Ann Ist Super Sanità 2022 | Vol. 58, No. 1: 25-33
DOI: 10.4415/ANN_22_01_04

A population-based cohort approach to assess excess mortality due to the spread of COVID-19 in Italy, January-May 2020

Maria Dorrucci[1], Giada Minelli[2], Stefano Boros[1], Valerio Manno[2], Sabrina Prati[3], Marco Battaglini[3], Gianni Corsetti[3], Xanthi Andrianou[1], Flavia Riccardo[1], Massimo Fabiani[1], Maria Fenicia Vescio[1], Matteo Spuri[1], Alberto Mateo-Urdiales[1,4], Martina Del Manso[1,4], Patrizio Pezzotti[1], Antonino Bella[1] and the Italian Integrated Surveillance COVID-19 Group[*]

[1]Dipartimento Malattie Infettive, Istituto Superiore di Sanità, Rome, Italy
[2]Servizio di Statistica, Istituto Superiore di Sanità, Rome, Italy
[3]Servizio Registro della Popolazione, Statistiche Demografiche e Condizioni di Vita, Istituto Nazionale di Statistica, Rome, Italy
[4]European Programme for Intervention Epidemiology Training (EPIET), European Centre for Disease Prevention and Control (ECDC), Stockholm, Sweden
**The members of the Italian Integrated Surveillance COVID-19 Group are listed before the References.*

Abstract
Aims. To assess the impact of the COVID-19 pandemic on all-cause mortality in Italy during the first wave of the epidemic, taking into consideration the geographical heterogeneity of the spread of COVID-19.
Methods. This study is a retrospective, population-based cohort study using national statistics throughout Italy. Survival analysis was applied to data aggregated by day of death, age groups, sex, and Italian administrative units (107 provinces). We applied Cox models to estimate the relative hazards (RH) of excess mortality, comparing all-cause deaths in 2020 with the expected deaths from all causes in the same time period. The RH of excess deaths was estimated in areas with a high, moderate, and low spread of COVID-19. We reported the estimate also restricting the analysis to the period of March-April 2020 (first peak of the epidemic).
Results. The study population consisted of 57,204,501 individuals living in Italy as of January 1, 2020. The number of excess deaths was 36,445, which accounts for 13.4% of excess mortalities from all causes during January-May 2020 (i.e., RH = 1.134; 95% confidence interval (CI): 1.129-1.140). In the macro-area with a relatively higher spread of COVID-19 (i.e., incidence rate, IR): 450-1,610 cases per 100,000 residents), the RH of excess deaths was 1.375 (95% CI: 1.364-1.386). In the area with a relatively moderate spread of COVID-19 (i.e., IR: 150-449 cases) it was 1.049 (95% CI: 1.038-1.060). In the area with a relatively lower spread of COVID-19 (i.e., IR: 30-149 cases), it was 0.967 (95% CI: 0.959-0.976). Between March and April (peak months of the first wave of the epidemic in Italy), we estimated an excess mortality from all causes of 43.5%. The RH of all-cause mortality for increments of 500 cases per 100,000 residents was 1.352 (95% CI: 1.346-1.359), corresponding to an increase of about 35%.
Conclusions. Our analysis, making use of a population-based cohort model, estimated all-cause excess mortality in Italy taking account of both time period and of COVID-19 geographical spread. The study highlights the importance of a temporal/geographic framework in analyzing the risk of COVID-19-epidemy related mortality.

Key words
- COVID-19 epidemic
- spread
- mortality
- excess mortality
- cohort
- surveillance
- Italy

Address for correspondence: Maria Dorrucci, Dipartimento Malattie Infettive, Istituto Superiore di Sanità, Viale Regina Elena 299, 00161 Rome, Italy. E-mail: maria.dorrucci@iss.it.

INTRODUCTION

Italy was the first European country to be hit by the COVID-19 pandemic. The first confirmed case of COVID-19 was reported on February 20, 2020 [1], and the first related death occurred in February; it was a man in his 70s who suffered from pre-existing cardiovascular disease. Since then and until the end of May 2020, Italy experienced a strong epidemic wave that, particularly in the months of March and April, put the national health services under pressure, especially in the regions with the highest cases of COVID-19 [2]. Excess mortality, defined as the difference between the total and expected mortalities, is one of the main epidemiological tools for the measurement of the effects of a pandemic [3].

In Italy [4, 5], as in the rest of Europe [6-8], numerous studies have already been conducted with the aim of estimating excess mortality caused by COVID-19. Studies that compared the estimates among European countries [8] demonstrated that during the first wave of the epidemic, Italy was among the countries that had the highest number of excess deaths. The differences in the estimates between countries depended, as already observed [8], on the different characteristics of the populations. However, studies that estimate excess mortality in relation to the geographical spread of the disease are scarce.

The data on the first wave in Italy, due to the high geographical heterogeneity of the COVID-19 incidence rates throughout the national territory [1], can help estimate excess mortality in relation to the geographical spread of the disease. On these grounds, we conducted an analysis of the overall mortality based on the data supplied by the National Institute of Statistics (Istituto Nazionale di Statistica, ISTAT) and the Italian National Integrated COVID-19 Surveillance System, managed by the National Institute of Health (Istituto Superiore di Sanità, ISS). This study aimed to describe all-cause excess mortality from January to May 2020, integrating all existing information on a national basis. In particular, we focused on the quantification of the excess of deaths from all causes in Italy in relation to the spread of COVID-19.

METHODS

Study population

The study population consisted of Italians living in Italy at the start of the study period, i.e., the resident population released by the Italian National Institute of Statistics on January 1, 2020. (ISTAT; http://dati.istat.it). In the current analysis, the resident population was considered by Italian administrative units (107 provinces), age (from 0-9 years to 90+ years by intervals of 10 years), and sex.

Deaths from all causes. ISTAT releases estimates per day of the number of deaths from all causes by age, sex, and municipality of residence drawing from population and taxpayers' registries (National Register of Resident Population and the information provided by the municipalities themselves). For the present analysis of the year 2020, deaths from all causes referred to 7,357 municipalities out of the 7,904 existing ones (for 93% of Italian municipalities corresponding to 95% of all Italian residents); for further details, see published report [9]. For the purpose of our study, the dataset of deaths from all causes was aggregated by day, provinces (107 provinces), age, and sex.

COVID-19 surveillance and deaths in COVID-19 cases. COVID-19 cases are microbiologically (SARS-CoV-2-positive nasopharyngeal swabs) diagnosed and collected daily by the Italian Regions through a dedicated web platform (Italian National Surveillance System for COVID-19). This surveillance system is monitored and updated daily by the ISS as per the case definition published and regularly updated online by the European Centre for Disease Prevention and Control (ECDC) [10-12]. Clinical and epidemiological information (e.g., age, sex, municipality/province of residence, comorbidity, vital status, and date of death) are also collected.

For the present analysis, we extracted from the dataset all cases and deaths (SARS-CoV-2 infections confirmed by RT-PCR) between January 1 and May 31, 2020, by province of residence (107 provinces), age classes (from 0-9 years to 90+ years, aggregated over a 10-year interval), sex, and day of death.

Primary outcomes: i) observed number of deaths from all causes per day in 2020; ii) expected number of deaths from all causes in the same day (estimated as the average number of deaths from all causes occurring daily in the previous 5 years, i.e., 2015-2019); iii) observed excess number of deaths per day in 2020, calculated as the difference of the first two outcomes: i)-ii).

Secondary outcomes: i) observed number of deaths "with a COVID-19 diagnosis" per day in 2020; ii) number of deaths "without a COVID-19 diagnosis" per day in 2020 (calculated as the difference between the total number of deaths from all causes per day in 2020 and the number of deaths "with a COVID-19 diagnosis" per day); and iii) incidence of COVID-19 cases per day (incidence rates × 100,000).

Study design

Retrospective cohort study.

Statistical analyses

All datasets were aggregated and merged by day, age group, sex, and province of residence. The count-time aggregate data were converted to frequency-weighted survival-time data. We considered in detail the survival times going from January 1, 2020, to the end of the study period, which was May 31, 2020, or to the date of death if the person died before the end of the study period.

We obtained interpolating curves for the outcomes per day (all-cause deaths, expected deaths, and difference between all-cause deaths and deaths with a COVID-19 diagnosis) using the locally estimated scatterplot smoothing method (LOESS) [13]. The LOESS method was further used as a supplementary analysis to compare the main characteristics (age and sex) of the outcomes.

To estimate the cumulative incidences of the outcomes, we employed the Kaplan-Meier method [14]. In addition, we applied Cox models [14] on aggregated data, as aforementioned, to estimate the relative hazards (RHs) of deaths from all causes in 2020 compared with the expected number of deaths from all causes in the same year, i.e., the RHs of excess deaths. The same model was used to estimate the RHs of deaths without a COVID-19 diagnosis, defined above, in 2020 compared with the expected deaths from all causes in the same year. The analyses were also repeated separately by geographical areas. For this purpose, we considered different levels of COVID-19 diffusion according to the tertiles of the COVID-19 incidence rates per 100,000 residents, and we classified provinces accordingly. In detail, we assumed "low spread of COVID-19" when the incidence rates by province were less than the first tertile, "moderate spread of COVID-19" when the rates ranged from the first to the second tertile, and "high spread" when the COVID-19 rates were over the third tertile.

Finally, we restricted the analysis to the period of March-May 2020 (first peak of the epidemic) to estimate again the RH excess mortality from all causes and calculate the RH of excess deaths for COVID-19 incidence rate increments.

Data were analyzed using the SAS software version 9.4.

RESULTS

Demographic characteristics of the cohort and the COVID-19 diagnoses

In total, 57,204,501 individuals were included in the analysis (*Supplementary Figure 1* available online and demographic description in *Table 1*). The geographical distribution of the excluded population (5%) was similar to that of the included population: of the excluded individuals, 41% lived in the north, 22% in the center, and 37% in the south; of those included, 44% lived in the north, 21% in the center, and 35% in the south.

The age and sex distributions of the cohort (*Table 1*) were similar to those of the whole Italian population (*Table 1*): young people accounted for less than 20% (people younger than 20 = 10,176,681, accounting for 17.8%), and people older than 60 accounted for almost 30% (25,513,208, accounting for 29.7%), with a male-to-female ratio of about 0.95.

Most of the deaths from all causes (*Table 1*) occurred in the age group of 80-89 years (121,055; 39.3%) and in females (158,613; 51.5%). The age distribution of COVID-19-related deaths was similar to that of all-cause deaths in the Italian population, except for individuals aged 70-79 years for whom the percentage was higher (26.7% *vs* 19.7%), as well as among individuals aged 90+ years for whom, however, the percentage was lower (17.6% *vs* 26.5%); with regard to gender, the majority of COVID-19 deaths were males (18,842; 58.5%).

Table 2 (part a) presents a brief description of the COVID-19 cases confirmed by the Italian National Surveillance System on COVID-19. The majority of cases of COVID-19 was concentrated in 36 provinces (33.6%), where the National Surveillance System reported during the studied period an incidence rate of at least 450 COVID-19 cases per 100,000 residents (high diffusion area). Conversely, lower spread of the disease was observed in the remaining provinces (*Table 2, part a,* low and moderate spread of COVID-19). It should be noted that the distribution by sex and age of COVID-19 cases was quite similar in the three COVID-19 diffusion macro-areas (*Table 2, part b*). In addition, it should be noted that among individuals aged 80+ years, the highest percentage was reported in the area with high incidence, especially in comparison with the estimate reported in low spread area (27.1% in the area with high incidence *vs* 14.1% in the area with low incidence).

Temporal patterns of the main outcomes according to the geographical spread of COVID-19

Overall, the ISTAT reported 307,809 deaths from all causes in Italy from January 1, 2020, to May 31, 2020

Table 1
Population description, corresponding to 94.9% of all residents in Italy at January 1st 2020; deaths occurred from Jan 1st to May 31 2020

Demographic features	Groups	Total cohort	Deaths for any causes	COVID-19 related death
Totals		**57,204,501**	**307,809**	**32,236**
		frequency (%)	**frequency (%)**	**frequency (%)**
Age	0-9	4,702,210 (8.2%)	564 (0.2%)	3 (<1%)
	10-19	5,474,471 (9.6%)	263 (0.1%)	0 (0%)
	20-29	5,846,300 (10.2%)	594 (0.2%)	14 (<1%)
	30-39	6,601,399 (11.5%)	1,248 (0.4%)	64 (0.2%)
	40-49	8,586,396 (15.0%)	4,178 (1.4%)	285 (0.9%)
	50-59	9,021,150 (15.8%)	12,080 (3.9%)	1,129 (3.5%)
	60-69	7,053,315 (12.3%)	25,850 (8.4%)	3,282 (10.2%)
	70-79	5,703,289 (10.1%)	60,537 (19.7%)	8,594 (26.7%)
	80-89	3,460,872 (6.0%)	121,055 (39.3%)	13,179 (40.9%)
	≥90	755,099 (1.3%)	81,440 (26.5%)	5,686 (17.6%)
Sex	males	27,858,592 (48.7%)	149,196 (48.5%)	18,842 (58.5%)
	females	29,345,909 (51.3%)	158,613 (51.5%)	13,394 (41.5%)

Table 2 (part a)
Description of COVID-19 cases occurred in Italy from Jan 1st to May 31 2020

Demographic features	
Total	**223,937**
	frequency (col %)
Age groups	
0-9	1,818 (0.8%)
10-19	3,364 (1.5%)
20-29	12,391 (5.5%)
30-39	17,248 (7.7%)
40-49	29,046 (13.0%)
50-59	40,208 (18.0%)
60-69	29,948 (13.4%)
70-79	32,243 (14.4%)
80-89	39,460 (17.6%)
≥90	18,211 (8.1%)
Sex	
males	102,640 (45.8%)
females	121,297 (54.2%)
Spread level of COVID-19*	
low (30-149 cases per 100,000)	18,517 (8.3%)
moderate (150-449 cases per 100,000)	42,323 (18.9%)
high (450-1,610 cases per 100,000)	163,097 (72.8%)

Table 2 (part b)
Description of COVID-19 cases by COVID-19 spread

Demographic features	**Low spread**	**Moderate spread**	**High spread**
Total	**18,517 (8.3%)**	**42,323 (18.9%)**	**163,097 (72.8%)**
	freq. (col. %)	**freq. (col. %)**	**freq. (col. %)**
Age groups			
0-9	315 (1.7%)	452 (1.1%)	1,051 (0.6%)
10-19	641 (3.4%)	903 (2.1%)	1,820 (1.1%)
20-29	1,646 (8.9%)	2,918 (7.0%)	7,827 (4.8%)
30-39	2,004 (10.9%)	3,579 (8.5%)	11,665 (7.1%)
40-49	2,726 (14.7%)	5,830 (13.8%)	20,490 (12.7%)
50-59	3,665 (20.0%)	8,129 (19.2%)	28,414 (17.4%)
60-69	2,721 (14.6%)	5,512 (13.0%)	21,715 (13.3%)
70-79	2,172 (11.7%)	5,311 (12.5%)	24,760 (15.2%)
80-89	1,933 (10.4%)	6,455 (15.2%)	31,072 (19.0%)
≥90	694 (3.7%)	3,234 (7.6%)	14,283 (8.8%)
Sex			
males	9,527 (51.5%)	18,673 (44.1%)	74,440 (45.6%)
females	8,990 (48.5%)	23,650 (55.9%)	88,657 (54.4%)

*Levels were obtained on the basis of the tertiles of the distribution of the incidence rates of COVID-19 by province of residence.
Note: the Italian provinces with *low incidence* were: Agrigento, Ascoli Piceno, Bari, Barletta, Benevento, Cagliari, Caltanissetta, Caserta, Catania, Catanzaro, Cosenza, Crotone, Enna, Foggia, Frosinone, Isernia, L'Aquila, Latina, Lecce, Livorno, Matera, Messina, Napoli, Nuoro, Oristano, Palermo, Potenza, Ragusa Reggio Calabria, Roma, Salerno, Siracusa, Sud Sardegna, Taranto, Trapani, Vibo Valentia, Viterbo; *moderate incidence*: Ancona, Arezzo, Brindisi, Campobasso, Chieti, Cuneo, Fermo, Ferrara, Firenze, Forlì-Cesena, Gorizia, Grosseto, La Spezia, Lucca, Macerata, Padova, Perugia, Pisa, Pistoia, Pordenone, Prato, Ravenna, Rieti, Rovigo, Sassari, Siena, Teramo, Terni, Treviso, Udine, Varese, Venezia, Vicenza; *high incidence*: Alessandria, Aosta, Asti, Belluno, Bergamo, Biella, Bologna, Bolzano, Brescia, Como, Cremona, Genova, Imperia, Lecco, Lodi, Mantova, Massa Carrara, Milano, Modena, Monza e della Brianza, Novara, Parma, Pavia, Pesaro e Urbino, Pescara, Piacenza, Reggio Emilia, Rimini, Savona, Sondrio, Torino, Trento, Trieste, Verbano-Cusio-Ossola, Vercelli, Verona.

(*Table 1*); of these deaths, 36,445 were in excess relative to the expected numbers in the same period (36,445 = 307,809-271,364; 271,364.2 = average in the previous 5 years (from January 1 to May 31 during the years: 2015-2019). Of these 36,445 excess deaths, the majority occurred in laboratory-confirmed COVID-19 cases, registered by the Italian National Integrated COVID-19 Surveillance System (32,236 "deaths with a diagnosis of COVID-19"; COVID-19 deaths as in *Table 1*), whereas the remaining 4,209 (11.5%) were deaths "without a COVID-19 diagnosis."

Figure 1 (part A) clearly shows an increase in the excess number of deaths (red line *vs* blue line) from early March 2020, with a peak at the end of March, which exactly followed the peak of "deaths with a COVID-19 diagnosis" (in March 27, *Supplementary Figure 2* available online). The excess number of deaths was lower in provinces reporting low to moderate levels of COVID-19 incidence (*Figure 1, parts B and C*) and higher in those with a higher diffusion of COVID-19 (*Figure 1, part D*).

When considering temporal patterns in the excess number of deaths by sex and age (*Supplementary Figure 3* available online), we observed that the excess number of all-cause deaths was higher in males than in females in all age groups, except for females aged 90+ years. Similarly, the number of deaths "diagnosed as non-COVID-19" was higher in males aged 80-89 years respect the number of females of the same age group. Furtherly, the number of deaths "diagnosed as non-COVID-19" was higher in females aged 90+ *vs* males aged 90+ years (*Supplementary Figure 3, parts B, C, E, and F* available online).

Estimates of the cumulative incidences of deaths from all causes and comparison with the expected number of deaths according to the COVID-19 diffusion over time

The observed and estimated cumulative mortalities from all causes of death are reported by month in *Supplementary Table 1* available online. The observed estimates of mortality were lower than expected until February 2020 (i.e., the observed mortality was 0.19% on February 2020 *vs* the expected mortality of 0.21%) and higher afterwards, ranging from 0.34 in March to 0.54 in May 2020 (*Supplementary Table 1* available online).

In areas with a high spread of COVID-19 (*Figure 2, part B*), the observed rates increased from 0.10% in January 2020 to 0.66% in May 2020, whereas in those with a lower spread, the all-cause mortality decreased from 0.10% to 0.44% over the same period, respectively.

These results were confirmed by the Cox models, which yielded RH of observed to expected numbers of

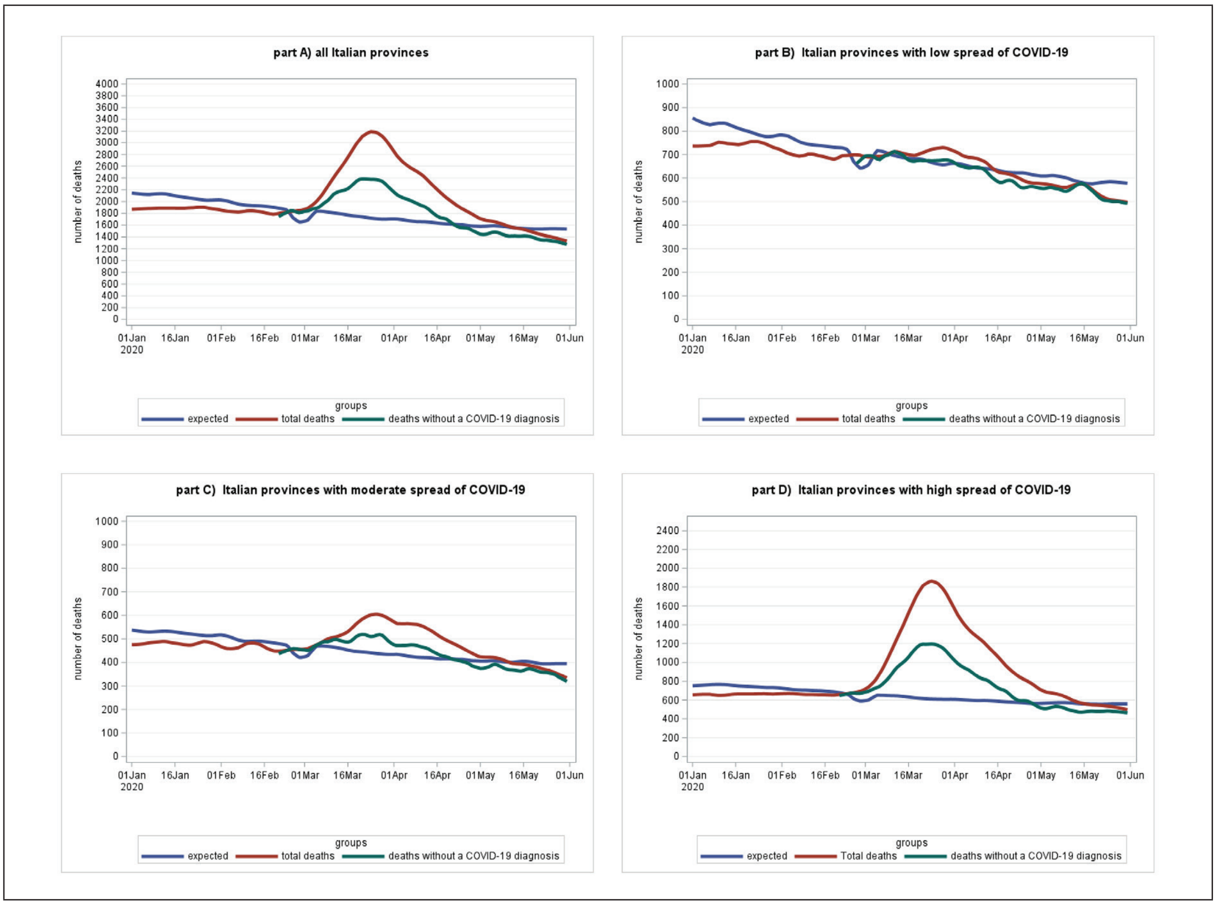

Figure 1
All-cause daily deaths from January 1 to May 31, 2020 (red curve); average number of deaths (blue curve) that occurred between 2015 and 2019 in the same study period (expected deaths); deaths without a COVID-19 diagnosis (green curve, i.e., difference between all-cause deaths and deaths with a COVID-19 diagnosis); curves fitted by locally estimated scatterplot smoothing (LOESS), local regression method. Described graph curves in: part A) all Italian provinces; part B) Italian provinces with low spread of COVID-19; part C) Italian provinces with moderate spread of COVD-19; part D) Italian provinces with high spread of COVID-19.

deaths from all causes of 1.375 (95% CI: 1.364-1.386), 1.049 (95% 1.038-1.060), and 0.967 (95% CI:0.959-0.976), respectively, in areas with a high, moderate, and low spread of COVID-19, for the year 2020 (January-May).

In *Figure 2 (part* E), the observed mortality from deaths without a COVID-19 diagnosis is graphed together with the expected mortality from all causes of death; therefore, the RH of the observed number of deaths without a COVID-19 diagnosis relative to the expected number of deaths was 1.016 (95% CI: 1.011-1.022), accounting for about 2% of excess mortality.

Restricting the analysis to the period of March-April 2020 (peak of the epidemic), the RH of excess deaths was 1.435 (95% CI: 1.424-1.446). We also estimated that the excess mortality increased by 35.2% for every 500 increase per 100,000 residents in the number of new cases of COVID-19 (*Table* 3).

DISCUSSION

Applying a population-based cohort approach, we estimated excess mortality attributable to the COVID-19 epidemic according to different geographical spread of the virus during the first 5 months in 2020. This was possible, thanks to the use of two nationwide data sources (ISS and ISTAT).

The study highlights an "excess" in mortality, an epidemiological term used to give an estimate of the difference between the observed number of deaths in a specific period and the expected number of deaths in the same period [3]. This could be due to both the direct and indirect effects of the pandemic. Direct effects take into consideration the number of deaths diagnosed as "COVID-19." All the excess deaths may be due to COVID-19, but they might have occurred before a possible diagnosis of SARS-CoV-2 infection (for instance, deaths from organ dysfunctions, such as heart or kidney failures, are likely to be triggered by the virus).

The indirect effects include the deaths of patients suffering from serious diseases other than COVID-19 due to the overload of the National Health Service, especially in the most affected areas, as also observed in the USA [15, 16]. The anticipated deaths of fragile people, such as the elderly, could be considered as another indirect effect of the pandemic. Only future analyses extended to the entire period of 2020-2021 will be able to assess the full impact of the pandemic, including a possible "collection" phenomenon, the so-called "har-

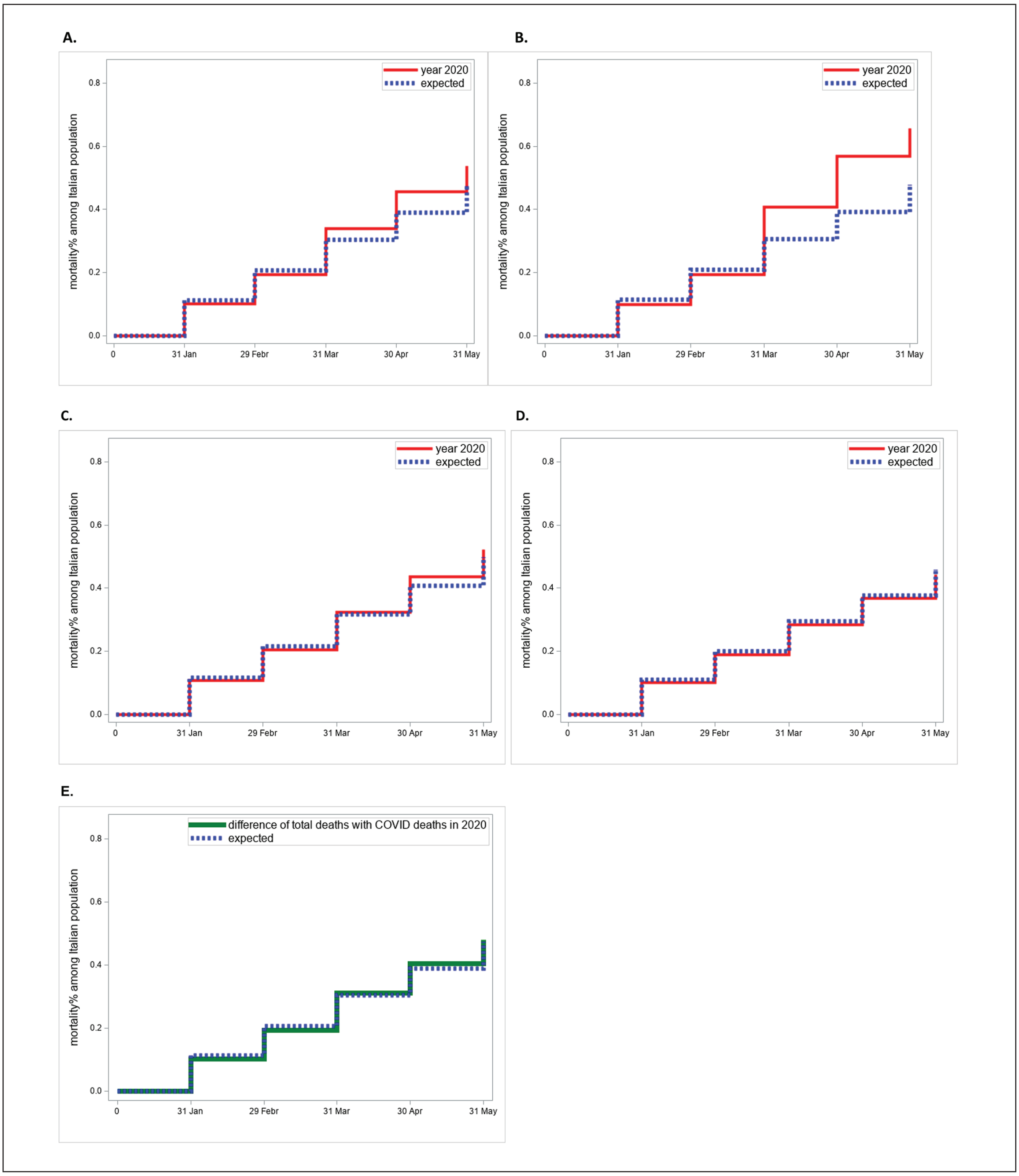

Figure 2
Part A) all-cause mortality in Italy during the study period; all provinces throughout Italy (part A), RH = 1.134 (95% CI: 1.129-1.140); part B) provinces with a high spread of COVID-19, RH of mortality observed in 2020 *vs* all-cause expected mortality = 1.375 (95% CI: 1.364-1.386); part C) provinces with moderate spread, RH of 2020 *vs* expected deaths = 1.049 (95% 1.038-1.060); part D) provinces with low spread, RH of 2020 *vs* expected = 0.967 (95% CI: 0.959-0.976); part E) cumulative incidences of deaths without a COVID-19 diagnosis *vs* expected deaths, RH = 1.016 (95% CI: 1.011-1.022).

vesting effect." Furthermore, the sanitary containment measures taken to tackle the health crisis during the first wave have significantly reduced violent deaths, such as traffic accidents.

In Italy, mortality from any cause started to rise quickly since the beginning of March 2020 and declined after a peak of more than 3,000 deaths in March 27. The same trends were observed for COVID-19-related deaths and deaths registered without a COVID-19 diagnosis. Therefore, the excess deaths gradually diminished after the peak in March 27.

We estimated a RH of excess deaths of 1.13 between January 1 and May 31/2020, i.e., an excess of 13% in the observed number of deaths from all causes com-

Table 3
Relative hazard (RH) of excess mortality (year 2020 vs expected) and RH for increments of 500 cases per 100,000 residents during the peak months of epidemic (March and April 2020)

	RH of death for any cause	95% CI
Year 2020 vs expected	1.435	(1.424-1.446)
Incidence rate of COVID-19 per increments of 500 x 100,000 residents	1.352	(1.346-1.359)

pared with the number of expected deaths from all causes. In the same period, the RH of the observed deaths without a COVID-19 diagnosis compared with the expected deaths from all causes was equal to 1.016, indicating that about 2% of the excess deaths occurred in subjects without a diagnosis of COVID-19. This percentage may partly reflect a delay in the diagnosis and/or access to health services (individuals may have died before being tested). It should be noted that these findings should be confirmed by analyzing all the causes of deaths. In this regard, the Italian National Institute of Statistics recently analyzed all the causes of death observed in March-April 2020 (the period of the first epidemic wave in Italy), considering deaths obtained from the death certificate data (https://www.istat.it/it/files//2021/04/Report-Cause-di-Morte_21_04_2021.pdf). They found that COVID-19 was the leading cause of death among males and the second (after neoplasms) among females.

When exploring excess deaths by sex and age groups, we observed that the all-cause excess deaths were higher in males than in females in all age groups, except for females aged 90 years or older. Notably, we observed in both sexes, aged 80 years or older, a higher number of deaths "diagnosed as non-COVID-19" during the peak months, especially in females aged 90 years or older. Unsurprisingly, this result reflected the structure of the Italian population, as in the rest of Europe, where females account for more than 70% of individuals aged 90 years or older (https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Population_structure_and_ageing). Furthermore, these results could prompt further research on the origins and causes of under-reporting of COVID-19 deaths.

We also demonstrated that the largest increase in mortality was observed in provinces with a COVID-19 incidence rate of 450-1,610 cases per 100,000 residents, with an overall increase in mortality of about 37%. Conversely, the RH of excess deaths decreased to 5% in provinces with a moderate level of COVID-19 incidence (150-449 cases per 100,000 residents); we observed an RH of death lower than 1 in provinces with a COVID-19 incidence lower than 30-149 cases per 100,000 residents, i.e., the mortality observed in 2020 was lower than expected. Our results are only in part consistent with a recent study conducted in Italy [17]. The authors reported "an estimated increase of 47,490 (95% eCI: 43,984-50,362) from the expected baseline numbers when accounting for temporal trends and differences in temperature distribution," corresponding to a percentage excess of 29.5% (95% eCI: 26.8-31.9). Even if not directly comparable because of differences in the periods analyzed and in the design and method used, we obtained a high percentage (43%) of excess mortality in March and April. Again, even if direct comparison with our results was impossible, our study partially replicates findings from a population-based cohort study conducted on a large sample of the UK population (3,862,012 of adults), in which the different scenarios hypothesized by the UK model approximately correspond to the three levels of COVID-19 spread in Italy (low, moderate, and high) [7]. We also estimated the increase in the RH for all-cause mortality with the increase in the incidence rate of COVID-19 (for increments of 500 cases per 100,000 residents), which was equal to 1.352 (95% CI: 1.346-1.359), corresponding to an increase of about 35% in the excess deaths for every 500 increase in the number of diagnoses per 100,000 residents. The last estimate could be a benchmark: for example, the excess mortality would be almost double (70%) for incidences of 1,000 new COVID-19 cases per 100,000, and this is what happened unfortunately in many municipalities in the north during the first wave [10, 17].

We also observed that among subjects aged 80 years, those who were living in areas with a high spread of COVID-19 were relatively more affected than those of the same age who were living in areas with a low spread of COVID-19 (27.1% *vs* 14.1%). This result highlights that, during the first wave, a high proportion of COVID-19 cases was reported in nursing homes, especially in the area with a high spread of the epidemic, i.e., in the northern regions [18].

Several studies have been published in Europe [6, 7], in USA [15, 16], and Italy [4, 5], taking into account excess deaths; however, this is one of the first studies to summarize all the available information on mortality, integrating data from two national data sources, namely, the National Institute of Statistics and COVID-19 Surveillance Data from the National Institute of Health. Therefore, the strength of this study relies on the fact that it was based on almost the entirety of the Italian resident population. Furthermore, the heterogeneity in the geographical spread of the epidemic allowed for the estimation of the different impacts on excess deaths in different areas according to different levels of COVID-19 incidence.

This study has some limitations. First, we applied a cohort approach without using individual data, but using aggregated data. While we could apply survival techniques as we knew the individual death dates, we were not able to be as accurate on COVID-19 exposure: given the association between aggregated cases/deaths at the province level, we cannot ignore the possible impact of ecological fallacy on our estimates of excess deaths. Furthermore, we assumed that the entire Italian population was exposed to the virus starting from an arbitrary date (January 1, 2020), and as a result, we have presumably underestimated the excess mortality. However, a recent study in Italy showed virus detection in wastewater since December 2019 in large cities in the north [20], confirming the appropriateness of this date as time zero of our analysis, especially as regards the northern regions.

Second, we used the COVID-19 incidence rate per 100,000 residents as a measure of the spread of the virus. Moreover, at the beginning of the epidemic, the incidence rates of COVID-19 were mostly representative of symptomatic people [1]. Therefore, when including asymptomatic cases, the estimates of COVID-19 incidence would have been much higher, as was the case in subsequent epidemic waves including the different geographical patterns of the Sars-Cov2 diffusion [21]. Finally, we defined deaths as "COVID-19-related" when occurring in patients who tested positive for SARS-CoV-2 by RT-PCR, regardless of pre-existing diseases that may have caused the death, as previously reported [1, 19]. This also reflected the unclearness of the definition of COVID-19-related death at the start of the epidemic.

In conclusion, our analysis, in which we used a population-based cohort approach, provided estimates of mortality excess due to COVID-19 in Italy, in particular during the first wave of the epidemic, when the surveillance system had just been established.

Authors' contributions

MD performed the statistical analyses and drafted the manuscript; GM contributed to the design of the study and revised the advanced draft of the manuscript; SB elaborated surveillance data and revised critically the manuscript; VM elaborated mortality data; SP coordinated and supervised National mortality data; MB and GC elaborated National mortality data and revised critically the manuscript; XA elaborated surveillance data and critically revised the manuscript, FR, MF, MFV elaborated surveillance data and critically revised the manuscript, MS, AMU and MDM elaborated surveillance data; PP is the head of the Italian coronavirus disease surveillance system and revised the manuscript; AB coordinated and supervised the surveillance data collection, contributed to the conception and design of the study, critically revising the manuscript; all authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

Acknowledgments

We thank Rocco Turroni for English language editing and the reviewers for useful suggestions that we added in the study. We also thank Flavia Chiariotti for useful suggestions in the discussion.

Funding

This research was funded by EU grant 874850 MOOD (MOnitoring Outbreak events for Disease surveillance in a data science context).

Conflict of interest statements

The Authors declare that they have no competing interests.

The Authors declare that data on excess mortality in year 2020 has been published previously, but not in the same form, with different objectives and statistical methods (Dorrucci M, Minelli G, Boros S, *et al.* Excess Mortality in Italy During the COVID-19 Pandemic: Assessing the Differences Between the First and the Second Wave, Year 2020. Front Public Health. 2021; doi: 10.3389/fpubh.2021.669209).

Patient consent for publication

Not required.

Ethics approval

This study was not submitted for approval to an ethical committee because the scientific dissemination of COVID-19 surveillance data was authorized by the Italian Presidency of the Council of Ministers on the 27th of February 2020.

Members of the Italian Integrated Surveillance COVID-19 group

Stefania Bellino, Maria Cristina Rota, Ornella Punzo, Paolo D'Ancona, Roberta Urciuoli, Stefania Giannitelli, Paola Stefanelli, Alessandra Ciervo, Corrado Di Benedetto, Marco Tallon (Istituto Superiore di Sanità). Regional representatives: Manuela Di Giacomo (Abruzzo), Michele La Bianca (Basilicata), Anna Domenica Mignuoli (Calabria), Pietro Buono (Campania), Erika Massimiliani (Emilia-Romagna), Fabio Barbone (Friuli Venezia Giulia), Francesco Vairo (Lazio), Camilla Sticchi (Liguria), Danilo Cereda (Lombardia), Lucia di Furia (Marche), Francesco Sforza (Molise), Maria Grazia Zuccaro (P.A. Bolzano), Pier Paolo Benetollo (P.A. Trento), Chiara Pasqualini (Piemonte), Lucia Bisceglie (Puglia), Maria Antonietta Palmas (Sardegna), Salvatore Scondotto (Sicilia), Emanuela Balocchini (Toscana), Anna Tosti (Umbria), and Mauro Ruffier (Valle D'Aosta); and Filippo Da Re (Veneto).

Received on 22 September 2021.
Accepted on 3 February 2022.

REFERENCES

1. Riccardo F, Ajelli M, Andrianou XD, et al. Epidemiological characteristics of COVID-19 cases in Italy and estimates of the reproductive numbers one month into the epidemic 2020. medRxiv preprint. doi: https://doi.org/10.1101/2020.04.08.20056861.
2. Armocida B, Formenti B, Ussai S, Palestra F, Missoni E. The Italian health system and the COVID-19 challenge. The Lancet Public Health. 2020;5:e253. doi: https://doi.org/10.1016/S2468-2667(20)30074-8.
3. Centers for Disease Control and Prevention (CDC). National Center for Health Statistics. Excess deaths associated with COVID-19. Available from: https://www.cdc.gov/nchs/nvss/vsrr/covid19/excess_deaths.htm.
4. Blangiardo M, Cameletti M, Pirani M, Corsetti G, Battaglini M, et al. Estimating weekly excess mortality at sub-national level in Italy during the COVID-19 pandemic. PLOS ONE. 2020;15:e0240286. doi: https://doi.org/10.1371/journal.pone.0240286.

5. Michelozzi P, De' Donato F, Scortichini M, et al. Mortality impacts of the coronavirus disease (COVID-19) outbreak by sex and age: rapid mortality surveillance system, Italy, 1 February to 18 April 2020. Euro Surveillance. 2020;25:2000620.
6. Stang A, Standl F, Kowall B, et al. Excess mortality due to COVID-19 in Germany. J Infect. 2020;5:797-801. doi: https://doi.org/10.1016/j.jinf.2020.09.012.
7. Banerjee A, Pasea L, Harris S, et al. Estimating excess 1-year mortality associated with the COVID-19 pandemic according to underlying conditions and age: a population-based cohort study. Lancet. 2020;395:1715-25.
8. Kontis V, Bennett JE, Rashid T, et al. Magnitude, demographics and dynamics of the effect of the first wave of the COVID-19 pandemic on all-cause mortality in 21 industrialized countries. Nat Med. 2020;26:1919-28. doi: https://doi.org/10.1038/s41591-020-1112-0.
9. Istituto Superiore di Sanità - Istituto Nazionale di Statistica. Terzo rapporto ISS-ISTAT sulla mortalità della popolazione residente, July 2020. Available from: https://www.ISTAT.it/it/archivio/245415.
10. Dipartimento della Protezione Civile. Ordinanza del Capo della Protezione Civile n. 640 del 27 febbraio 2020.
11. Coronavirus disease 2019 (COVID-19): situation report-57. Published March 17, 2020. Accessed March 18, 2020. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200317-sitrep-57-covid-19.pdf?sfvrsn=a26922f2_2.
12. European Centre for Disease Prevention and Control. Case definition and European surveillance for COVID-19, as of 2 March 2020 [Internet]. Available from: www.ecdc.europa.eu/en/case-definition-and-european-surveillance-human-infection-novel-coronavirus-2019-ncov.
13. Cleveland WS, Grosse E. Computational methods for local regression. Statistics and Computing. 1991;1:47-62.
14. Collett D. Modeling survival data in medical research. London: Chapman and Hall; 1994.
15. Weinberger DM, Cohen T, Crawford FW, et al. Estimating the early death toll of COVID-19 in the United States. Preprint. bioRxiv. 2020; doi:10.1101/2020.04.15.20066431.
16. Woolf SH, Chapman DA, Sabo RT, Weinberger DM, Hill L. Excess deaths from COVID-19 and other causes, March-April 2020. JAMA. 2020, research letter.
17. Scortichini M, Schneider dos Santos R, De' Donato F, et al. Excess mortality during the COVID-19 outbreak in Italy: a two-stage interrupted time-series analysis. Intern J Epidemiol. 2020;49(6):1909-17. doi: https://doi.org/10.1093/ije/dyaa169.
18. Lombardo FL, Bacigalupo I, Salvi E, et al. The Italian national survey on coronavirus disease 2019 epidemic spread in nursing homes. Int J Geriatr Psychiatry. 2021 Jun;36(6):873-82. doi: 10.1002/gps.5487.
19. Onder G, Rezza G, Brusaferro S. Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. JAMA. 2020;323(18):1775-6. [Internet]. March 23. doi:10.1001/jama.2020.4683.
20. La Rosa G, Mancini P, Bonanno Ferraro G, et al. SARS-CoV-2 has been circulating in northern Italy since December 2019: Evidence from environmental monitoring. Sci Total Environ. 2021;750:141711. doi: 10.1016/j.scitotenv.2020.141711.
21. Epicentro. Epidemia COVID-19 Aggiornamento nazionale 29 dicembre 2020. Available from: www.epicentro.iss.it/coronavirus/bollettino/Bollettino-sorveglianza-integrata-COVID-19_29-dicembre-202.pdf.