



OPEN COVID-19 mortality in Italy varies by patient age, sex and pandemic wave

Francesca Minnai¹, Gianluca De Bellis¹, Tommaso A. Dragani²✉ & Francesca Colombo¹

SARS-CoV-2 has caused a worldwide epidemic of enormous proportions, which resulted in different mortality rates in different countries for unknown reasons. We analyzed factors associated with mortality using data from the Italian national database of more than 4 million SARS-CoV-2-positive cases diagnosed between January 2020 and July 2021, including > 415 thousand hospitalized for coronavirus disease-19 (COVID-19) and > 127 thousand deceased. For patients for whom age, sex and date of infection detection were available, we determined the impact of these variables on mortality 30 days after the date of diagnosis or hospitalization. Multivariable weighted Cox analysis showed that each of the analyzed variables independently affected COVID-19 mortality. Specifically, in the overall series, age was the main risk factor for mortality, with HR > 100 in the age groups older than 65 years compared with a reference group of 15–44 years. Male sex presented a two-fold higher risk of death than female sex. Patients infected after the first pandemic wave (i.e. after 30 June 2020) had an approximately threefold lower risk of death than those infected during the first wave. Thus, in a series of all confirmed SARS-CoV-2-infected cases in an entire European nation, elderly age was by far the most significant risk factor for COVID-19 mortality, confirming that protecting the elderly should be a priority in pandemic management. Male sex and being infected during the first wave were additional risk factors associated with COVID-19 mortality.

COVID-19, the disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is responsible for a worldwide pandemic of enormous proportions, in terms of the numbers of both infections and deaths. SARS-CoV-2 infection may be asymptomatic or symptomatic, with severity ranging from a few flu-like symptoms to severe respiratory manifestations requiring supplemental oxygen and to multi-organ failure^{1–3}. COVID-19 mortality has been widespread throughout the world, but with considerable temporal and geographical variations^{4–7}. Although the reasons for these differences are not yet fully elucidated, according to a meta-analysis⁸ the predictors of mortality from COVID-19 that have been repeatedly observed in different case series from different countries are advanced age, male sex, and pre-existing comorbidities; in addition, some abnormal values of laboratory biomarkers have been associated with poor prognosis.

Italy was the first country after China to be heavily affected by the pandemic, and it has registered excess overall mortality and more than 127 thousand COVID-19-related deaths by July 25, 2021^{9,10}. The trends of cases and deaths, in most countries, have not been constant over time, but have fluctuated, with peaks of incidence called "waves". The first wave, in Italy and other European countries, began in January 2020 and lasted until summer 2020^{7,11}. Indeed, during the summer there were few cases and deaths, but since the fall the numbers of cases and deaths rose again with subsequent waves⁹, due in part to virus variants.

There is ongoing discussion about the effects of different waves on mortality in COVID-19 patients¹¹. In particular, a few studies showed that mortality decreased after the first wave (i.e. after June 2020)^{12,13}, but these results have not yet been confirmed by the analysis of country-wide mortality data. Therefore, we analyzed factors associated with COVID-19 mortality in the Italian national case series. In particular, we analyzed the effect of sex, age and period of infection on the overall survival of hospitalized and not hospitalized COVID-19 patients, using a Cox model that is the most commonly used method to deal with time to event data and analyze prognostic factors¹⁴.

¹Institute of Biomedical Technologies, National Research Council (ITB-CNR), Segrate, MI, Italy. ²Department of Research, Fondazione IRCCS Istituto Nazionale Dei Tumori, Milan, Italy. ✉email: tommaso.dragani@istitutotumori.mi.it

Materials and methods

Ethical statement. Data used in this study were collected by the Istituto Superiore di Sanità (ISS; Rome, Italy), an agency of the Italian government, for the Italian national integrated COVID-19 surveillance program, and kept in a confidential database. Ethical approval of data collection was handled by ISS. Data were anonymized prior to release or use. The scientific dissemination of these data was authorized by the Italian Presidency of the Council of Ministers on February 27, 2020 (Ordinance n. 640) and August 4, 2020 (Ordinance n. 691). Data were provided to us upon request (protocol no. AOO-ISS-19/04/2921-0014810).

Dataset. Data on people infected with SARS-CoV-2 in Italy (Italian COVID-19 epidemiological surveillance data) were obtained from ISS after filling a request on April 15, 2021, at <https://www.iss.it/richiesta-dati-covid19>. The series included people whose PCR diagnosis of infection or first symptoms were recorded from January 28, 2020, to July 25, 2021. The data regarded each person's age at diagnosis, sex, nationality, date of PCR-confirmed diagnosis, presence/absence of symptoms, date of hospitalization (if pertinent), the alive/dead status on the day of data sharing (i.e., July 25, 2021) and, if relevant, the date of death. Data on comorbidities were not provided.

Survival analyses. Survival was analyzed in the entire cohort and, separately, for non-hospitalized and hospitalized patients. For the entire cohort and for non-hospitalized persons, we assessed mortality within 30 days of the date of infection detection (positive PCR test). Instead, for hospitalized patients, we assessed mortality within 30 days of the date of hospitalization. In all analyses, we considered only those persons for whom complete data were available regarding age, sex, date of infection detection or of hospitalization (depending on the group), and status (alive or dead) after 30 days. Patients who died on the day of infection detection or hospitalization were assigned 1 day of follow-up. Patients whose date of death was erroneously reported in the dataset as being before the date of infection detection or hospital admission were excluded from analysis.

Survival analyses considered the effects of age, sex and pandemic wave. In the choice of age groups, we took as reference the age group 15–44 years, instead of the youngest class (0–14 years). The reasons for this decision were that the age group 15–44 years has undergone only marginal alterations in overall mortality during the pandemic in European countries¹⁵, and it is larger than the younger age class (0–14 years). Therefore, its use as reference provides good stability for risk estimates. For pandemic waves, we defined the first wave as that from the beginning of the dataset (January 28, 2020) until June 30, 2020. We also defined a second period starting on July 1, 2020 and ending on December 31, 2020, and a third period from this latter date (approximately when vaccination started in Italy) to July 25, 2021. These latter two periods are here called the second and third semesters.

Associations between demographic-clinical features (i.e. age, sex and pandemic wave) and survival were evaluated, using the survival package in R environment (R version 3.6.0), to draw Kaplan–Meier curves and run the log-rank test. The hazard proportionality assumptions were verified through the function “cox.zph()” of the survival package. The variables that were found to impact upon survival in log-rank test (with $P < 0.05$) were analyzed in a multivariable weighted Cox analysis to account for non-proportional hazards^{16,17}. We did this using the coxphw R package applying the Average Hazards Ratio method, by setting the parameter *template* = “AHR”. Cox and log-rank test P values < 0.05 (two-sided) indicated sufficient statistical significance.

Results

The Italian cohort of SARS-CoV-2-positive individuals included 4,333,014 persons of median age 46 years, with a slight predominance of female cases (Table 1). Overall, 50.4% of the cases were symptomatic and 415,390 were hospitalized, with 13.7% of them requiring intensive care. The median age of non-hospitalized patients was much lower than that of hospitalized cases (44 vs. 70 years). Of the entire case series, 240,850 were diagnosed during the first wave, 1,907,690 became infected between July 1, 2020 and December 31, 2020, and 2,104,894 were diagnosed in the first seven months of 2021. At the end of the study, 127,524 subjects were dead, with 35,837 having been diagnosed as infected before June 30, whereas 49,120 and 40,755 dead patients had been diagnosed in the second and third semesters, respectively (for 1812 subjects, information about the date of infection detection was missing). Over 31 thousand non-hospitalized people had died by July 25, 2021, while the total number of deaths among the hospitalized people was 95,907, including 25,320 during the first wave. The median age at death of non-hospitalized patients was 86 years, with 94% of them being ≥ 65 years old (not shown). Instead, the median age of deceased hospitalized patients was 81 years, with 90% of them ≥ 65 years old (not shown). Surprisingly, more than 38 thousand people who were hospitalized were listed in the dataset as being asymptomatic.

To investigate the factors affecting mortality after SARS-CoV-2 infection, we first drew Kaplan–Meier curves for the whole series and, separately, for non-hospitalized and hospitalized patients. For the whole series, the analysis was limited to 4,224,698 persons, after eliminating 108,316 persons with incomplete data (including 1744 persons for whom the date of death was erroneously listed as being before the date of diagnosis or hospitalization). For non-hospitalized and hospitalized persons, these numbers were 3,816,311 and 412,942, respectively. We tested the effects of age, sex, and pandemic wave (first wave, before June 30, 2020, was taken as reference) on the risk of death 30 days after infection detection or hospitalization. Highly significant associations (log-rank test, $P < 2 \times 10^{-16}$) for all three variables were observed, in the whole series and in the two subsets (Fig. 1). The probability of survival decreased with increasing age and was lower for males than females and for persons who were diagnosed in the first wave.

Since the proportionality hazard assumptions were not verified in our series (Schoenfeld residuals test, $P < 2 \times 10^{-16}$), we carried out weighted multivariable Cox analyses to deal with non-proportional hazards. Multivariable analysis of mortality in the whole series demonstrated that age, male sex, and the first pandemic wave were highly significant, independent risk factors for death (Table 2). This analysis showed a sharp increase in mortality risk with increasing age, especially in persons ≥ 65 years old and with a tremendously high risk [hazard

Feature	All individuals (n = 4,333,014)	Not hospitalized (n = 3,917,611)	Hospitalized (n = 415,390)
Age, years, median (range)	46 (0–109)	44 (0–109)	70 (0–109)
Age group (years)			
0–14	433,250 (10.0)	428,290 (10.9)	4960 (1.2)
15–44	1,608,907 (37.1)	1,566,579 (40.0)	42,326 (10.2)
45–64	1,396,476 (32.2)	1,278,774 (32.6)	117,702 (28.3)
65–74	398,268 (9.2)	312,247 (8.0)	86,021 (20.7)
75–84	296,057 (6.8)	198,179 (5.1)	97,878 (23.6)
≥ 85	199,948 (4.6)	133,445 (3.4)	66,503 (16.0)
Sex, n (%)			
Female	2,211,778 (51.0)	2,034,154 (51.9)	177,624 (42.8)
Male	2,121,206 (49.0)	1,883,438 (48.1)	237,766 (57.2)
Not available	30 (0)	19 (0)	0 (0)
Nationality, n (%)			
Italian	3,488,073 (80.5)	3,161,283 (80.7)	326,789 (78.7)
Other	338,567 (7.8)	304,496 (7.8)	34,071 (8.2)
Not available	506,374 (11.7)	451,832 (11.5)	54,530 (13.1)
With symptoms, n (%)			
Yes	2,184,992 (50.4)	2,081,590 (53.1)	376,943 (90.7)
No	2,119,830 (48.9)	1,808,048 (46.2)	38,239 (9.2)
Not available	28,192 (0.7)	27,973 (0.7)	208 (0.05)
Hospitalized, n (%)			
Yes	415,390 (9.6)	0 (0)	415,390 (100)
No	3,917,611 (90.4)	3,917,611 (100)	0 (0)
Not available	13 (0)	NA	NA
In intensive care unit, n (%)			
Yes	57,083 (1.3)	0 (0)	57,083 (13.7)
No	4,275,918 (98.7)	0 (0)	358,307 (86.3)
Not available	13 (0)	0 (0)	0 (0)
Infection detection, n (%)			
Before June 30, 2020	240,850 (5.6)	152,595 (3.9)	88,632 (21.3)
From July 1 to December 31, 2020	1,907,690 (44.0)	1,754,510 (44.8)	
After December 31, 2020	2,104,894 (48.6)	1,936,024 (49.4)	325,290 (78.3)
Not available	79,580 (1.8)	74,482 (1.9)	1,468 (0.004)
Deceased at study end, n (%)^a			
Yes	127,524 (2.9)	31,617 (0.8)	95,907 (23.1)
No	4,177,522 (96.4)	3,858,983 (98.5)	318,538 (76.7)
Not available	27,968 (0.6)	27,011 (0.7)	945 (0.2)
Deceased at 30 days ^b , n (%) ^c	109,605 (2.5)	27,409 (0.7)	85,503 (20.6)

Table 1. Demographic and clinical features of 4,333,014 SARS-CoV-2-positive individuals in the Italian national database, from January 28, 2020 to July 25, 2021. NA not applicable. ^aJuly 25, 2021. ^bAnd with data available for survival analysis. ^cFor the entire cohort and for non-hospitalized persons, 30 days after infection detection (positive PCR test). For hospitalized persons, 30 days after admission.

ratio (HR) = 691; 95% CI, 634 to 753] in the age ≥ 85 years group. Children (0–14 years old) showed a much lower risk of death (HR = 0.120) than both the reference age group and the older groups. Male sex was associated with a ~ twofold risk of death (HR = 2.05), compared to females. The first pandemic wave (before June 30, 2020) was associated with an excess risk of death of almost threefold, compared to the subsequent periods (HR = 0.379 and 0.346 in the second and third semesters, respectively).

Multivariable analysis of non-hospitalized individuals, 30 days after diagnosis, also showed that age, sex, and pandemic wave were all independent poor prognostic factors (Table 3). Again, age conferred the highest risk of death, with an HR > 130 in the groups of subjects 65 years or older and HR > 1400 for those 85 years and older. Indeed, the vast majority of deaths (94%) among non-hospitalized patients regarded patients ≥ 65 years old. The risk estimates associated with male sex and pandemic wave are similar to those for the whole series, except for non-hospitalized patients of the third semester, who had an even lower risk of mortality (HR = 0.21).

Finally, multivariable Cox analyses of the smaller subgroup of hospitalized patients also showed that age, sex, and pandemic wave were significantly associated with the risk of death 30 days after hospitalization (Table 4). The estimates of the risk of death in this subgroup were lower than in non-hospitalized patients, as evidenced by the smaller values of HR reaching 59 even in the oldest age group, when compared to the reference group.

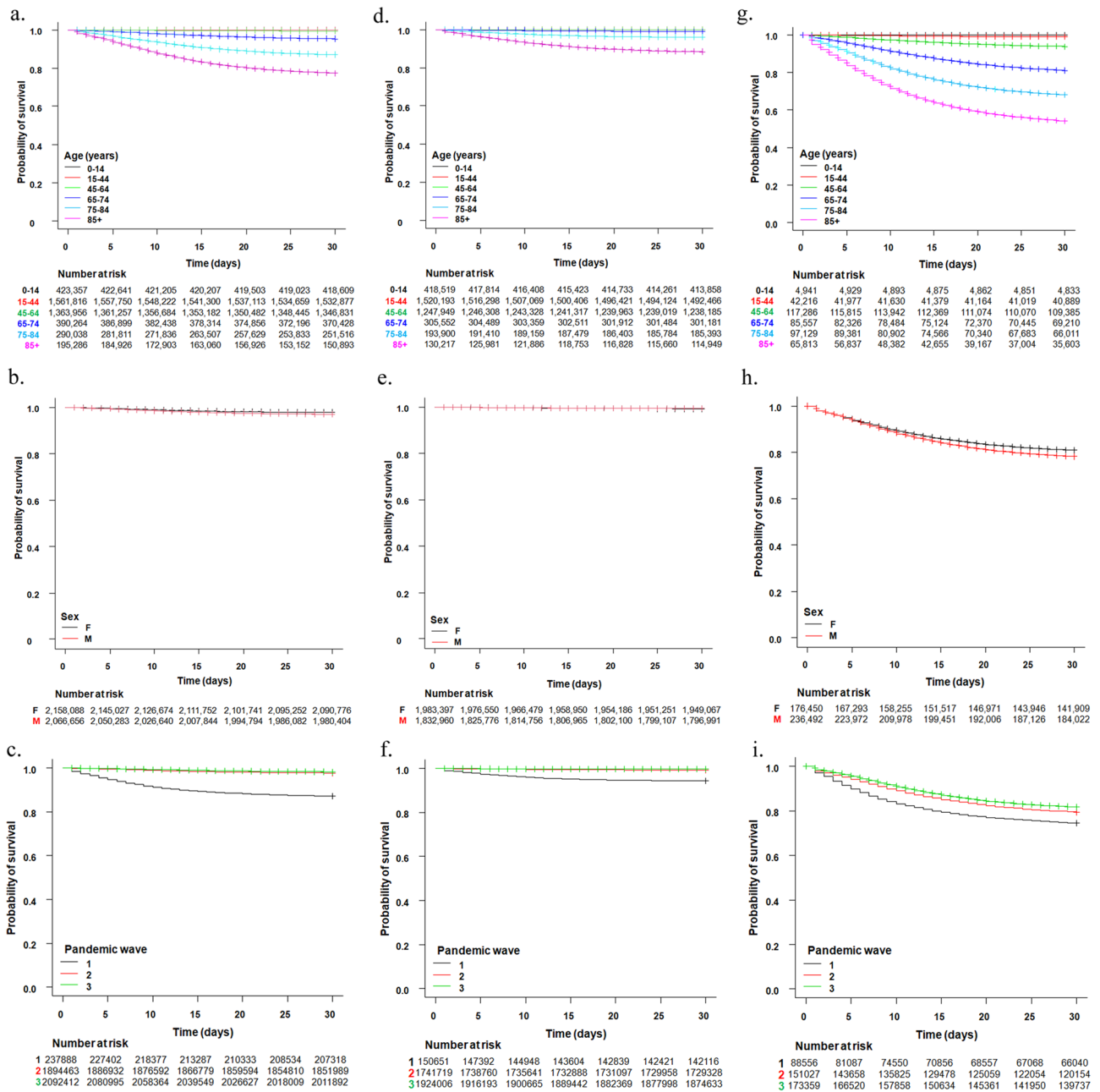


Figure 1. Kaplan–Meier survival curves for the entire Italian series of SARS-CoV-2-positive persons (panels A–C), for non-hospitalized persons (panels D–F), and for hospitalized COVID-19 patients (panels G–I) in the Italian national COVID-19 series, by age group, sex, and pandemic wave (1 = before June 30, 2020, 2 = from July 1 to December 31, 2020, and 3 = after December 31, 2020). Crosses denote censored samples. Numbers of patients at risk for each group, at each time point, are reported under each plot. Log-rank test $P < 2 \times 10^{-16}$, except for panel (E) for which $P = 1.2 \times 10^{-11}$.

Finally, being hospitalized after the first wave was associated with a lower risk of death (HR was about 0.7 in both semesters after June 30, 2020), but this effect was less intense than that observed among non-hospitalized persons. Also, the poor prognostic role of male sex was confirmed in these hospitalized patients (HR = 1.44).

Discussion

The Italian COVID-19 epidemiological surveillance dataset analyzed here contained information on over 4 million persons molecularly diagnosed with a SARS-CoV-2 infection until July 25, 2021. The dataset included information about age, sex, date of diagnosis, presence vs. absence of symptoms, date of hospitalization (if pertinent), date of death (if pertinent), and a few other data. Survival analyses on the whole series and on subsets of non-hospitalized and hospitalized patients strongly confirmed the pivotal role of age in the probability of survival of COVID-19 patients. The analysis by age category, adjusted for sex and pandemic wave, showed that age groups

Feature	Hazard ratio (95% confidence interval)	P-value ¹
Age group (years)		
0–14	0.12 (0.074–0.194)	$< 2 \times 10^{-16}$
15–44	1.00	
45–64	16.9 (15.5–18.4)	$< 2 \times 10^{-16}$
65–74	126 (116–137)	$< 2 \times 10^{-16}$
75–84	358 (328–390)	$< 2 \times 10^{-16}$
≥ 85	691 (634–753)	$< 2 \times 10^{-16}$
Sex		
Female	1.00	
Male	2.05 (2.03–2.08)	$< 2 \times 10^{-16}$
Pandemic wave		
Before June 30, 2020	1.00	
From July 1 to December 31, 2020	0.379 (0.374–0.385)	$< 2 \times 10^{-16}$
After December 31, 2020	0.346 (0.340–0.352)	$< 2 \times 10^{-16}$

Table 2. Factors associated with death in the whole Italian series of people infected by SARS-CoV-2. Of the entire series, 108,316 cases were excluded due to incomplete or erroneous data, for a total of 4,224,698 SARS-CoV-2-positive subjects analyzed and 109,605 deaths within 30 days of diagnosis. ¹Multivariable weighted Cox analysis (non-proportional hazards model).

Feature	Hazard ratio (95% confidence interval)	P-value ¹
Age group (years)		
0–14	0.2959 (0.1443–0.6074)	9×10^{-4}
15–44	1.000	
45–64	16.19 (13.29–19.72)	$< 2 \times 10^{-16}$
65–74	134.9 (111.2–163.7)	$< 2 \times 10^{-16}$
75–84	527.7 (435.7–639.3)	$< 2 \times 10^{-16}$
≥ 85	1457 (1203–1765)	$< 2 \times 10^{-16}$
Sex		
Female	1.000	
Male	1.859 (1.813–1.904)	$< 2 \times 10^{-16}$
Pandemic wave		
Before June 30, 2020	1.000	
From July 1 to December 31, 2020	0.3492 (0.3391–0.3602)	$< 2 \times 10^{-16}$
After December 31, 2020	0.2112 (0.2041–0.2194)	$< 2 \times 10^{-16}$

Table 3. Factors associated with death in non-hospitalized subjects in the Italian national database. Among non-hospitalized persons, 101,300 cases were excluded due to incomplete or erroneous data, for a total of 3,816,311 SARS-CoV-2-positive subjects analyzed and 27,409 deaths within 30 days of diagnosis. ¹Multivariable weighted Cox analysis (non-proportional hazards model).

older than 65 had mortality risks that were hundreds of times greater than that of the 15- to 44-year-old reference class. The 0–14 years age group had a mortality risk that was about 10 times less than that of the reference class. Male sex was also confirmed to be a poor prognostic factor, but with a much smaller effect. Additionally, our analysis demonstrated that being diagnosed during the first pandemic wave (until June 2020) was associated with an approximately threefold higher mortality risk than being diagnosed later.

In non-hospitalized patients, the mortality risk associated with age was greater than that for the whole series. This difference might be explained by the observation that most deceased non-hospitalized patients were very old, with a median age of 86 years. As a possible interpretation, we suppose that some elderly persons deteriorated rapidly and died before they could be hospitalized.

In hospitalized patients, old age was associated with an excess risk of death, as in the whole series, although the statistical estimates were lower. For example, for the age group ≥ 85 years old, the HR was 58.7 for hospitalized patients and 687 for the whole series. The difference may be explained by the facts that hospitalized patients were much older than subjects of the whole series (median age, 70 vs. 46 years), and that hospitalization itself poses an excess risk of death, as age is a known risk factor for hospitalization^{18,19}, including in our series (not shown).

Our finding of age being a risk factor for COVID-19 mortality is in agreement with that of a meta-analysis by Shi et al.⁸ on 27 studies (including 24 from China, two from the United States, and one from Italy) and a

Feature	Hazard ratio (95% confidence interval)	P-value ¹
Age group (years)		
0–14	0.154 (0.076–0.3118)	2×10^{-7}
15–44	1.00	
45–64	5.49 (4.99–6.03)	$< 2 \times 10^{-16}$
65–74	18.08 (16.46–19.86)	$< 2 \times 10^{-16}$
75–84	34.5 (31.4–37.9)	$< 2 \times 10^{-16}$
85+	59.0 (53.8–64.8)	$< 2 \times 10^{-16}$
Sex		
Female	1.00	
Male	1.44 (1.42–1.47)	$< 2 \times 10^{-16}$
Pandemic wave		
Before June 30, 2020	1.00	
From July 1 to December 31, 2020	0.713 (0.701–0.725)	$< 2 \times 10^{-16}$
After December 31, 2020	0.675 (0.663–0.686)	$< 2 \times 10^{-16}$

Table 4. Factors associated with death in hospitalized COVID-19 patients in the Italian national database. Among hospitalized persons, 2448 cases were excluded due to incomplete or erroneous data, for a total of 412,942 SARS-CoV-2-positive subjects analyzed and 85,503 deaths within 30 days of diagnosis. ¹Multivariable weighted Cox analysis (non-proportional hazards model).

meta-analysis by Booth et al.²⁰ on 66 studies with > 17 million patients from 14 countries. Both meta-analyses found an association between old age and excess risk of mortality from COVID-19, although the quantitative risk estimates differ. Of note, these meta-analyses did not report HRs associated with survival, since no Cox analyses were done. To the best of our knowledge, only one other nation-wide study, conducted in France by Semenzato et al.²¹, used Cox models to analyze the effects of age on the risk of mortality in a large number of hospitalized COVID-19 patients. Although the age groups differ between the two studies, the risk estimates are similar, with HRs > 50 in elderly patients in both studies.

Several immunological mechanisms responsible for the increased risk of death from COVID-19 in the elderly can be hypothesized. One study demonstrated that pre-existing T-cell immunity induced by circulating human alpha- and beta-coronaviruses is present in young adults but virtually absent in older adults²². Consequently, older adults had a minimal baseline frequency of cross-reactive T cells directed toward the novel SARS-CoV-2; for this reason, they may be at higher risk of severe COVID-19 disease and death. Moreover, the phenomenon of immunosenescence, which involves age-related changes in innate and adaptive immunity, has been imputed as being associated with the increased mortality of older adults infected with SARS-CoV-2²³. The elderly exhibit a deficient immunologic response to SARS-CoV-2 infection, which may be another reason for their increased risk of severe disease and death²⁴.

In the Italian nationwide COVID-19 series, male sex was an unfavorable prognostic factor for survival, with a risk that was twofold higher than for females in the whole series, and > 80% and ~ 50% higher in non-hospitalized and hospitalized patients, respectively. This result is in agreement with those of several other studies^{8,20}, although the quantitative risk estimates differ. The HRs for male sex calculated in this study, which range from 1.46 to 2.05, are similar to those reported by Semenzato et al.²¹. Additionally, in an analysis of the excess number of deaths, standardized by age in 29 high-income countries, men were more affected than women in almost all countries²⁵.

The mechanism by which sex is an unfavorable prognostic factor for COVID-19 is not yet known. Most likely, several sex-related factors contribute to the higher risk of males for poorer COVID-19 outcomes. A study of 1683 Italian patients who underwent chest computed tomography at admission showed that men had a higher prevalence of cardiovascular comorbidities, more coronary calcifications, and a higher coronary calcium score than females²⁶. Notably, the higher coronary calcific burden of men appeared to be associated with higher mortality. A study of about 3000 COVID-19 patients in a single center in China observed that the level of inflammatory cytokines in peripheral blood was higher in males than in females²⁷. Also, the percentages of CD19+ B cells and CD4+ T cells were generally higher in female patients during the course of the disease. Overall, males had greater inflammation, lower lymphocyte counts, and lower and delayed antibody responses during SARS-CoV-2 infection and recovery than females. Finally, from the perspective of an immunological mechanism, it has been hypothesized that chronic, subclinical, systemic inflammation, characteristic of aging, and immunosenescence contribute to the excess risk of COVID-19 mortality in elderly men²⁸.

Our multivariable analysis provides strong support for the hypothesis that mortality from COVID-19 was much greater during the first wave (January to June 30, 2020) than later. Indeed, taking the first wave as reference, in the subsequent periods we observed a ~ threefold reduced risk of death, both in the whole series and in non-hospitalized patients. In hospitalized patient, the excess risk of death was ~ 30% lower in the two semesters after the first wave. The excess risk of death associated with pandemic wave was first reported in an Italian study of hospitalized patients¹², and then confirmed by studies of Massachusetts healthcare workers¹³, patients of the U.S. Veterans Affairs healthcare system²⁹, and UK patients³⁰. The reasons for this effect could include the initial lack of preparedness of national health systems for pandemic management, the lack of knowledge about the

most effective therapies for COVID-19 patients with severe disease, and the possibility that frailer people were more affected at the beginning of the pandemic than the rest of the population³¹. Another possible explanation for a lower risk of mortality after June 30, 2020, and in particular, during the third semester of the pandemic, compared with the first semester, may be mass vaccination, which began in January 2021; indeed, vaccines are associated with a reduced risk of severe COVID-19 and mortality³².

In the Italian COVID-19 epidemiological surveillance dataset, more than 2 million infected persons were symptomatic (50.4% of all cases). Modeling studies on the prevalence of infection in different populations suggested that the total number of SARS-CoV-2-positive individuals exceeds symptomatic cases by an order of magnitude or more^{33–35}. If this holds true for the Italian population, then ~ 20 million people in Italy have been infected by SARS-CoV-2, i.e., 10 times the 2 million symptomatic cases. Why some infections are asymptomatic and others lead to severe COVID-19 has not yet been elucidated. Cross-reactive immunity, pre-existing in individuals who had been exposed to other coronaviruses, could be one of the mechanisms for asymptomatic and moderate courses of SARS-CoV-2 infection in many individuals³⁶.

A limitation of our study is the lack of data about COVID-19 patients' comorbidities, which are important risk factors for outcome⁸. Other possible confounders that we cannot consider in our model are, for instance, demographic, geographic and environmental factors^{37,38}. This lack of information prevented us from analyzing other risk factors for death. Moreover, the reasons why some hospitalized patients were classified as asymptomatic are not known, but their hospitalization may have been due to reasons other than COVID-19. For example, in 5432 cases, the date of SARS-CoV-2 infection detection was after the date of hospitalization, and in 13,144 patients the diagnosis was on the same day. An additional limitation could be an underestimation of the number of cases and deaths during the first wave, due to initial unpreparedness of the health system to deal with the crisis caused by the pandemic and, thus, the initially limited testing capacity³⁹. This might have determined a poorer data collection in the first wave as compared to the following periods.

Overall, this study confirms that age and male sex are independent risk factors for COVID-19 mortality for both hospitalized and non-hospitalized patients. Because age was found to be the most impactful negative prognostic factor, it should be considered in pandemic management, by giving priority to strategies aimed at protecting elderly people. Additionally, this is the first country-wide study to demonstrate a higher risk of death during the first pandemic wave than later. Similar nation-wide studies in different countries, to the best of our knowledge, have not been published. Thus, we cannot compare our study with those from other nations with different mortality rates, and we cannot exclude that such differences are due to unequal pandemic management in the first wave, considering that Italy was the first Western nation to be affected. Our study also suggests that the medical research that started with the pandemic onset and that led to the development of increasingly more effective clinical protocols contributed to improving COVID-19 patient survival²⁵. Despite the limitations of this study, principally due to the lack of some clinical data (e.g. about comorbidities and environmental factors), this study demonstrates the usefulness of a national database for studying a new disease such as COVID-19. Efforts should be made in Italy to create a more detailed national database like those of the United Kingdom⁴⁰ and France⁴¹ that collect more data on demographics, symptoms, diagnostic tests and treatments. National health databases, especially when accompanied by a national biobank of blood samples, offer great possibilities for biomedical research. They allow the construction of cohorts with unparalleled statistical power and help study risk factors for common diseases, rare diseases, and new emerging diseases such as COVID-19. Their availability could impact treatment and public health. Therefore, the creation of such databases in countries that do not yet have them and the creation of European databases are desirable.

Data availability

ISS data that have been used for the present study are available upon request at this web link: <https://www.iss.it/richiesta-dati-covid19>.

Received: 15 October 2021; Accepted: 1 March 2022

Published online: 17 March 2022

References

- Long, Q.-X. *et al.* Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nat. Med.* **26**, 1200–1204 (2020).
- Guan, W.-J. *et al.* Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med.* **382**, 1708–1720 (2020).
- White-Dzuro, G. *et al.* Multisystem effects of COVID-19: A concise review for practitioners. *Postgrad. Med.* **133**, 20–27 (2021).
- Kontis, V. *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.* **26**, 1919–1928 (2020).
- Michelozzi, P. *et al.* Temporal dynamics in total excess mortality and COVID-19 deaths in Italian cities. *BMC Public Health* **20**, 1238 (2020).
- Rostami, A. *et al.* SARS-CoV-2 seroprevalence worldwide: a systematic review and meta-analysis. *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.* **27**, 331–340 (2021).
- Achilleos, S. *et al.* Excess all-cause mortality and COVID-19-related mortality: a temporal analysis in 22 countries, from January until August 2020. *Int. J. Epidemiol.* <https://doi.org/10.1093/ije/dyab123> (2021).
- Shi, C. *et al.* Predictors of mortality in patients with coronavirus disease 2019: A systematic review and meta-analysis. *BMC Infect. Dis.* **21**, 663 (2021).
- Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet. Infect. Dis* **20**, 533–534 (2020).
- Dorrucci, M. *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 Heal.* **9**, 669209 (2021).
- Alicandro, G., Remuzzi, G. & La Vecchia, C. Italy's first wave of the COVID-19 pandemic has ended: no excess mortality in May, 2020. *Lancet (London, England)* **396**, e27–e28 (2020).

12. Ciceri, F. *et al.* Decreased in-hospital mortality in patients with COVID-19 pneumonia. *Pathog. Glob. Health* **114**, 281–282 (2020).
13. Lan, F.-Y. *et al.* Evolving virulence? Decreasing COVID-19 complications among Massachusetts healthcare workers: A cohort study. *Pathog. Glob. Health* **115**, 4–6 (2021).
14. Benítez-Parejo, N., Rodríguez del Águila, M. M. & Pérez-Vicente, S. Survival analysis and Cox regression. *Allergol. Immunopathol. (Madr)*. **39**, 362–373 (2011).
15. Euromomo. Graphs and maps. (2021). Available at: <https://www.euromomo.eu/graphs-and-maps/>. (Accessed: 15th September 2021)
16. Schemper, M., Wakounig, S. & Heinze, G. The estimation of average hazard ratios by weighted Cox regression. *Stat. Med.* **28**, 2473–2489 (2009).
17. Dunkler, D., Ploner, M., Schemper, M. & Heinze, G. Weighted cox regression using the R package coxphw. *J. Stat. Softw.* **84**, 1–26 (2018).
18. Ko, J. Y. *et al.* Risk Factors for Coronavirus Disease 2019 (COVID-19)-Associated Hospitalization: COVID-19-Associated Hospitalization Surveillance Network and Behavioral Risk Factor Surveillance System. *Clin. Infect. Dis. an Off. Publ. Infect. Dis. Soc. Am.* **72**, e695–e703 (2021).
19. Jehi, L. *et al.* Development and validation of a model for individualized prediction of hospitalization risk in 4536 patients with COVID-19. *PLoS One* **15**, e0237419 (2020).
20. Booth, A. *et al.* Population risk factors for severe disease and mortality in COVID-19: A global systematic review and meta-analysis. *PLoS One* **16**, e0247461 (2021).
21. Semenzato, L. *et al.* Chronic diseases, health conditions and risk of COVID-19-related hospitalization and in-hospital mortality during the first wave of the epidemic in France: A cohort study of 66 million people. *Lancet Reg. Heal. Eur.* **8**, 100158 (2021).
22. Saletti, G. *et al.* Older adults lack SARS CoV-2 cross-reactive T lymphocytes directed to human coronaviruses OC43 and NL63. *Sci. Rep.* **10**, 21447 (2020).
23. Cunha, L. L., Perazzio, S. F., Azzi, J., Cravedi, P. & Riella, L. V. Remodeling of the immune response with aging: Immunosenescence and its potential impact on COVID-19 immune response. *Front. Immunol.* **11**, 1748 (2020).
24. Bajaj, V. *et al.* Aging, Immunity, and COVID-19: How Age Influences the Host Immune Response to Coronavirus Infections? *Front. Physiol.* **11**, 571416 (2020).
25. Islam, N. *et al.* Excess deaths associated with covid-19 pandemic in 2020: age and sex disaggregated time series analysis in 29 high income countries. *BMJ* **373**, n1137 (2021).
26. Cereda, A. *et al.* The hidden interplay between sex and COVID-19 mortality: the role of cardiovascular calcification. *GeroScience* 1–15 (2021). <https://doi.org/10.1007/s11357-021-00409-y>
27. Huang, B. *et al.* Sex-based clinical and immunological differences in COVID-19. *BMC Infect. Dis.* **21**, 647 (2021).
28. Bonafè, M. *et al.* Inflamm-aging: Why older men are the most susceptible to SARS-CoV-2 complicated outcomes. *Cytokine Growth Factor Rev.* **53**, 33–37 (2020).
29. Ioannou, G. N. *et al.* Trends over time in the risk of adverse outcomes among patients with SARS-CoV-2 infection. *Clin. Infect. Dis. an Off. Publ. Infect. Dis. Soc. Am.* (2021). <https://doi.org/10.1093/cid/ciab419>
30. Doidge, J. C. *et al.* Trends in Intensive Care for Patients with COVID-19 in England, Wales, and Northern Ireland. *Am. J. Respir. Crit. Care Med.* **203**, 565–574 (2021).
31. Grosso, F. M. *et al.* Decreasing hospital burden of COVID-19 during the first wave in Regione Lombardia: an emergency measures context. *BMC Public Health* **21**, 1612 (2021).
32. Lopez Bernal, J. *et al.* Effectiveness of the Pfizer-BioNTech and Oxford-AstraZeneca vaccines on covid-19 related symptoms, hospital admissions, and mortality in older adults in England: test negative case-control study. *BMJ* **373**, n1088 (2021).
33. Peirlinck, M. *et al.* Visualizing the invisible: The effect of asymptomatic transmission on the outbreak dynamics of COVID-19. *Comput. Methods Appl. Mech. Eng.* **372**, 113410 (2020).
34. Pullano, G. *et al.* Underdetection of cases of COVID-19 in France threatens epidemic control. *Nature* **590**, 134–139 (2021).
35. Havers, F. P. *et al.* Seroprevalence of Antibodies to SARS-CoV-2 in 10 Sites in the United States, March 23–May 12, 2020. *JAMA Intern. Med.* <https://doi.org/10.1001/jamainternmed.2020.4130> (2020).
36. Loyal, L. *et al.* Cross-reactive CD4(+) T cells enhance SARS-CoV-2 immune responses upon infection and vaccination. *Science* <https://doi.org/10.1126/science.abh1823> (2021).
37. Fontal, A. *et al.* Climatic signatures in the different COVID-19 pandemic waves across both hemispheres. *Nat. Comput. Sci.* **1**, 655–665 (2021).
38. Páez-Osuna, F., Valencia-Castañeda, G. & Rebolledo, U. A. The link between COVID-19 mortality and PM(2.5) emissions in rural and medium-size municipalities considering population density, dust events, and wind speed. *Chemosphere* **286**, 131634 (2022).
39. Modi, C., Böhm, V., Ferraro, S., Stein, G. & Seljak, U. Estimating COVID-19 mortality in Italy early in the COVID-19 pandemic. *Nat. Commun.* **12**, 2729 (2021).
40. Sheikh, A. *et al.* Health information technology and digital innovation for national learning health and care systems. *Lancet. Digit. Heal.* **3**, e383–e396 (2021).
41. Moulis, G. *et al.* French health insurance databases: What interest for medical research?. *La Rev. Med. Int.* **36**, 411–417 (2015).

Acknowledgements

The authors acknowledge the contributions of Valerie Matarese, PhD, who provided scientific editing.

Author contributions

T.A.D., F.C. and G.D.B. conceived and designed the study. F.M. and F.C. analyzed the data. All authors interpreted the results. T.A.D. and F.C. wrote the first drafts of the manuscript and, then, all the authors contributed to the writing of the manuscript. All the authors agreed with the results and conclusions of the manuscript. F.M. and F.C. had full access to raw data. All authors had full access to the statistical reports and tables in the study.

Funding

No funding was received specifically for this work. Editing and publication costs were paid for by T.A.D.'s research funds at Fondazione IRCCS Istituto Nazionale dei Tumori. The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to T.A.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022