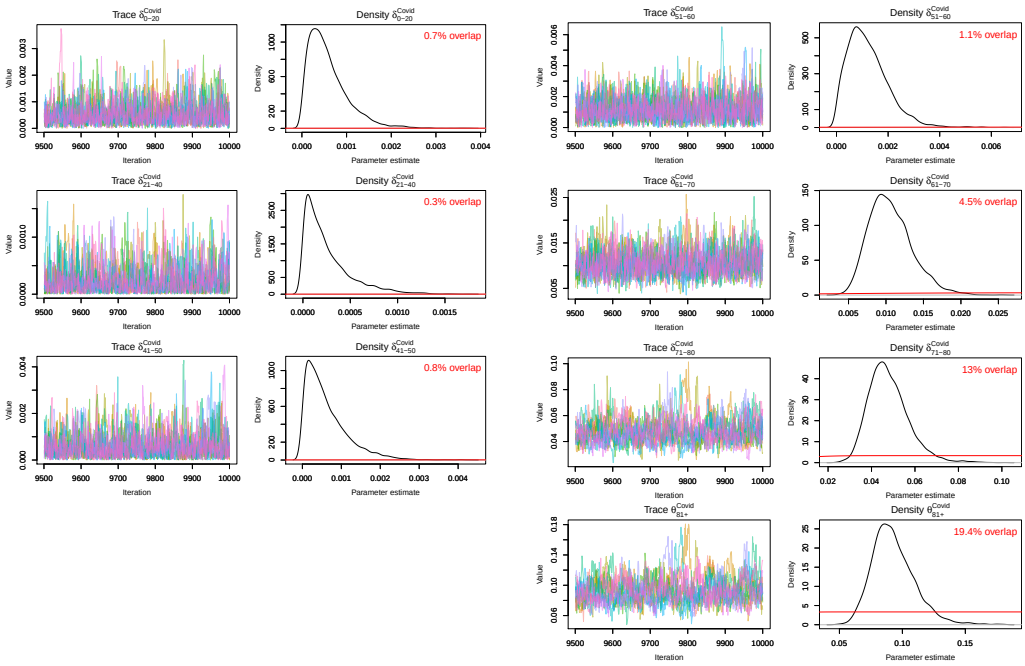



For each proportion of the population affected between 10% and 80% on the horizontal axis, we estimate a restricted version of the model described in Section 2.3 in which the infection rate $\theta_i$ is set equal for each municipality at that value. Solid lines represent the median of each posterior while the surrounding bands report 95% credible intervals from the 2·5 to the 97·5 percentiles of the posterior distributions. The Overall estimate is obtained by weighting posteriors by population shares in 2019.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070912; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

# Appendix A. Markov Chain Monte Carlo Estimation diagnostics

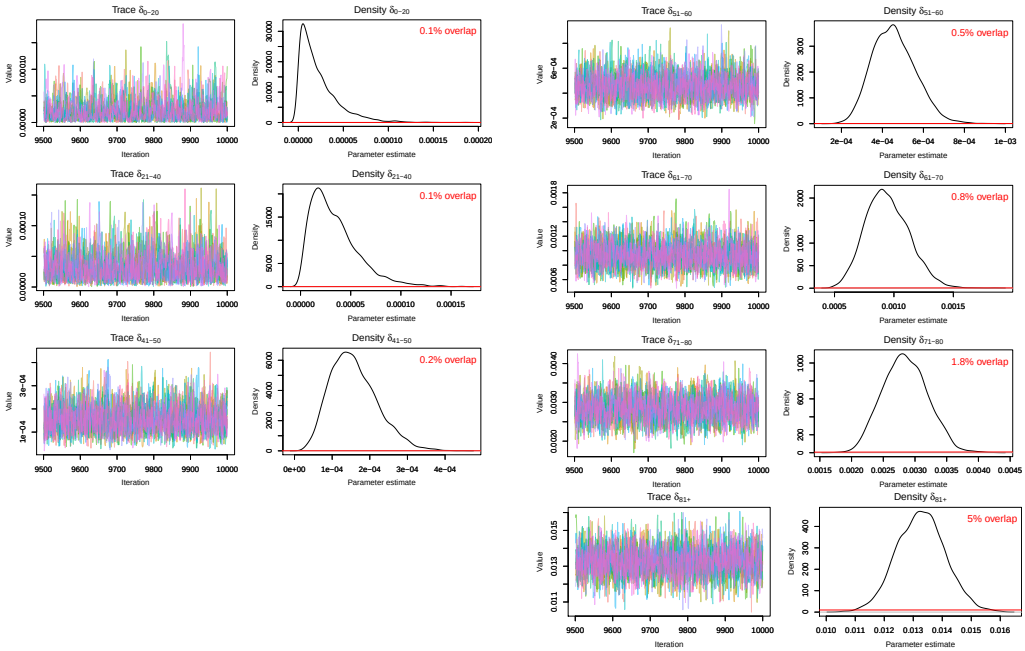
**Figure Appendix A.1:** Trace and density plots for MCMC posteriors of $\delta^{Covid}$ for each age range




For each of the seven age ranges $a$, this figure reports diagnostic plots for the MCMC simulation of the model describe in Section 2.3 for the COVID-19 infection mortality rate parameter: $\delta_a^{Codiv}$. The left panels report trace plots of the last 5000 draws from the posterior to check convergence. The right panels report the corresponding posterior distribution estimate (black solid line) together with the prior distribution for that parameter (red solid line). The % overlap reported in red is the PPO (prior posterior overlap).

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070912; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

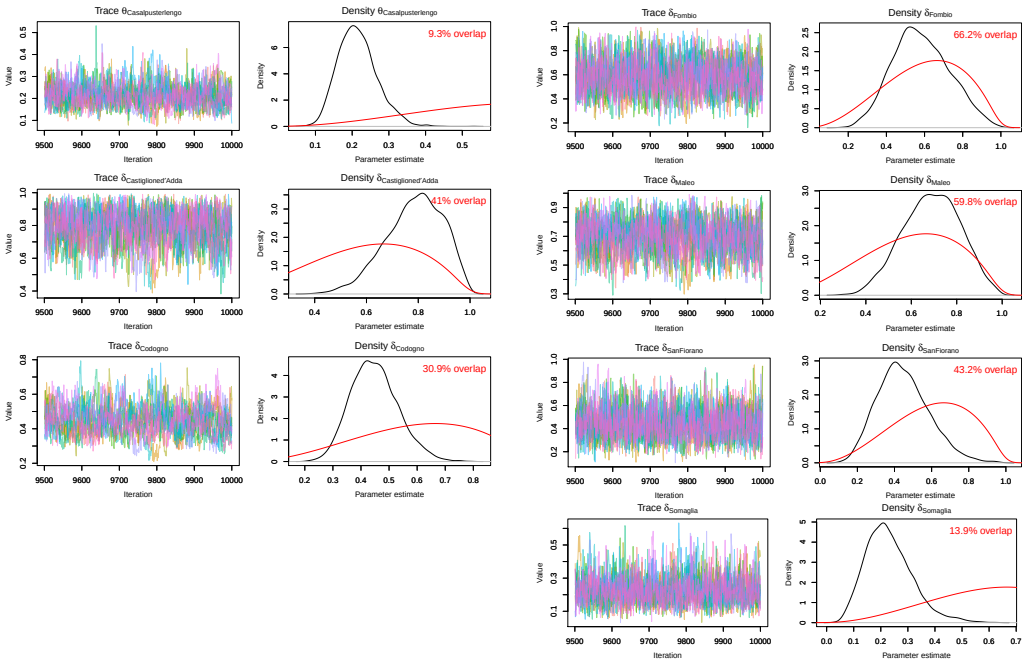
**Figure Appendix A.2:** Trace and density plots for MCMC posteriors of $\delta$ for each age range




For each of the seven age ranges $a$, this figure reports diagnostic plots for the MCMC simulation of the model describe in Section 2.3 for the baseline mortality rate parameter: $\delta_a$. The left panels report trace plots of the last 5000 draws from the posterior to check convergence. The right panels report the corresponding posterior distribution estimate (black solid line) together with the prior distribution for that parameter (red solid line). The % overlap reported in red is the PPO (prior posterior overlap).

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070912; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

**Figure Appendix A.3:** Trace and density plots for MCMC posteriors of $\theta_i$ for each municipality




For each of the seven municipalities $i$, this figure reports diagnostic plots for the MCMC simulation of the model describe in Section 2.3 on the baseline mortality rate parameter: $\theta_i$. The left panels report trace plots of the last 5000 draws from the posterior to check convergence. The right panels report the corresponding posterior distribution estimate (black solid line) together with the prior distribution for that parameter (red solid line). The % overlap reported in red is the PPO (prior posterior overlap).