Federico Ricci-Tersenghi

Follow   13 Followers   About

An estimate of direct and indirect deaths related to the COVID-19 epidemic in Italy.

Federico Ricci-Tersenghi · May 1, 2020 · 25 min read

We often hear that the number of victims of the SARS-CoV-2 epidemic in Italy is much higher than the figures provided daily by the Civil Protection.

Furthermore, we read more and more news related to deaths not directly due to the virus but indirectly due to the epidemic. In essence, they are often people with serious pathologies, that have nothing to do with COVID-19, but who have difficulty or reluctance to timely enter in hospitals, sometimes dying for that reason.

Unfortunately, to date, accurate estimates of these two numbers, i.e., the deaths due directly and indirectly to the virus, are not available.

In this note we try to make up for this shortcoming, providing a reliable estimate of these deaths through the statistical analysis of the data made available by the Italian National Statistical Service (Istat) on the total deaths that occurred in Italy until April 4, 2020. We also develop different scenarios starting from the observation of the lower rate of women dying of COVID -19.

by Enrico Bucci (1), Luca Leuzzi (2), Enzo Marinari (3), Giorgio Parisi (3) and Federico Ricci-Tersenghi (3)
(1) Sbarro Institute — Temple University
(2) Nanotech — CNR & Sapienza University
(3) Sapienza University & INFN & Nanotech — CNR

many Italian regions. We limit ourselves to those with coverage greater than 50% of the population (Lombardy, Liguria and Emilia-Romagna). The use of these data to estimate the number of deaths in poorly sampled regions can lead to severe errors.

- The excess of deaths with respect to a year without the epidemic shows a temporal trend clearly linked to the epidemic, growing from the last week of February and showing a peak around March 20.

- A comparison of the Istat-derived excess deaths with the data reported by the Italian Civil Protection Department for the deaths attributed to the COVID epidemic shows that the latter is hugely underestimated in the regions most affected by the epidemic (about 7,000 fewer deaths in Lombardy and about 1,000 in Emilia-Romagna). In the most affected provinces (e.g., Bergamo) we estimate that the real number of deaths is more than twice that reported by the Civil Protection and more than five times the deaths without the COVID-19 epidemics.

- A comparison of the temporal evolution of the deaths estimated from Istat data and those provided by the Civil Protection suggests that the latter are not only underestimated but also somehow time-delayed. This delay may significantly modify the time dependence of the COVID deaths, with consequences on the estimation of epidemiological parameters.

- COVID-19 is well known to display a larger fatality rate for men than women dying in hospitals. This gender imbalance, contrasting the natural gender balance for all-cause deaths, allows us to estimate the possible percentage of fatalities that occurred indirectly for the epidemic, that is, not directly caused by the virus, but by other factors connected to the virus outbreak.

Introduction

The SARS-CoV-2 coronavirus epidemic spreading in Italy in the first months of 2020 produced a high number of deaths certified as positive for the virus (more than 27,000 to date). These mainly include hospital deaths; the real number, however, is likely to be much higher. In this note, we try to provide a first estimate of the actual number up to April 4, 2020.

Every day the Italian Civil Protection Department (DPC) [1] provides a report with the

these numbers (in particular, the number of deaths) are strongly underestimated [2]. Recently Istat has made available the number of total deaths occurred up to April 4, 2020, in 1,689 Italian municipalities [3]. By comparing these data with those of previous years from the same towns, we can observe that in many cities in the most affected regions the number of deaths is significantly higher than the seasonal average. The difference is also much higher than the number of deaths certified by the Civil Protection as due to COVID-19.

Several authors have tried to make a statistical analysis of the mortality data provided by Istat [4] inferring the actual number of deaths and their possible causes. However, most of these studies have some limitations: either do not infer properly the total number of deaths or do not split the excess of deaths between those due directly to the coronavirus and those due to 'collateral effects' of the epidemic outbreak.

At present, the most comprehensive study analysing the Istat mortality data is by Modi et al. [5]. Nonetheless, the authors of this study do not consider properly the fact that Istat provided data with *bias* and they obtain the total number of deaths by just "scaling up by the completeness factor": this kind of analysis provides too large estimates, that we are going to correct by properly estimating the number of deaths not provided by Istat.

Moreover, the scaling-up strategy is extremely noisy and risky when applied to regions that have a small coverage by Istat data. The authors of Ref. [5] apply the scaling-up strategy also to regions like Sardegna, Toscana and Marche where the Istat data coverage is smaller or close to 30%: in these cases the scaling up from 30% to total, without taking into account that the sampled fraction of the population is the most affected by the epidemic, may produce a wrong result. An example making this evident is the case of Marche region: here the only province which is well sampled is 'Pesaro e Urbino' because this is the province most affected by the epidemics in Marche region. Thus extending to the entire region the mortality rate of that province produces a total number of deaths which is too large.

Finally, the authors of Ref. [5] "attribute all the excess deaths to COVID-19 fatalities", which we show to be very unlikely exploiting the gender imbalance in deaths due to COVID-19.

deaths resulting from undetected infections (typically many of the deaths occurred at home), and to 'collateral' deaths, that is deaths due to the heavy stress of the health system. Patients with other serious pathologies in some cases had difficulty accessing hospitals. The average time for ambulance interventions had significantly increased. The visits to check patients with other pathologies, even severe ones, have been partially suspended. Accessing the emergency room became more complex. Due to these facts, sometimes even the health services attempted to persuade the patients not to contact the hospitals unless in severe conditions, which are, however, difficult to evaluate for those without medical knowledge. Knowledge of the stress state of the sanitary service, moreover, can be identified as a psychological reason that prevents non-COVID patients from going to the emergency room, or even to their general health practitioner.

The deaths recorded by Istat from February 22 to April 4

The reference period of our analysis extends from February 22nd, 2020 (date of the first death in Italy officially attributed to COVID-19) to April 4th, 2020 (last day of the period covered by the data made available by Istat at the time of writing this contribution). One should keep in mind that the data provided by Istat contain all the deaths that occurred in the reference period, that we divide for convenience into three categories:

- **base** deaths are those that would happen under normal conditions, i.e. in the absence of the epidemic;

- the deaths caused **directly** by COVID-19 (part of these are those officially **certified** and forming the statistics provided by the Civil Protection);

- deaths caused **indirectly** by the epidemic, that is, not caused by the COVID-19 virus, but which would not have occurred in normal conditions and which are presumably the consequence of the critical conditions in which the health system had to operate in some regions.

Of course, the confinement also led to other changes in total deaths: probably more domestic fatalities and less on the roads and at work. However, all these categories of deaths have minimal impact on the totals (respectively about 20, 10 and 3 daily deaths throughout Italy). So on a first approximation, we neglect them.

categories in which we divide the deaths.

DEATHS - ITALY 22/2-4/4 2020

BASE	EXCESS		
	COVID-19 CERTIFIED	COVID-19 NON-CERTIFIED	INDIRECT

The data on deaths provided recently by Istat allow us to obtain a statistically solid estimate of excess deaths in the reference period. Unfortunately, the sample of Istat data is partial and with a *bias*, because it represents a subset of 1,689 municipalities, distributed unevenly throughout Italy, which meet the criteria set out in the Istat methodological note [6]: at least ten deaths in the first months of 2020 and an increase in mortality in March 2020 compared to the 2015–2019 average higher than 20%. These criteria induce some *bias*, which must be taken into account when processing the data. According to some scholars, this bias introduced by Istat would not allow a significant estimate of the real number of deaths [7]. Still, as we show in this note, this estimate can be obtained using the right precautions and techniques discussed in detail in the methodological note at the end.

REGION	Total population 1/1/2019	ANPR municip. population 1/1/2019	% ANPR municip. population	% sampled municip. population
Lombardia	10060000	8369992	83,2%	72,7%
Liguria	1550640	1296389	83,6%	68,1%
Emilia-Romagna	4459477	4006532	89,8%	59,4%
Valle d'Aosta	125666	125666	100,0%	42,6%
Veneto	4905854	4067466	82,9%	39,2%
Piemonte	4356406	3980929	91,4%	38,8%

Toscana	3729641	3273677	87,8%	31,3%
Umbria	882015	648933	73,6%	22,3%
Sardegna	1639591	1313788	80,1%	21,8%
Trentino-Alto Adige	1072276	345927	32,3%	21,6%
Puglia	4029053	2537516	63,0%	18,4%
Abruzzo	1311580	806477	61,5%	12,7%
Basilicata	562869	131863	23,4%	9,5%
Friuli-Venezia Giulia	1215220	406104	33,4%	8,5%
Molise	305617	221846	72,6%	7,5%
Sicilia	4999891	2943125	58,9%	7,0%
Calabria	1947131	748489	38,4%	5,0%
Campania	5801692	2943680	50,7%	4,3%
Lazio	5879082	1846069	31,4%	3,3%

Table 1. Istat data coverage by region.

Our estimates of deaths are based on a selection of the 1,689 municipalities sampled by Istat.

First, we have selected only provinces and regions where the sampling made available by Istat constitutes a broad coverage of the total. The municipalities monitored by Istat are a subset of those that have joined the ANPR (National Resident Population Registry), whose coverage of the territory in terms of population is shown in *table 1* for the regions and in *table 2* for some provinces.

For each region and the best-sampled provinces, we report in the tables the percentage

municipalities).

PROVINCE	Total population 1/1/2019	ANPR municip. population 1/1/2019	% ANPR municip. population	% sampled municip. population
Parma	451631	417867	92,5%	90,7%
Milano	3250315	2987063	91,9%	87,7%
Massa-Carrara	194878	174000	89,3%	86,2%
Lodi	230198	219193	95,2%	83,7%
Rimini	339017	317981	93,8%	80,2%
Genova	841180	794967	94,5%	79,3%
Cremona	358955	317078	88,3%	78,7%
Piacenza	287152	245035	85,3%	76,0%
Monza e della Brianza	873935	755593	86,5%	75,1%
Bergamo	1114590	863498	77,5%	74,6%
Mantova	412292	334719	81,2%	73,3%
La Spezia	219556	173113	78,8%	71,2%
Brescia	1265954	942364	74,4%	71,2%
Biella	175585	161389	91,9%	70,4%

| Modena | 705393 | 576696 | 81,8% | 69,1% |
| Pesaro e Urbino | 358886 | 296094 | 82,5% | 67,8% |

Table 2. Istat data coverage by province.

The best estimates of deaths are obtained in Lombardy and Emilia-Romagna.
To reduce the bias due to the sampling of the data provided by Istat and the statistical uncertainties in our estimates, we prefer to focus mainly on two of the three regions that have a coverage of more than 50% in the sampled municipalities, namely Lombardy and Emilia-Romagna. Since these are the regions most affected by the epidemic, we believe that our analysis is extremely significant, although restricted to these regions. Within these selected regions, we then focus only on the provinces with the highest coverage, shown in *table 2*, where we show data for the 17 provinces with population coverage greater than 65%.

Note that some regions are particularly poorly represented in the data provided by Istat (for example Lazio — the region of Rome — and Campania have coverage below 5%): using these data to try to evaluate the real impact of the epidemic in those regions would produce statistically poor results.

Results on the excess of deaths

We refer to the methodological note at the end for a detailed explanation of the procedure we have used to estimate the number of total deaths in the reference period of the year 2020.

In *figure 1* we show the comparison between the number of deaths occurred in 2016 in Lombardy and those taking place in 2020 as we estimate it from the Istat data. The choice of the year 2016 as the reference year comes from the observation that it is the one that best approximates the data for the year 2020 in January and February (among the 2015–19 five-year period made available by Istat): it is, therefore, a valid baseline for the following months. In the methodological note at the end, we show that other choices for the reference baseline curve lead to the same conclusions.
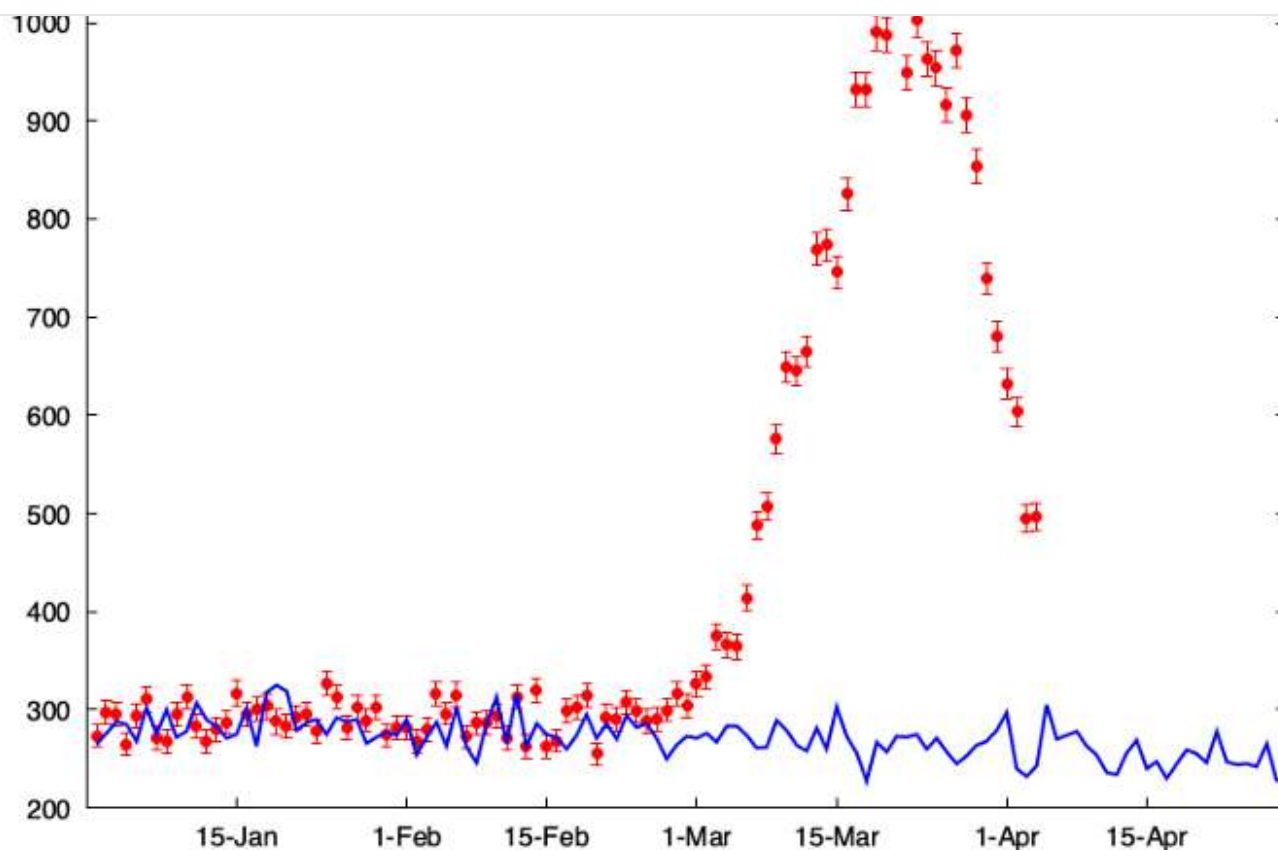
Lombardy

Figure 1. Total deaths in Lombardy day by day: comparison between those that occurred in 2016 (blue curve) and those that we estimate to have occurred in 2020 (red data with error).

The data in *figure 1* clearly show the effects of the epidemic attack that has caused victims in Lombardy since the last week of February. They also tell us that **the peak in the total number of deaths was around March 20**. Furthermore, all deaths above the blue reference curve are excess deaths due to the ongoing COVID-19 epidemic. We will try later to estimate how many are directly due to the virus attack and how many are indirectly due to the virus outbreak. As of now, we just compare the excess deaths with those officially certified as positive for COVID-19 and provided daily by the Department of Civil Protection.
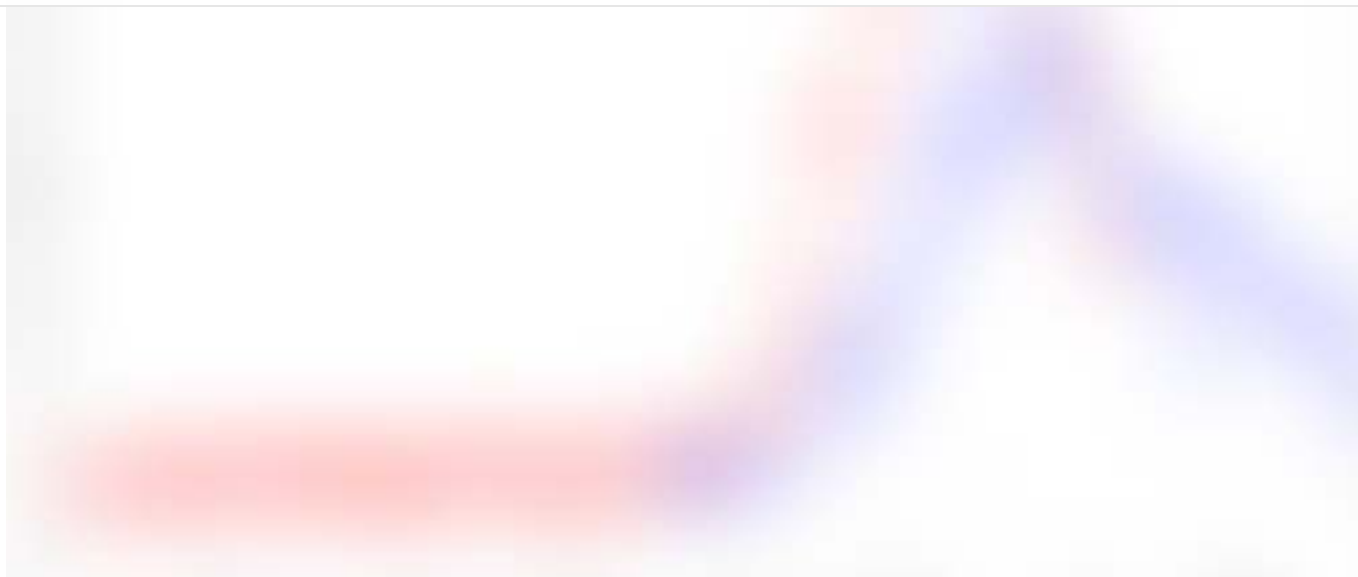
Figure 2. The number of deaths in excess estimated for Lombardy by Istat data (red points with error) and the number of certified deaths positive to COVID-19 provided by the Civil Protection (blue points).

In *figure 2,* we show this comparison for Lombardy, and we deduce that the official data collected and communicated by the Civil Protection are probably subject to some systematic error. The most evident is that the **official number of deaths due to COVID-19 is hugely underestimated** unless we assume that the difference between the two curves in *figure 2* is entirely due to indirect deaths (later we discuss this in detail). In the first days of April, the red curve (total deaths) goes below the blue curve (COVID certified deaths). This is impossible, proving that at least one of the two curves does not follow the real data. Unfortunately, we have indications that both datasets may differ appreciably from reality.

Positive certified deaths from COVID-19 are probably reported in the database managed by the Civil Protection with some delay (for example the Piedmont region has repeatedly announced that the fatalities communicated in a given day had not all occurred in the previous 24 hours). Under the hypothesis that the death counts enter the Civil Protection database with some variable delay, the peak in the real number of deaths from COVID-19 would become much broader. We remind that there is an ongoing debate on why the daily deaths in many Italian regions are almost constant for several weeks, that is a sort of *plateau* is being observed instead of a sharp peak, and this may be a possible explanation.

Moreover, we also notice an anomaly in the data provided by Istat: a cross-comparison

Istat are systematically lower in the last days of the period considered. This may induce the effect of a too fast decrease after the peak. We, therefore, need to wait for an update of the Istat data to confirm the scenario of the rapid decrease that we are currently seeing.

The analysis that we have shown in *figures 1 and 2* for Lombardy can be carried out only for provinces where the number of daily deaths is large enough. For example, in *figure 3* we show the total number of deaths estimated for the province of Bergamo, the most affected in Italy.

Figure 3. Total deaths in the Province of Bergamo.

Extending deaths estimates to more regions and provinces

Willing to study a larger number of regions and provinces, we prefer to consider just the total number of deaths in the entire epidemic period, instead of their time evolution as

take into consideration all the deaths that happened from that date onwards. The results for the three regions and the 17 provinces with a large enough coverage are reported in *table 3* and *figure 4*. We show the *baseline* deaths with a green bar, the *excess* ones with a red bar, and the *official* COVID-19 positive deaths with a blue bar.

Figure 4. Graphical representation of the data in table 3. Total deaths are divided into the `base` (green) and the `excess` (red). Of the latter, only a part is certified as due to COVID-19 (blue).

In all the cases that we analyzed, the deaths officially attributed to COVID-19 are only a fraction of the excess deaths in the period examined. The value of this fraction

Bergamo, where the real number of excess deaths is more than **twice the official ones**, and almost **five times** as big as they would have been without the epidemic.

Can we extend the estimate of deaths in excess at the national level?
Before answering this question, we should look at *table 1* again*,* and notice that most regions are not sufficiently represented by the Istat sample to allow a correct estimate. Precisely for this reason, it is more prudent to consider only the excess of the deaths we estimated from Istat data in the regions with high coverage. For the others, one may want to consider the data reported by the DPC. In this way we obtain that on April 4, 2020, a very reasonable estimate of excess deaths due to the epidemic (directly and indirectly) was close to 25,000, that is **10,000 more than the official numbers** reported by DPC. At the time this manuscript is written (April 30, 2020), the estimated excess number of deaths is larger than 36,000 units.

This analysis confirms that, like the number of people infected, the number of deaths caused by COVID-19 is largely underestimated. This result is consistent with the fact that in the most affected areas many symptomatic people never had access to the necessary hospital care during the disease and died at home without even having a swab made to test their positivity to the virus.

How could COVID-uncertified deaths be related to the epidemic?
It is natural to ask whether all the excess deaths uncertified for being positive to the virus are, however, directly ascribable to COVID-19 syndrome. A sanitary service in crisis jeopardizes the health and survival of all patients, subject to any disease, significantly reduces emergency assistance, suspends the regular monitoring of even severe health conditions. It is very probable that, in a limited period and in specific areas of Italy, which were dramatically attacked by the epidemic, many people, in the absence of immediate care or assistance, did not survive. Add also the fact that hospitals have become hubs for the spreading of the virus and many people have delayed their access to the emergency room for fear of infection: in this way, further deaths occurred, deaths

purely psychological reasons, if citizens with serious non-COVID diseases refuse assistance from the hospital structures. There are signs that this phenomenon may happen, and it is undoubtedly necessary to take action to prevent it.

Gender imbalance in COVID deaths may help in estimating the mortality related to COVID or other causes.

The gender imbalance in hospital lethality of the COVID-19 coronavirus is well documented at all levels, both nationally and internationally. For example, in China [8] the fraction of women among certified positive virus deaths, which we will call f C, it is equal to 36.2%, while in Spain [9] it is equal to 36.6% and in Italy [10] it is equal to 35.8%. In many other countries [11], including France and Germany, this fraction is close to 40%. Significant exceptions are Belgium, Canada, South Korea and others, with certified female death rates close to 48–49%. It is important to clarify that we speak here of certified deaths and therefore mainly of hospital fatalities. In Lombardy, on the other hand, this gender imbalance seems to be even more severe in the deaths certified by the DPC and is equal to 30.6% on April 5, 2020. The gender imbalance remains large in the various age groups, as shown in *figure 5*.

Figure 5. Gender imbalance in certified positive COVID-19 deaths in Lombardy by age group (data as of April 5, 2020).

a proportion of about one to one. Of course, to do this, we need a further hypothesis, on the gender imbalance in COVID deaths occurring outside hospitals: it is not at all obvious that the two percentages are equal. We will then analyze four scenarios, where among the COVID deaths (which therefore include both hospital certified and non-certified, but virus-positive home deaths) the percentage of female deaths is respectively 30% (average in Lombardy), 36% (average in Italy), 40% (average in some European countries) and 47% (average in Istat data on all deaths).

In practice, we have two types of deaths which are distinguished by the mortality ratio between men and women.
A first type is 'non-COVID' deaths, in which women appear proportionally to how many are in the total population: we call fNC the fraction of female deaths in this typology, and we note that it varies slightly from area to area, remaining around the average value of 52% (in the analysis we consider the actual fNC value measured in a given area in a non-epidemic time).
A second type is 'COVID' deaths, in which women are a fraction fC. The four scenarios outlined above correspond to fC= 0.3, 0.36, 0.4 and 0.47).
At this point, knowing the ratio between male and female deaths in all excess deaths (those estimated by Istat data), it is immediate to calculate how many of these deaths are COVID and non-COVID using the formula:

fraction of non-COVID deaths = (x -fC) / (fNC -fC)

where x is the fraction of female deaths among all excess deaths. Note that the uncertainty in estimating the fraction of non-COVID deaths becomes very large when the fractions fC and fNC approach each other. Therefore, the stronger the gender imbalance in deaths due to COVID-19, the more precise the estimate of the percentage of indirect deaths.
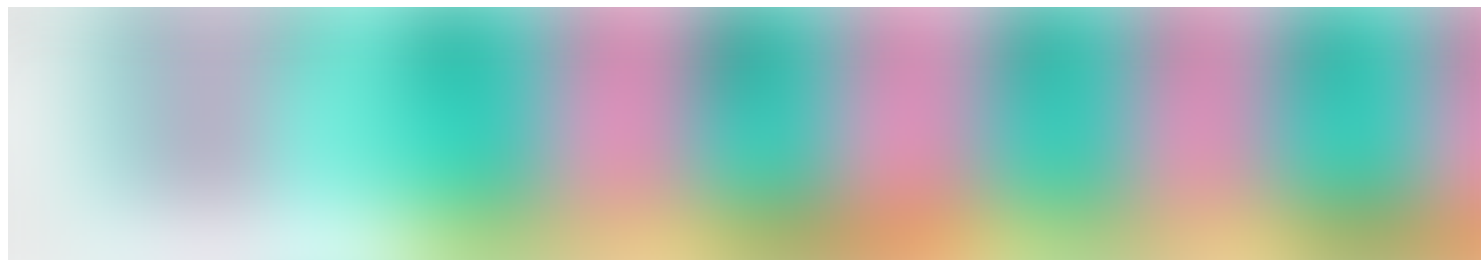
Table 4. Partition of the excess deaths between the 'non-certified COVID' and 'non-COVID' categories for the statistically most significant regions and provinces. The numerical ranges shown in the table correspond to a standard deviation. If the lower limit of the range is negative, the null value is highly probable.

In *table 4* we consider **4 different scenarios** in which we vary the percentage *fC* of women who died from the virus. We report the distribution of excess deaths in two categories: uncertified COVID-19 deaths and indirect deaths. Their accurate reading can give us valuable information on how many deaths have occurred due to the virus, but have not been recognized as such (non-certified COVID deaths) and instead how many collateral damages of the epidemic (non-COVID deaths) occurred for the causes discussed earlier.

- **The first scenario in which the percentage of women's deaths is 30%** can only make sense for Lombardy and the Lombardy provinces where this percentage was actually measured in certified deaths due to COVID-19.
  If this scenario turns out to be the correct one, then in Lombardy there has been a number of collateral deaths equal to about half of those due directly to the virus (about 5,000 non-COVID deaths against about 10,000 COVID deaths). This should make us reflect deeply on how much the organization of the emergency and the maintenance of essential services allows to considerably reduce the impact of an epidemic.

- **The second and third scenarios in which the percentage of female deaths due to COVID-19 is 36% or 40%** are perhaps the most plausible since these percentages are those measured in many different countries. The conclusions reached in these scenarios is that the epidemic caused both a high number of indirect deaths and a high number of deaths due to COVID-19 not certified as such. The percentage of

and tells us that **about a third of deaths are 'collateral damage' to the epidemic**. The percentage of uncertified COVID deaths compared to the total deaths due to the virus varies in Lombardy between 15% and 30%, while in Emilia-Romagna it varies between 10% and 22%. This result also seems reasonable, based on the observation that in the areas most affected by the epidemic, COVID deaths not certified with a swab are in a higher proportion. We believe these are very important numbers that need to be validated in future studies.

- **The fourth scenario (47% of female deaths)** implies that most of the deaths are directly due to the virus. We observe that by setting fC to 47% of the percentage of female deaths among those due to the infection, we obtain a number of non-COVID excess deaths that are always compatible with zero (except for the provinces of Cremona and Parma). In this scenario, we obviously have a vast number of deaths due to COVID-19 that have not been certified. Unfortunately, in this scenario, the estimates have a high uncertainty since *fNC-fC* is small. We note that this situation would imply a picture in which the deaths of men and women for COVID-19 are much more balanced than those in hospitals.

We would like to stress that the most reliable predictions are those for the provinces of Parma (Emilia-Romagna) and Milan (Lombardy), which have very high coverage rates (90.7% and 87.7% respectively).

Conclusions

Our main message is that **the imbalance between male and female deaths due to COVID-19 is a crucial piece of information**, which can significantly help us improve our understanding of the development of the epidemic.

We have evidence for an imbalance between male and female COVID-19 fatalities and, thanks with this, a signal for the presence of many deaths due to non-COVID-19 diseases and the crisis of the health system.

Applying our analysis scheme to more accurate data, which we hope will be available soon, will allow us to investigate the real distribution of deaths, by gender and age. Moreover, we could accurately quantify the lethality of the disease and, together, of the health crisis connected to the epidemic spread. Hopefully, more accurate data will also

The directions to proceed with this type of study are many. For example, the analysis that considers the different age groups separately is of undoubted interest. We note that the ratio between male and female deaths in Lombardy and Germany is very similar under 60 years. Still, increasing the age, the difference grows. In Lombardy, the ratio is higher than in Germany, as if in Lombardy a component equally distributed between the two sexes was missing. This is an interesting phenomenon that should be studied in-depth, making comparisons also with other countries.

Acknowledgements We thank Diego Alberici for providing us with the data of the deaths for the provinces of Emilia-Romagna and our friends. They promptly reported an important error in one of the tables in a preliminary version of the text.

Methodological note

Unfortunately, the data provided by Istat do not refer to all Italian municipalities, but a subset of them that we call *sampled municipalities* and which satisfy the following three conditions:

1. having taken over by 31/12/2019 in the ANPR (National Registry of Resident Population): this is 5295 out of a total of 7913 Italian municipalities;

2. having sent the registry changes in time;

3. having registered a number of deaths from 1/1/2020 to 4/4/2020 not less than 10 and an increase in mortality of at least 20% starting from March (i.e., in the period 1/3/2020–4/4/2020) compared to the corresponding average for the five years 2015–2019.

To ensure continuity in the analysis, the municipalities that had entered this list in the first survey (up to 28/3) did not come out in the second survey (relative to 4/4) even if they have lost some of these characteristics.
In the attempt to provide an unbiased estimate of the number of excess deaths over the

criteria.

We assume that condition 1 is unrelated to the number of excess deaths and we call *F1* the fraction of the population in the municipalities that satisfy condition 1. We also call *D1* the number of deaths in the reference period in these municipalities.

Condition 2 is probably negatively correlated with excess deaths since a municipality severely affected by the epidemic could have more problems updating its data. Therefore assuming that all the municipalities have transmitted the data, we are ignoring this correlation, and the estimate of the number of excess deaths we obtain is probably smaller than the actual one. By calling *F2* the population fraction in the municipalities that satisfy the first two conditions, we are, thus, assuming *F2 = F1* and *D2 = D1* where *D2* is the number of deaths in these municipalities.

Condition 3, on the other hand, is positively correlated with excess deaths and would provide an estimate higher than their true value if we did not correctly take into account the bias with which Istat selected a sampled municipality. We call *F3* the percentage of the population in the sampled municipalities, and we approximate *F3* with the fraction of deaths that occurred in these municipalities in the 2015–19 five-year period (on a large scale the deaths are certainly proportional to the population). We also call *D3* the number of deaths in these municipalities in the reference period: this number can be directly read off from the data provided by Istat.

The total number of deaths

The best estimate of the total number of deaths that occurred in the reference period in the year 2020 is obtained by adding to the deaths in the sampled municipalities, *D3*, an estimate of the deaths that occurred in the ANPR municipalities for which Istat did not provide the numbers, equal to *D1-D3*. Normalizing for the coverage that the ANPR (*F1*) municipalities provide on the totality of all Italian municipalities we get:

D = (D3 + (D1-D3)) / F1

The number of deaths can always be obtained by multiplying the mortality times the population and therefore to estimate the deaths in the non-sampled municipalities we can write:

where *P1* and *P3* are, respectively, the populations of the municipalities that satisfy only condition 1 (ANPR municipalities) and those that also satisfy condition 3 (sampled municipalities), *Mnot3* is the mortality in the municipalities that DO NOT satisfy condition 3. At the same time, *M* is the average mortality in the absence of an epidemic and *B* is the baseline number of deaths, i.e., the total deaths in a period without the epidemic (we discuss two different choices for this baseline value below). The mortality *Mnot3* of the municipalities that do not satisfy condition 3 is a bit smaller than the average mortality *M*. The ratio *Mnot3/M* depends in principle on the year: we use the one measured in the five years 2015–19, where the data are complete, as an approximation for that in the year 2020. We believe that this provides a very reasonable estimate because the municipalities that do not meet condition 3 are those least affected by the epidemic (therefore with minor variations between the five years 2015–19 and the year 2020). Furthermore, to obtain a good lower bound rather than the exact number of deaths, it is enough to assume that even in the municipalities that do not meet condition 3, the deaths in 2020 are not less than those of the previous years (the baseline). Under this hypothesis, the formula we use for the calculation of *D* provides a very reasonable estimate, and a good lower bound, to the real number of deaths.

The definition of the baseline

The so-called 'baseline value' *B,* appearing in the formula for estimating total deaths *D*, is the average number of deaths in times without the epidemic.
The same baseline value is also needed to compute the excess deaths from the total deaths. The baseline value can be estimated in different ways that are worth discussing.

The winter of 2019–20, up to the outbreak of the epidemic, was unusually mild and there were fewer deaths due to the flu epidemic than the average of the previous five years. So using the 2015–19 five-year average as a baseline value fo the year 2020 would be a bad choice.
In a first approach, we, therefore, defined the baseline value as the average of deaths in the years 2015–2019 suitably rescaled by a multiplication factor that made this average coincide with the deaths that occurred in 2020 before the epidemic (i.e., in the period 1/1–22/2). Alternatively, we observed that among the five previous years, 2016 was also unusually mild in terms of the number of deaths, with a trend very similar to that of

Comparison between estimates of deaths obtained with different baseline values.

As we can see in the table for the Lombardy region, the most unfortunate and emblematic case of the current epidemic crisis in Italy, changing the definition of the baseline value, the estimates of total deaths are very similar, and compatible with each other within the uncertainty intervals reported, corresponding to one standard deviation. The estimated numbers of excess deaths (once the baseline is taken out) are still similar, although they deviate by more than one standard deviation. For further prudence, we used the second choice in the discussion of the results, the one in which the baseline is the 2016 value, which leads to an average excess smaller than 4%.

A check of the data provided by Istat

Given the hurry in which Istat processed the data, we believe that a check on their consistency is useful. We can carry out this check by crossing the data provided by Istat with those presented by the Daily Mortality Surveillance System (SISMG) in its weekly report.
We have chosen to do it for four cities that are among the municipalities sampled by Istat showing a clear deviation from the baseline: Milan, Brescia, Bologna and Genoa.
In the following plots, we have superimposed on the graph extracted from the last SISMG report [12] the data provided by Istat (red dots).
In principle for a single municipality sampled by Istat the value indicated by the red dots should be exact since it does not require any extrapolation.
Nonetheless, we observe that the consistency of the two data sets is never perfect: for some cities, such as Bologna, the trend is qualitatively similar, while for others the last points differ considerably.
We would like to highlight the case of Milan in which the number of Istat deaths seems to display a significant decrease in the last week. This bahevior does not appear at all in

This is the discrepancy that leads us to be careful before being too optimistic on the Istat data: the sharp drop after the peak could be in part due to the underestimated numbers in the last days provided by Istat.

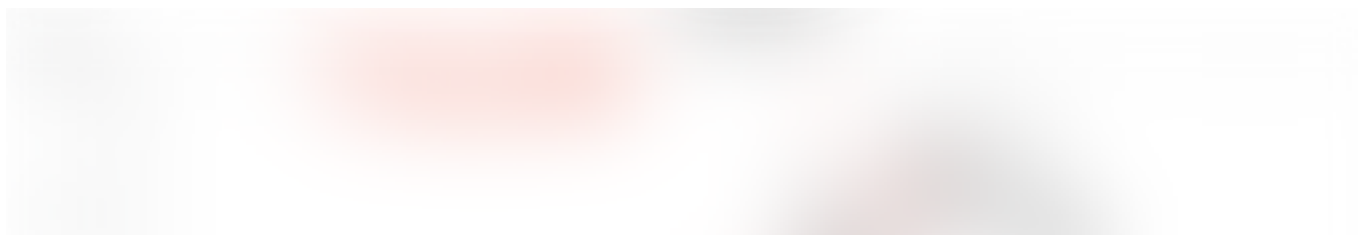Comparison between Istat and SISMG mortality data for Genoa.
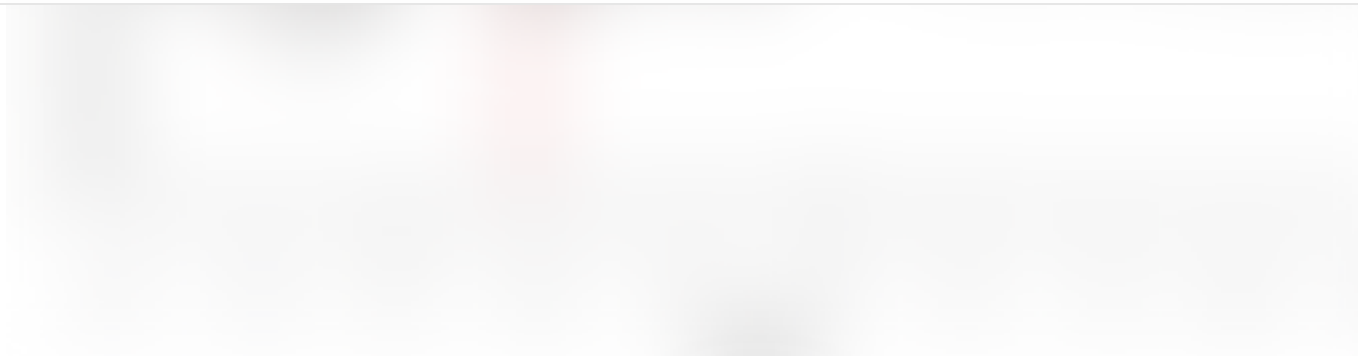
Comparison between Istat and SISMG mortality data for Bologna.

Comparison between Istat and SISMG mortality data for Brescia.

Comparison between Istat and SISMG mortality data for Milan.

References

[1] http: //www.protezionecivile.gov.it/attivita-rischi/rischio-sanitario/eme …

[2] https://www.scienzainrete.it/icle/epidemiologia-dei-necrologi/luca -c … https: //www.scienzainrete.it/ articolo/confermato-eccesso-di-mortalit%C3% …

[3] https://www.istat.it/it/archivio/240401

[4] https: //naturalstupidity.ghost.io/cosa-ci-dicono-i-dati-istat-in-piu-ris … https://medium.com/@pmeridian/le-morti-correlate-al-covid-19-eed78ee421e7 https://www.ft.com/content/6bd88b7d-3386-4543-b2e9-0d5c6fac846c

[5] https://www.medrxiv.org/content/10.1101/2020.04.15.20067074v2

[6] https: //www.istat.it/it/files//2020/03/Il-punto-sui-decessi_al_16-aprile …

[7] https://www.lavoce.info/archives/65171/ coronavirus-dead-stones-on-the -…

[8]http: //weekly.chinacdc.cn/en/article/id/e53946e2-c6c4–41e9–9a9b-fea8db1a …

[9] https://covid19.isciii.es/

[10] https: // www. epicentro.iss.it/coronavirus/sars-cov-2-decessi-italia

[11] https://globalhealth5050.org/covid19/

[12] http://www.deplazio.net/images/stories/SISMG/SISMG_COVID19 .pdf