How many people really die from Covid-19? Lessons from Italy and some international evidence
April 5th, 2020

As of April 2nd, Italy has recorded 13915 number of deaths from Covid-19. But how much can we trust this number? Some simple calculations based on data just made publicly available from the Italian Office for National Statistics (ISTAT) reveals a dramatically different picture.

On march 31st, ISTAT released the number of total deaths in the weeks since January 1st and up to March 21st in the Italian municipalities most affected by Covid-19, which account for about 20% of Italian population. A comparison of these data with the time series of the total number of deaths in the same weeks of the previous years (e.g. 2015-19) can be used to estimate the number of unexplained deaths in 2020. This can be confronted with the number of covid-19 deaths officially reported by the government to obtain an estimate of the extent to which the official numbers may under-report the presence of the virus in the Italian population.

So, how can one infer the actual number of deaths due to Covid-19? Key to this enterprise is to construct a counterfactual time series for 2020, namely what the number of total deaths would have been if the pattern of deaths over the weeks of 2020 was the same as the pattern observed in the very same weeks of the previous five years. The first covid-19 death was reported in Italy on February 24th, so will take this date as the starting date of our counterfactual analysis while for February 23rd we will use the actual number of total deaths. Our results are summarised in figure 1 and table 1. The solid black line in figure 1 represents the actual cumulated number of total deaths in Italy since February 23rd (the day before the first reported covid-19 death) and March 21st (last data point available from ISTAT).

The dashed black line refers to our counterfactual experiment. To construct the dashed black line, we first compute the growth rates of the cumulated number of total deaths over each pair of consecutive weeks in the years 2015 to 2019 since February 23rd and then, average each week-on-week growth rates for the same week over these previous five years. Then, we apply these historical average growth rates to the total number of death recorded on 23rd February 2020 to construct, going forward, a counterfactual time series of what the weeks of 2020 after February 23rd would have looked like if the number of total deaths in 2020 evolved as it did on average over the previous five years. This is the dashed black line in Figure 1. By construction, all lines in the chart start with the same value on February 23rd.[1]

Finally, the green line is simply the sum of the dashed black line plus the number of covid-19 death officially reported. As the graph shows, the number of officially recorded deaths in this restricted sample of most affected Italian municipalities was 4825, which is --by construction-- the gap between the green line and the dashed black line. Unfortunately, however the green line is well below the actual cumulated number of total deaths recorded by ISTAT since February 23rd. In other words, by comparing the cumulated number of deaths in Italy in 2020 with the historical evolution of the years 2015-2019 over the

1 An alternative way to construct the counterfactual would be to use the levels of the cumulative number of total deaths in the previous five years. However, those levels can be very different from year to year because of seasonal factors. For example, in 2018 the levels of cumulative deaths in France in the month of March was much higher than 2019. This was due to a particularly strong influenza that caused more deaths than in typical years. Using the levels of the cumulative number of total deaths to construct the counterfactual would place a high weight to the anomalous level of deaths during the 2018 influenza and thus would distort inference based on historical levels. In contrast, growth rates of the cumulative number of deaths are more similar across years and this is why we prefer to draw inference based on an historical comparison of growth rates. Again, for the case of France, both in 2018 and 2019 the number of daily and cumulated deaths follow a similar concave trajectory over March, with slopes flattening out around March 10th.

same weeks between February 23rd and March 21st, there are a further 4282 deaths which are unexplained, being completely unusual by any historical standard.

A plausible interpretation of these extremely unusual and unexplained further 4282 deaths is that their vast majority is directly related to Covid-19. These could be people who die home or in nursing homes and that more generally may not get tested for Covid-19. Another interpretation for a few of those fatalities is that the health system in many countries is operating at full capacity, if not over-stretched. Hence, there may be an abnormal increase in deaths with causes other than Covid-19 (e.g. heart attacks or road accidents), which nonetheless can be indirectly attributed to the epidemic. Our approach does NOT allow us to distinguish those. Rather, it gives an idea of the total death toll of the disease coming from both direct and indirect deaths.

Our analysis suggests that in Italy, over the period between February 23rd and March 21st of 2020, for every officially recorded covid-19 death, there may have been another one that went undetected. The implication is not simply that the number of covid-19 deaths may be, in fact, double relative to what officially recorded in Italy, but also that the number of infected, which is typically estimated using the inverse of the fatality rate, may also be significantly larger than previously thought.

At this point, one may ask: "how specific to Italy is this finding?" This is what applied researchers refer to as external validity. In what follows we replicate our analysis for Portugal, France and United Kingdom. We stress that the analysis of United Kingdom is very preliminary as the data are still limited. We are now moving to analyse US data, on which we hope to be able to report soon.

Portugal. The Ministry of Health releases publicly real-time data on mortality from all causes. The first registered Covid-19 death in Portugal has been reported on the 16th of March 2020. So, our analysis covers the period starting on the 16th of March until the 3rd of April. Figure 2 reports a graphical illustration of the result, and a summary of the result is in table 1. Portugal experienced 984 excess deaths relative to what implied by the growth rates of the previous years. Of those 984 deaths, only 266 (the difference between the black and the green curve) have been officially recorded as Covid-19 deaths. This is roughly a quarter of the 984 excess deaths.

France. We use the data on total deaths from INSEE and the data on Covid-19 registered data from Sante Publique France. The analysis is for the period starting on the 2nd of March (first recorded Covid-19 death) until the 23rd of March (last available observation for 2020 deaths). When we construct the counterfactual series for 2020 (dotted black line), we impute growth rates using historical data for the years 2015 to 2019, which are available from the INSEE website. Based on the historical growth rates for these years, we find that, relative to the 2020 counterfactual number of cumulated deaths, on March 23rd there are a further 2565 unexpected deaths. This is three times the official number of Covis-19 deaths, which was 860 on March 23rd. In other words, for every officially recorded death by Covid-19 in France, there may be as many as other two covid-19 related deaths that went undetected. This result is reported in figure 3 and summarized in table 1.

United Kingdom. We have repeated the same analysis for the whole United Kingdom, and then focused on the most affected area, namely London. On March 31st, the ONS has released the total number of deaths up to the week ending on March 20th for each region of the United Kingdom. We have started our analysis in the week of March 13th (the eleventh week of 2020), namely the first week after the first death was reported in the U.K. We remark that our analysis at this stage is only illustrative for the United Kingdom and London as it covers only one week, namely the week up to March 20th, when the ONS data ends. The ONS has indicated on its webpage that new data about the week after March 20th will be released on April 7th and we will update and re-evaluate our analysis as soon as this new data will become available.

Despite the few available observations and all the caveats discussed above, the findings for the U.K. and London would appear as a possible prima faciae evidence that the British count of Covid-19 deaths may be suffering a similar under-reporting issue than the one we have documented for Italy and other European countries. More specifically, while the officially recorded covid-19 deaths up to march 21st was 102 and 44, for the U.K and London respectively, our analysis suggests that the actual number of total covid-19 deaths may, in fact, have been potentially as large as 248 for the whole U.K. and 87 for London. Table 1 summarises the result for the U.K.

WARNING ON INTERPRETATIONS OF THIS ANALYSIS. It is crucial to appreciate that in the vast majority of countries the number of officially recorded covid-19 deaths provided by the government necessarily reflects mostly (if not only) deaths that occur in hospitals (and similar structures). A main reason is that those are the places where governments are concentrating their efforts and thus most tests have been so far conducted. As such, it is very hard for any government to keep track of all covid-19 deaths and thus produce in real-time an accurate aggregate estimate of the actual number of total deaths most likely associated with Covid-19. The analysis in this article and the codes that we are making publicly available at the link below develop a very simple tool that is explicitly intended to support governments in any country of the world to monitor and track the spread of the contagion, both nationally and, perhaps even more importantly, regionally and across demographic groups. Our work is not meant to replace rigorous epidemiological modelling of the number of excess deaths due to Covid-19. Rather, it aims to provide a simple, quasi-real time (subject to data releases) and transparent calculation. We welcome criticism and discussion of our approach and results. To aide in this, we are making all aspects of our analysis public.

Providing a more accurate number for the total deaths associated with Covid-19 is important because all economic models quantifying the costs of the recession curve during a pandemic rely on an estimate of the replication number from epidemiology models. But the replication number in any epidemiology model crucially depends on the fatality rate, which in turn is a function of the number of deaths. We hope that the tools provided in this article, notwithstanding all the caveats we have highlighted, could play their small role to feed governments' and researchers' calculations on the intensity and spread of the contagion curve in their own country and therefore help to evaluate the costs of the recession curve.

Andrea Galeotti, London Business School, agaleotti@london.edu
Sebastian Hohmann, Wheeler Institute, sebas.hohmann@gmail.com
Paolo Surico, London Business School, psurico@london.edu
An updated version of A user-guide to Covid 19

Additional contributors
Luis Fonseca (London Business School) for the analysis of Portugal
Riccardo Trezzi for the analysis of Italy

Additional information

This project aims at estimating Covid-19 mortality from official statistics on total deaths in 2020 and their growth rates in the previous years over the same period. We aim to constantly update the analysis as new data becomes available. We also aim to add robustness checks and alternative methods to accurately determine Covid-19 deaths. Hence, this article will be updated as often as feasible.

We would like to replicate the analysis for as many countries as possible. If you have data or you know where we can access data in a country, please share them with us. We are happy to replicate the analysis, share the results with you, and update the analysis so that other people can access this information.

If you have the data and you want to perform the analysis yourself, in what follows you will find all material you need. However, we would appreciate if you share with us the results, with a summary of how you conducted the analysis. In particular, with explicit reference to changes in the codes or methodology. This is to assure that results are comparable across countries.

You can access the relevant documentation: https://github.com/sebastianhohmann/covid19_total_death

Table and graphs

TABLE 1. DETECTED AND UNDETECTED COVID-19 DEATHS

COUNTRY (sample period)	Total excess deaths (black minus dashed line)	Covid-19 official deaths (green minus dashed line)	Unexplained deaths (black minus green line)
ITALY (23rd February to 21st March)	9107	4825	4282
PORTUGAL (16th March to 3rd April)	984	266	718
FRANCE (2nd March to 23rd March)	2565	860	1705
UNITED KINGDOM 13th March to 20th March	248	102	146

Figure 1: Italy

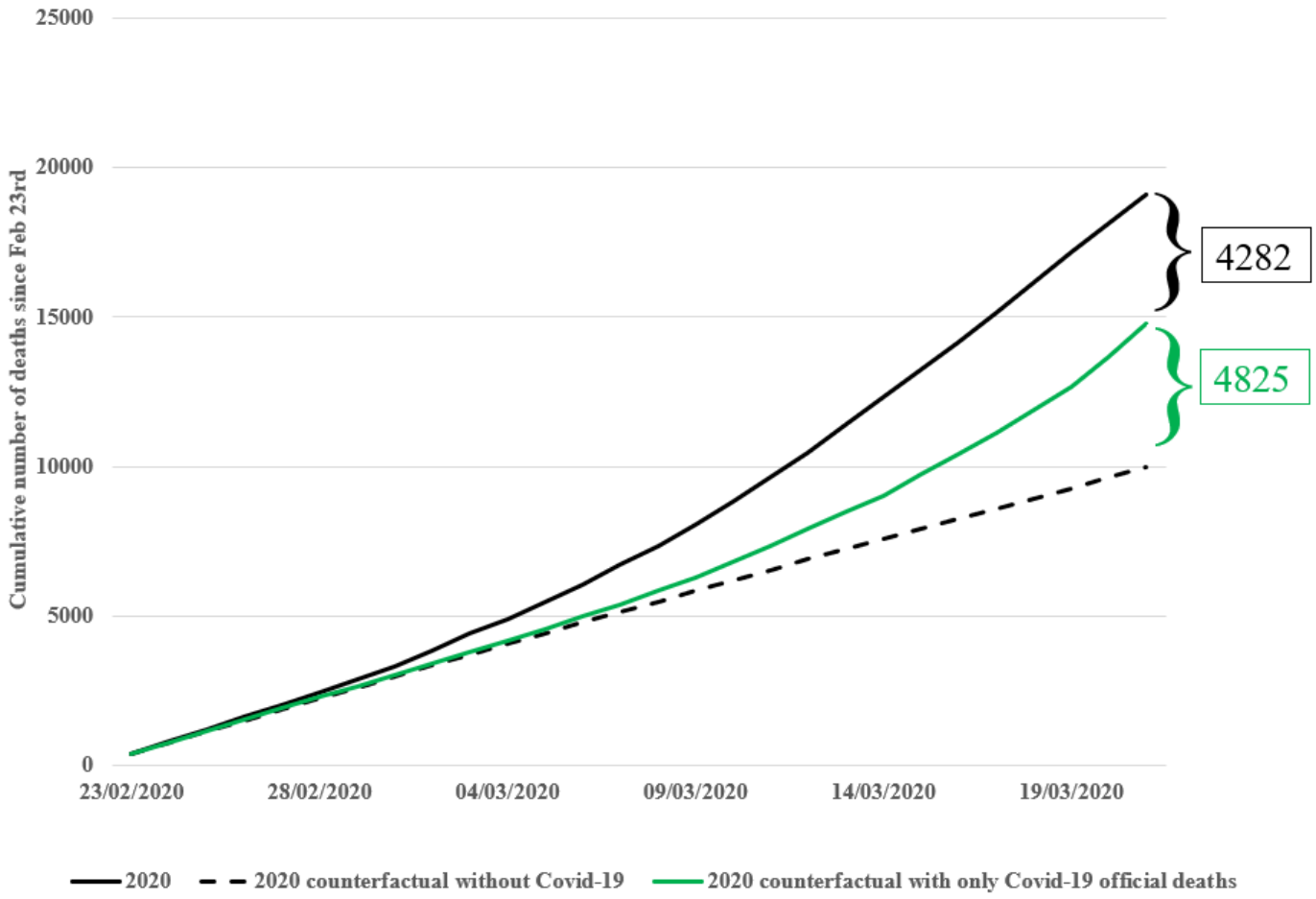

Figure 2: Portugal

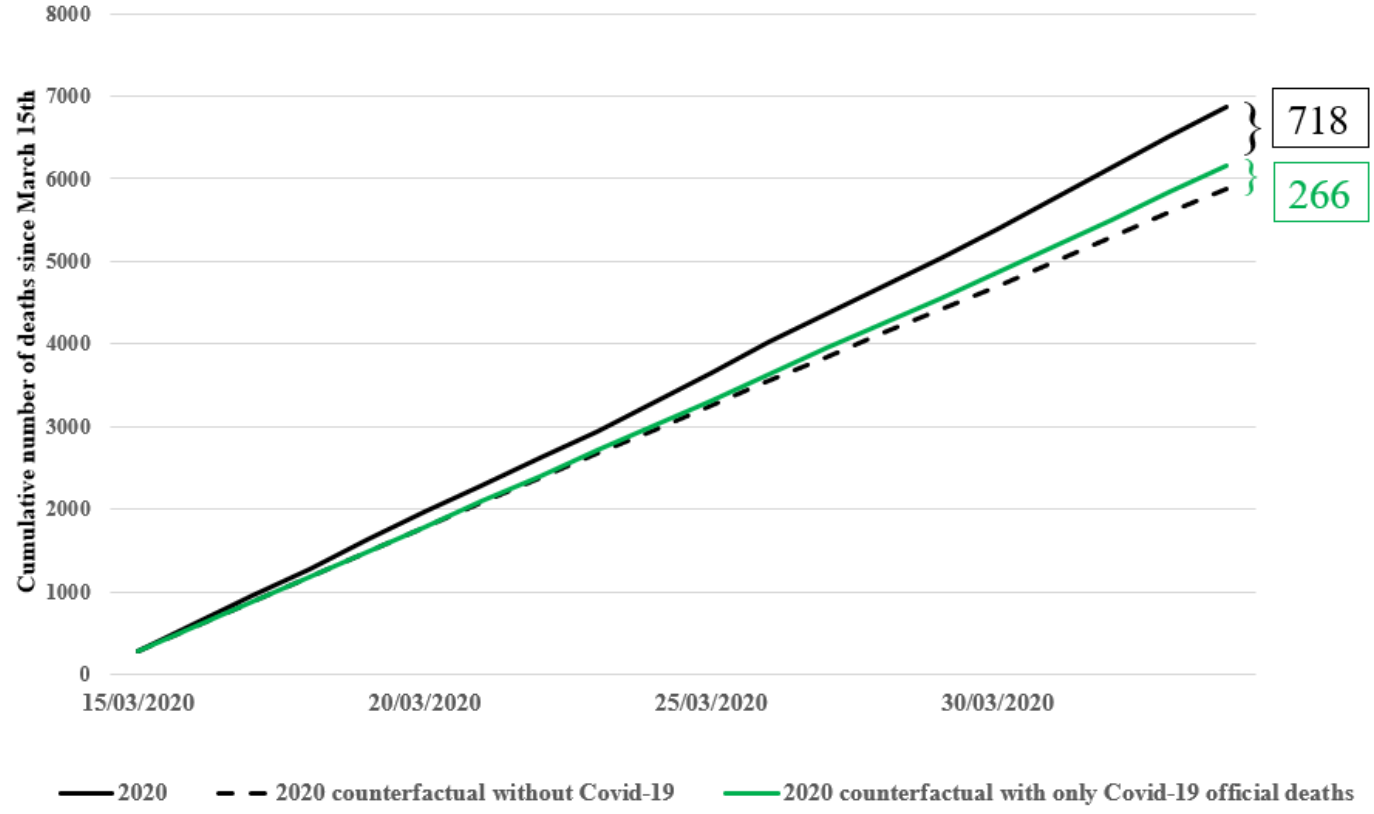

Figure 3: France