

Age and gender distribution of COVID-19 infected cases in Italian population

Ivan Arisi (✉ i.arisi@ebri.it)

Bioinformatics Facility, European Brain Research Institute (EBRI) Rita Levi-Montalcini , Viale Regina Elena 295, 00161, Rome , Italy <https://orcid.org/0000-0002-0536-2508>

Elide Mantuano

Institute of Translational Pharmacology CNR, Via Fosso del Cavaliere,100, 00133 Rome, Italy
<https://orcid.org/0000-0003-1656-5543>

Research Article

Keywords: SARS-CoV-2, COVID-19, Infection rate, Population structure, Sex and Gender.

Posted Date: September 8th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-72021/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Since the SARS-CoV-2 epidemic started, it became clear that the impact of the infection incidence and fatality rate were closely related to the population structure. Our analysis was devoted to the distribution of the infected cases in the Italian population stratified by age and sex in order to define the differences in gender impact of COVID-19 in each age class. Data on infected cases were extracted from the Italian EpiCentro (ISS) web site from March 12 to May 20, 2020. Data were pooled in ten years' groups. Odds ratio (OR) men versus women was evaluated by the Fisher exact test. Logistic regression was used to investigate the combined effect of age and sex on infection incidence. Statistical analysis, performed by R-Bioconductor, highlights differences in age-dependent susceptibility to infection between men and women. In elderly class (50+) generally man result more infected than women, with the exception of the oldest women (90+). In age classes <50 OR was about 1.0, while an intriguing exception is the age group 20-29 in which the ratio was unbalanced in favour of men. This analysis supports a strong influence of biological sex and environmental factors related to age in COVID-19 infection by SARS-Cov-2.

Introduction

COVID-19 epidemic started in December 2019 in Wuhan (China), and soon spread throughout the world [1]. The causative agent was identified as the Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2), closely related to the human SARS betacoronavirus (~80% homology) and MERS-CoV (Middle-East Respiratory Syndrome) betacoronavirus (~40% homology). The first western country affected was Italy, where, after the identification of the first Italian patient on February 21, 2020, the epidemic expanded very rapidly. The government decided that public health should have priority and declared the lockdown (March 11, 2020) to minimize transmission of COVID-19 through social distancing. Schools, travels and commercial activities have been shut down, except some essential services.

It was immediately clear that the impact of the infection was closely related to the demographic composition of the population because deaths were concentrated at older ages [2, 3]. Furthermore, COVID-19 fatality rate appears to be higher in men than in women all over the world [4].

Official bulletins and literature reported percentages of infected and dead patients, stratified data by age and sex, but referred to the total of infections. Some authors suggested to calculate this ratio using the population size stratified by age and sex as the denominator to discriminate specific infection incidence and fatality rates for each class, especially when the rate of the disease is highly variable by age and sex as in the SARS-CoV-2 infection [5].

Populations show different demographic profiles: developing countries have a higher proportion of younger people (like China), while in industrialized countries, like Italy, the share of older people is larger than the younger one. Also sex ratio is often age-dependent: generally there are more women among the elderly since they have a longer life expectancy, but at birth the ratio is reversed, with more surviving boys than girls [6].

Frequently diseases manifest differently in men and women. These differences of symptoms, diagnosis, severity, duration of the disease involve biological sex (genetic, anatomy, hormones and physiology) as well psychological and cultural behavior (ethnicity, social, and religious background, environmental and lifestyle) changing throughout the life course [7]. The concept of gender is controversial; it is often associated to the socially constructed roles and behaviors or referred to a complex interrelation and integration of both biological and cultural aspects. The “gender” medicine considers both aspects and represents the first step to the “precision” medicine.

The World Health Organization (WHO) pays particular attention on gender differences and their impact on diagnosis to shed light on the risk factors causing disease in men and women and address the corresponding specific treatments [8].

Here we analyze the distribution of the SARS-CoV-2 infected patient-cases in the Italian population stratified by age and sex to evaluate the gender impact on COVID-19 in each age class.

Methods

Data sources.

The most recent detailed data (January 1, 2019) on resident population in the Italian municipalities are available from the National Institute of Statistics of Italy (ISTAT) [9]. Although these data are more than a year old, the actual population underwent only a contraction of less than 0.19% [10]. We pooled these data, stratified by sex, into ten years' groups. The actual demographic structure of Italian population (Fig.1 and Supplementary Table S1) shows a slight deviation of male/female ratio before the age of 40 in favor of a slightly higher male component. After the age of 40 the percentage of men decreases slowly until the older ages when it is reverted (men/women odd ratio= 0.65 in 80-89 years class).

Data for people infected by the SARS-CoV-2 virus in Italy were extracted from official bulletins of the Italian “Istituto Superiore di Sanità” (ISS) from March 12 to May 20, 2020 [11]. The first bulletin with sex-disaggregated data was on March 12, 2020. Data are combined into ten years bins. Every bulletin was an upgrade of the total cumulative number of infected cases at the date of publication.

To our knowledge, few publications on COVID-19 take into account the different age classes, just mentioning the population structure, but do not really include in the statistical analysis the population profile for each class [12].

Data Analysis.

Positive patient-cases (pos) in the age groups (0-9, 10-19, ... , 90+) were extracted from official data sources detailed above, while negative patient-cases (neg) were calculated as the difference between total Italian population and positive patient-cases in the corresponding age group. Significance of odds ratio (OR) Male versus Female = $(M_pos/M_neg)/(F_pos/F_neg)$, was evaluated by the Fisher exact test on 2x2 contingency tables. Lethality was computed as the ratio (num_deaths/num_positive). We used

logistic regression to test the combined effect of age and sex on infection and death. Models: $\text{Logit(Positive)} \sim \text{Sex} + \text{Age} + \text{Sex}:\text{Age}$; $\text{Logit(Dead)} \sim \text{Sex} + \text{Age} + \text{Sex}:\text{Age}$. The Positive and Dead response variables are encoded as follows: positive=1, negative=0 and dead=1, alive=0 respectively. The Age predictor corresponds to the age groups {0-9, 10-19,..., 90+} discretized as {1,2,..., 10}, the Sex variable {M, F} is binarized as M= -1, F= +1 respectively. Dataset was randomly split: 70% to train the models, 30% as independent validation set. Area under the curve (AUC) of ROC curve computed on the validation set was used as a measure of model goodness-of-fit. Statistical analyses were performed using R-Bioconductor. Plots and charts were created using R-Bioconductor and Microsoft Excel.

Results

Ratio of infection between sexes evaluated on population structure.

A thorough investigation on COVID-19 epidemic needs age- and sex-disaggregated data, normalized to the total population size of the corresponding class (age group and sex) [5, 13].

This is necessary to disaggregate also the epidemic indexes (incidence and fatality) to really understand how age and sex contribute to infection.

An example of two different results is reported in Table 1, from the official COVID-19 bulletin of March 23, 2020. The data were stratified by age and sex, but the percentages of affected cases were obtained using in A) total patient-cases for ages as denominator while in B) total population size for ages as denominator, as suggested by Bhopal [5]. The results obtained in B) are more informative on the real impact of the infection in each class of age and sex.

Random versus non-random virus spreading.

Epidemic data analysis in Italy is complex due to the evolution of the infection and the outbreak of epidemic in the last period. In the first period, probably from the end of December 2019 to the end of February 2020, the infection was not identified as caused by Coronavirus 2 (CoV-2), so the virus was free to circulate and infect people without any constraint [14]. When the first COVID-19 Italian patient was identified, government declared complete lockdown to the whole country (March 11). Lockdown imposed social distancing to minimize an uncontrolled transmission of COVID-19. Two additional events characterized the Italian epidemic after end of March: about 60% of Italian cases came from Lombardy region where many patients contracted infection in hospital or emergency room. Moreover, at the end of March the infection spread also in the Elderly Care Homes (Italian RSA). It is likely that in these locations elderly people represented the majority of subjects. In this view it is possible to consider that until March 11 the epidemic diffused randomly, while after March 11 (complete lockdown) there was a non-random diffusion of the virus both among the younger age classes (school closure) and the older ones (infection in hospitals and RSA). It is to note that the average virus incubation is around 14 days and the infection has a duration between 30-40 days, depending on the severity of the symptoms, therefore the effects of the random virus circulation could be clearly detected until the end of March [15]. For these reasons all

ISS official bulletins until March 26, 2020, can be considered equally informative to analyze the differential spreading of the infection between age and sex classes of population, because the epidemic was essentially driven by an exponential diffusion of the infection without any further confounding factors, which came into play after that date. Figure 2 shows the OR Male/Female for cumulative incidence data based on the population structure stratified by age: all data points are statistically significant, except for the younger age classes, where the few recorded cases led to a poor data sampling. The corresponding OR for new cases are reported in Supplementary Fig.S1, with essentially the same trend. The evolution of the epidemic in Italy is confirmed when computing the ratio of Male/Female incidence and cumulative incidence in the same time points (Supplementary Fig. S2, S3). In all figures a visual separation between the two phases (before and after March 26, 2020) is highlighted by a blank (Supplementary Tables S2-S6 for details of statistical testing results, plotted in Fig. 2, 3, Supplementary Fig. S1-S3 respectively). The choice of March 26 as a separation time-point, between random and non-random virus circulation, is also justified by the correlation pattern of male/female ratios, where a kind of phase transition can be observed between March 26 and March 30, 2020 (Supplementary Fig. S4).

Data from March 12 to March 26, 2020: random virus circulation. Cumulative data collected from March 12 to March 26, 2020, during the exponential phase of the epidemic, shows that there are slightly more infected boys than girls in the 0-9 age class and more men than women in the 50-59 age class. A large difference in OR is present in the age classes from 60 to 90+ where the percentage of infected men is about double than women of the same age. In the 10-19 and 30 to 49 age classes both men and women are equally likely infected, while the opposite occurs in the 20-29 class, in which the percentage of sick women was significantly higher than men. All ORs for this range of time, except 10-19 age class, are statistically significant (FDR corrected p-value <0.05, see Fig. 2 and Supplementary Fig. S1, S2).

Data from March 30 to May 20, 2020: non-random virus circulation. The data collected from March 30 to May 20, 2020, show that ORs changed particularly within the elderly. While age classes 0-19 and >50 gradually reach the OR=1.0 (same percentage of infected men and women) the age classes from 30-49 inverted their OR values. It is to note that in the 20-29 age class the OR remains in favor of a lower percentage of infected men across the analyzed period.

Gender fatality rate (risk of death).

To analyze the effects of the infection on deaths stratified by age and sex, we report the fatality rate computed from official data (Fig. 3). In this report, we do not consider the mortality rate (i.e., the ratio to the whole population). The statistical analysis of cumulative data shows that men suffer a significant higher fatality rate than women are at every age class, with no differences in pre- or post-lockdown period (Fig. 3). Data for age classes 0 to 29 are often too under-sampled for a reliable statistical analysis.

Differences between pre- and post-lockdown phase. We analyzed data on March 26 and May 20, 2020 by logistic regression to highlight the different effects of Age:Sex interaction on infections and deaths by COVID-19 before and after the lockdown (Table 2). The effects of Age and Sex predictors and their

interaction are largely significant, except for the interaction Sex:Age on Deaths at March 26: this suggests that before the lockdown the sex difference was only weakly dependent on the age (Fig. 4).

Discussion

In this paper we analyze shares of men and women infected by SARS-Cov-2, normalized on the demographic structure of the population, to investigate differences in sex and gender infection incidence.

The first Italian patients were identified in Italy at the end of February 2020 and soon after the lockdown was declared. Observing the OR Male/Female of patient-cases stratified by age (Fig. 1), it is possible to follow the trend during epidemic from the beginning to May 20, 2020 and the differences between age classes. Overall, the OR shows a different evolution in time for each age class due to different factors: endogenous (biological sex and age physiology), and exogenous (sociality, habits, job and lifestyle). As previously detailed, we assume that infection data were determined by a random spread of the virus until March 11, 2020, with detectable consequences until two weeks after, and by a non-random virus circulation after that date. For this reason, we analyzed separately the two periods: until March 26, 2020 (hereafter named "pre-lockdown") and the second one until May 20, 2020 (named "post lockdown").

We first analyze the differences between pre- and post-lockdown periods, and subsequently, we discuss the gender issue during the pre-lockdown period.

Male/Female infection ratio in each age class during epidemic.

Age class 0-9: the OR was at first slightly in favor of more infected boys than girls, but then decreased towards 1.0 (boys and girls tended to be infected at the same level). This change was likely due to habits for boys of staying outside in larger groups more often than the girls and the propensity to practice team sports like football: during lockdown they had to reduce these physical contacts.

Age class 10-19: the value of OR remains essentially the same in pre- and post-lockdown. As for teenager peer groups are essential to social and general development and the social distancing imposed by lockdown did not change the infection ratio, we suppose that this was mostly due to biological factor (sex, hormones).

Age class 20-29: maybe the most interesting. Indeed, the OR remains constantly significantly in favor to more infected women than men during the whole reporting period, likely for biological reasons that were not affected by and thus independent from a different lifestyle during lockdown.

Age classes 30 to 49: data started with an OR~1.0 to then reach values significantly <1.0, towards a larger proportion of infected women.

Age class 50-59: the trend is similar to the previous one, though with starting value of OR significantly >1.0.

It is possible that this increase of the middle age infected women in post-lockdown was due to a larger portion of women that are healthcare-worker, compared to men (84% for European region) that were selectively infected in RSA[16]. In Italy on April 28, 2020, 69% of infected healthcare-workers were women [17].

Age classes >60: a huge difference in OR men/women is present in the pre-lockdown period, with men infected even more than twice than women. The trend shows a rapid evolution towards lower values after lockdown, particularly for people in the age class 80-89, where, at the end, men and women appear equally likely infected and, particularly for people aged 90+ with a clearly higher proportion on infected women. We suppose that the rapid change toward a larger incidence in women was heavily due to a non-random diffusion of the virus in the elderly. At the end of March, the virus diffused in hospitals often through emergency room and in the RSA, in which older people are admitted (usually age 65 and over). This happened in a framework where women account for the majority of population (Fig. 1), male/female population sex-ratio decreases from 0.92 to 0.37 in 60-69 to 90+ age classes respectively, and in such older Italian population cohort about 56% are women over 65 years old (63% for age >80) (Supplementary Table S1). ISS official bulletins from April 23 to May 20, 2020 report that RSA host 40-60% of new diagnosed cases with a large majority of women.

Differential male/female risk of death.

Scientific literature and official web sites on COVID-19 report a higher fatality rate in men than women are across the whole age spectrum [18]. Our statistical analysis confirms this difference in fatality rate and shows that this trend does not changes between pre- and post-lockdown period and even increases a bit (Fig. 3) although in some age classes women are more likely infected than men. To highlight this aspect we choose two representative dates of pre- and post-lockdown phases for disease incidence and death cases: March 26, 2020, the end of period with data reflecting the exponential phase of the infection spread, and May 20, 2020, for the post-lockdown phase. Figure 4 shows the cumulative counts of patient-cases and deaths in these two time points. It is to note that, although the incidence of infection increases in women from March 26 to May 20, 2020 (Supplementary Fig. S2), the fatality rate remains higher in men than women (Fig. 3) probably because the COVID-19 disease in men tends to be more severe than in women [19].

Analysis of infection incidence in exponential virus spreading.

As discussed above, SARS-CoV-2 virus spread freely, with an exponential growth of infection until March 11, 2020. After lockdown the virus was characterized by a forced non-random spreading due to social distancing, Thus the weeks between February and the end of March allow to analyze the virus spreading in the different age and sex classes of population, with minima confounding effects (dates March 12-19-23-26, 2020 in Fig. 2).

We should consider several aspects possibly involved in sex-specific response to SARS-CoV-2.

Female gender is known to have a stronger immune response to viral infections compared to male gender, due to more robust innate and adaptive immune responses [7]. Immune system is under the sex hormones influence: as a general rule, estrogens promote both innate and adaptive immune responses, which result in a better and faster response to pathogens. Instead androgens have an immune suppressive effect which may explain the greater susceptibility to infectious diseases observed in men, but a minor incidence of asthma in boys than girls [20]. X-chromosome inactivation in women causes an imbalance of genes involved in immune response, such as CD40L and TLR7, and ACE2 that exert a protective function in pathologies like hypertension, cardiovascular diseases and acute respiratory distress syndrome, which are concurrent conditions representing a major risk of worse prognosis in COVID-19 [7, 21]. ACE2 expression is higher in young people than in elderly individuals and higher in women than in men [22]. Other factors may be involved in infection such as hormone-regulated expression of genes or environmental factors like smoking, drinking and personal care [23]. Chronic activation of innate and adaptive immune functions increases with age and leads to a decline of the immune system response, causing a greater sensitivity to infections and chronic diseases [24].

On the basis of these few elements, we try to explain our results.

The 0-9 age class presents a larger proportion of infected male children. This result could be due to a better response of female gender immune system and higher level of ACE2 in female than male gender.

In the 10-19 age class, sex hormones change the OR bringing it closer to 1.0. Although estrogens give a better immune performance than androgens, it is possible that other unknown mechanisms occur, possibly linked to sex maturation. These mechanisms may be crucial since in the 20-29 age class, when the sexual identity is fully defined, the OR is significantly inverted, with clearly more infected women than men. This is a relevant result also because it is reported in foreign European and extra-European countries [25]. It is to note that no environmental factors like social distancing due to lockdown change the OR in the 20-29 age class during the epidemic (May 20, 2020, last date reported in this work).

The population aged 30-49 shows a more balanced ratio of infected man and women, though still with a slightly larger women proportion.

The population aged 50-59, and mainly from 60 to 90+ shows a larger presence of infected men, up to more than two-fold than women,

In the older age, the presence of overexpressed X-linked genes might reveal its greater importance. Women are more protected by infection by their strong immune system, and they have higher level of ACE2 than men. In addition, women in midlife and over usually assume Vitamin D to help the menopausal transition and to prevent osteoporosis due to the lack of estrogens, while in older men vitamin D deficiency is not adequately evaluated. It is known that adequate vitamin D levels would contribute to reduce inflammation and acute respiratory tract infection. Many authors speculate that, among other factors, the different susceptibility to virus infection may be the consequence of lower vitamin D levels in men in their sixties, compared with age-matched women [26].

Conclusion

This is the first preliminary study investigating the role of sex and gender in SARS-CoV-2 infection, related to the population structure from infancy to elderly.

The sex and gender approach to health (gender medicine) is an innovative discipline emerging from new considerations of biological (sex specific aspects) and gender (sociocultural, environmental habit) differences in predisposition and manifestation of disease. These multifactorial domains interact in health and diseases and evolve during ageing.

The SARS-CoV-2 epidemic highlighted a huge difference in age and gender virus impact both in incidence of cases and in fatality rate. The severe effects of this epidemic underline the need to understand the different sex reaction to viral agent causing the disease. Our results suggest complex modulations of the response to infection, specific for age and sex groups and indicate new perspectives in future gender research to understand the disease mechanisms and develop suitable therapies.

References

1. Paules, C.I., Marston, H.D & Fauci, A.S. Coronavirus infections—More than just the common cold. **323**, 707–8 (2020).
2. Dowd, J.B., et al. Demographic science aids in understanding the spread and fatality rates of COVID-19. *Proceedings of the National Academy of Sciences*. **117**, 9696-8; <https://doi.org/10.1073/pnas.2004911117> (2020).
3. Dudel, C., Riffe, T., Acosta, E., van Raalte, A.A, & Myrskylä, M. Monitoring trends and differences in COVID-19 case fatality rates using decomposition methods: Contributions of age structure and age-specific fatality. Preprint at <https://www.medrxiv.org/content/10.1101/2020.03.31.20048397v2> (upgrading on May 18, 2020).
4. Global Health 50/50 at The UCL Centre for Gender and Global Health. <https://globalhealth5050.org/covid19/sex-disaggregated-data-tracker/> (Accessed May 26, 2020).
5. Bhopal, R. Covid-19 worldwide: we need precise data by age group and sex urgently. *BMJ*. **369**, m1366; <https://doi.org/10.1136/bmj.m1366> (2020).
6. Ritchie, H. & Roser, M. - "Gender Ratio". 2019. Published online at OurWorldInData.org. Retrieved from: ' <https://ourworldindata.org/gender-ratio> ' [Online Resource] (2019).
7. Bartz, D., et al. Clinical Advances in Sex- and Gender-Informed Medicine to Improve the Health of All: A Review. *JAMA Intern Med*. **180**, 574–83; <https://doi.org/10.1001/jamainternmed.2019.7194> (2020).
8. World Health Organization:WHO. https://www.who.int/topics/womens_health/en/.
9. ISTAT (Istituto Nazionale di Statistica of Italy). <http://demo.istat.it/pop2019/index.html>. (Accessed March 15, 2020).

10. ISTAT (Istituto Nazionale di Statistica of Italy). <https://www.istat.it/en/archivio/238460>. (Accessed March 15, 2020).
11. ISS (Istituto Superiore di Sanità of Italy). <https://www.epicentro.iss.it/coronavirus/sars-cov-2-sorveglianza-dati> (Accessed May 26, 2020).
12. Gujski, M., et al. Epidemiological Analysis of the First 1389 Cases of COVID-19 in Poland: A Preliminary Report. *Med Sci Monit.* 26, e924702 ; <https://doi.org/10.12659/MSM.924702> (2020).
13. Global Health 50/50 at The UCL Centre for Gender and Global Health. <https://globalhealth5050.org/mission-vision-values/> (Accessed May 26, 2020).
14. Stefanelli, P., et al. Whole genome and phylogenetic analysis of two SARS-CoV-2 strains isolated in Italy in January and February 2020: additional clues on multiple introductions and further circulation in Europe. *Euro Surveill.* 25, 2000305; <https://doi.org/10.2807/1560-7917.ES.2020.25.13.2000305> (2020).
15. Zhou, B., She, J., Wang, Y. & Ma, X. The duration of viral shedding of discharged patients with severe COVID-19. *Clin Infect Dis.* ciaa451; <https://doi.org/10.1093/cid/ciaa451> (2020).
16. WHO report on Gender equity in the health workforce: Analysis of 104 countries Health Workforce Working Paper 12019. https://www.who.int/hrh/resources/gender_equity-health_workforce_analysis/en/ (2019).
17. United Nation Women. <https://data.unwomen.org/resources/covid-19-emerging-gender-data-and-why-it-matters> (2020).
18. Global Health 50/50 at The UCL Centre for Gender and Global Health. <https://globalhealth5050.org/covid19/> (2020).
19. Jin, J-M., et al. Gender Differences in Patients With COVID-19: Focus on Severity and Mortality. *Public Health.* 8,152; <https://doi.org/10.3389/fpubh.2020.00152> (2020).
20. Sharifi, N. & Ryan, C.J. Androgen hazards with COVID-19. *Endocr Relat Cancer.* 27:E1-E3; <https://doi.org/10.1530/ERC-20-0133> (2020).
21. Asselta, R., Paraboschi, E.M., Mantovani, A. & Duga, S. ACE2 and TMPRSS2 variants and expression as candidates to sex and country differences in COVID-19 severity in Italy. *Aging (Albany NY).* 12, 10087-10098; <https://doi.org/10.18632/aging.103415> (2020).
22. Cheng, H., Wang, Y.& Wang, G.Q. Organ-protective effect of angiotensin-converting enzyme 2 and its effect on the prognosis of COVID-19. *J Med Virol.* 92, 726-30; <https://doi.org/10.1002/jmv.25785> (2020).
23. Gebhard, C., Regitz-Zagrosek, V., Neuhauser, H.K., Morgan, R. & Klein S.L. Impact of sex and gender on COVID-19 outcomes in Europe. *Biol Sex Differ.* 11, 29; <https://doi.org/10.1186/s13293-020-00304-9> (2020).
24. Márquez, E.J., Trowbridge, J., Kuchel, G.A., Banchereau, J.& Ucar, D. The lethal sex gap: COVID-19. *Immun Ageing.* 17, 13; <https://doi.org/10.1186/s12979-020-00183-z> (2020).

25. Global Health 50/50 at The UCL Centre for Gender and Global Health.
<https://globalhealth5050.org/covid19/age-and-sex-data/> . (Accessed May 26, 2020).
26. La Vignera, S., et al. Sex-Specific SARS-CoV-2 Mortality: Among Hormone-Modulated ACE2 Expression, Risk of Venous Thromboembolism and Hypovitaminosis D. *J. Mol. Sci.* **21**, 2948 (2020).

Declarations

Authors' contributions: E.M. devised the study, performed preliminary analyses, collected the reference literature and produced the first draft of the manuscript. Both authors collected epidemic datasets. I.A. performed the statistical analysis and prepared figures and tables. Both authors wrote the manuscript.

Funding: The study was not funded.

Conflicts of interest: The authors declare no competing interests (ICMJE declaration).

Ethics approval: Not applicable

Consent to participate: Not applicable

Availability of data and material: Data used in this paper were present on official public web sites. The authors elaborated these data and have final responsibility for the decision to submit the results for publication.

Code availability: Statistical analyses were performed using R-Bioconductor. Plots and charts were created using R-Bioconductor and Microsoft Excel.

Tables

age classes	(cases male)/(total cases per age)	(cases male)/(total cases per age)	(cases male)/ (cases female)	(cases male)/(total male per age)	(cases female) / (total female per age)	(cases male)/ (cases female)
	A	A	A	B	B	B
0-9	57.3%	42.7%	1.34	0.0069%	0.0055%	1.27
10-19	52.1%	47.9%	1.09	0.0062%	0.0072%	0.86
20-29	45.7%	54.3%	0.84	0.0306%	0.0390%	0.78
30-39	50.5%	49.5%	1.02	0.0559%	0.0555%	1.01
40-49	50.2%	49.8%	1.01	0.0784%	0.0768%	1.02
50-59	56.2%	43.8%	1.28	0.1371%	0.1024%	1.34
60-69	67.1%	32.9%	2.04	0.1977%	0.0890%	2.22
70-79	66.4%	33.6%	1.98	0.2735%	0.1169%	2.34
80-89	57.9%	42.1%	1.38	0.3532%	0.1662%	2.12
90+	35.3%	64.7%	0.55	0.3239%	0.2197%	1.47

Table 1. Estimates of Male/Female sex-ratio from data of official COVID-19 bulletin, Mar 23, 2020. * Data on infected cases from “Bollettino-sorveglianza-integrata-COVID-19”, Mar 23, 2020. (A) Percentage of infected Male and Female cases (referred to total cases) and their ratio, disaggregated by age class. (B) Percentage of infected Male and Female patient-cases (referred to total population) and their ratio, disaggregated by age class. The ratio Male/Female derived in (A) and (B) are different, since the ratio in (B) reflects the population structure.

Date	Logistic Model		Main predictors			Interaction
			Intercept	Sex	Age	Sex:Age
2020 Mar 26	A) Positive ~ Sex + Age + Sex:Age AUC=0.720 (95% CI 0.717-0.723)	coefficient	-8.7162	0.2379	0.3365	-0.0711
		p-value	<2e-16	<2e-16	<2e-16	<2e-16
		std error	0.0152	0.0152	0.0022	0.0022
	B) Dead ~ Sex + Age + Sex:Age AUC=0.712 (95% CI 0.709-0.715)	coefficient	-15.9827	-0.6692	0.9610	0.0030
		p-value	<2e-16	4.13e-10	<2e-16	0.808 n.s.
		std error	0.1071	0.1071	0.0124	0.0124
2020 May 20	C) Positive ~ Sex + Age + Sex:Age AUC=0.700 (95% CI 0.698-0.702)	coefficient	-7.5763	0.1282	0.3374	-0.0198
		p-value	<2e-16	<2e-16	<2e-16	<2e-16
		std error	0.0087	0.0087	0.0012	0.0012
	D) Dead ~ Sex + Age + Sex:Age AUC=0.692 (95% CI 0.690-0.694)	coefficient	-15.3822	-1.0455	1.0760	0.0737
		p-value	<2e-16	<2e-16	<2e-16	<2e-16
		std error	0.0518	0.0518	0.0059	0.0059

Table 2. Contribution to cases and deaths of Age:Sex combined factors. *We analyzed data on Mar 26 and May 20, 2020, by the following models: (A,C) $\text{Logit(Positive)} \sim \text{Sex} + \text{Age} + \text{Sex:Age}$; (B,D) $\text{Logit(Dead)} \sim \text{Sex} + \text{Age} + \text{Sex:Age}$. Data refer only to Covid-19 certified infections and deaths, as detailed in ISS bulletins. The Positive and Dead response variables are encoded as positive=1, negative=0 and dead=1, alive=0 respectively. The Age predictor corresponds to the age group {0-9, 10-19,..., 90+} discretized as {1,2,..., 10}, the Sex variable {M, F} is binarized with integer values {-1,1} respectively. The effect of Age and Sex predictors and their interaction, is largely significant, except for the interaction Sex:Age on Death at Mar 26, 2020. AUC for models, computed on independent sets, are expressed as: value (95% Conf Int)

Figures

Fig.

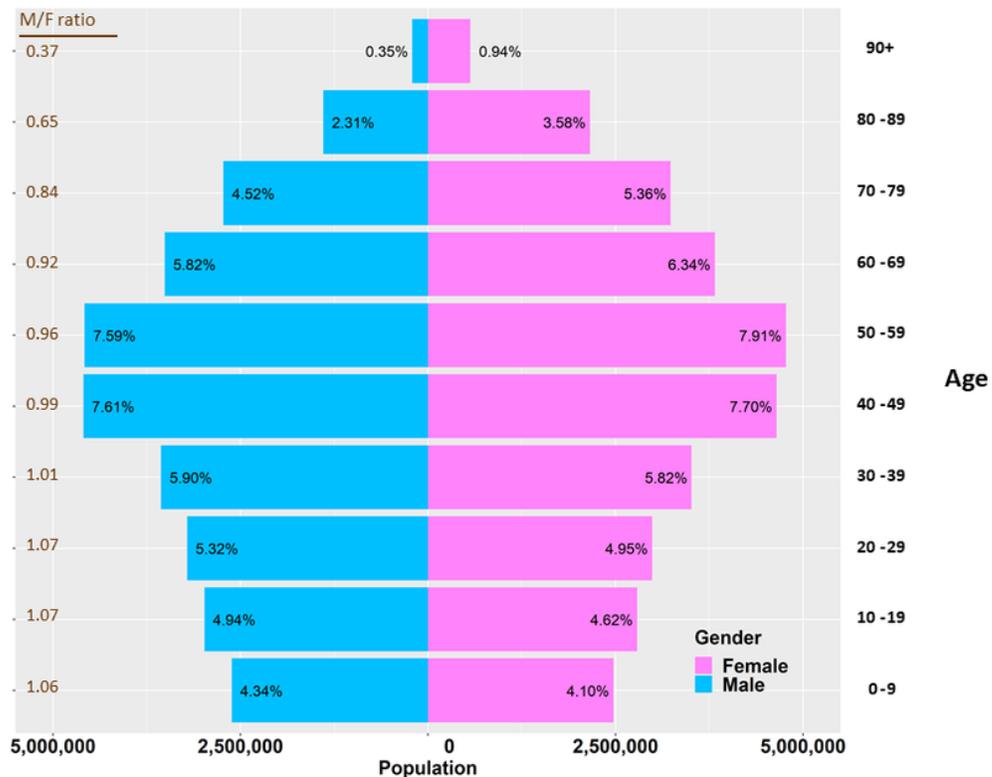


Figure 1

Italian population stratified by age and gender. The percentages refer to the whole Italian population on Jan 1, 2019 (see Methods). The ratios Male versus Female population are shown on the left side.

Fig. 2

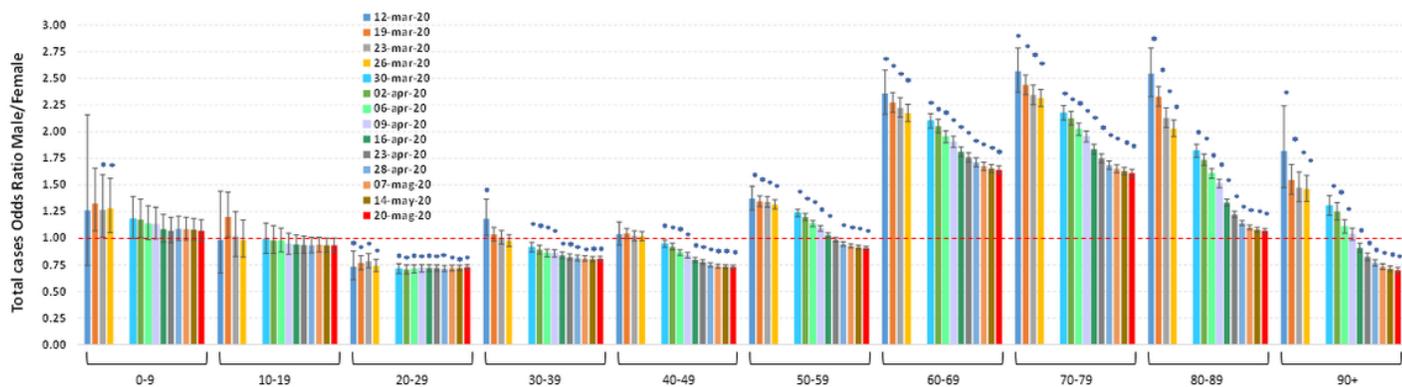


Figure 2

Odds Ratio Male versus Female of cumulative SARS-Cov-2 positive cases, stratified by age. The OR Male/Female was computed at all the dates (14 time points) for which official statistics were available: $OR = (Male_pos/Male_neg)/(Female_pos/Female_neg)$, where $_pos$ and $_neg$ refer to the total cumulative positive and negative cases on the corresponding date. The blanks between bars on Mar 26 and Mar 30, 2020, separate the first exponential phase of infection spread, and the post lockdown phase, whose effects start to be visible about two weeks after the official lockdown date (Mar 11, 2020). The statistical significance of OR (indicated by “*”) was assessed by the Fisher exact test; p-values were adjusted for multiple testing by FDR correction. The red horizontal line highlights the value $OR=1.0$ (Null hypothesis of no difference between Male and Female). Error bars correspond to standard 95% Confidence Interval of OR. Details of statistical testing and significance are reported in Supplementary Table S2.

Fig. 3

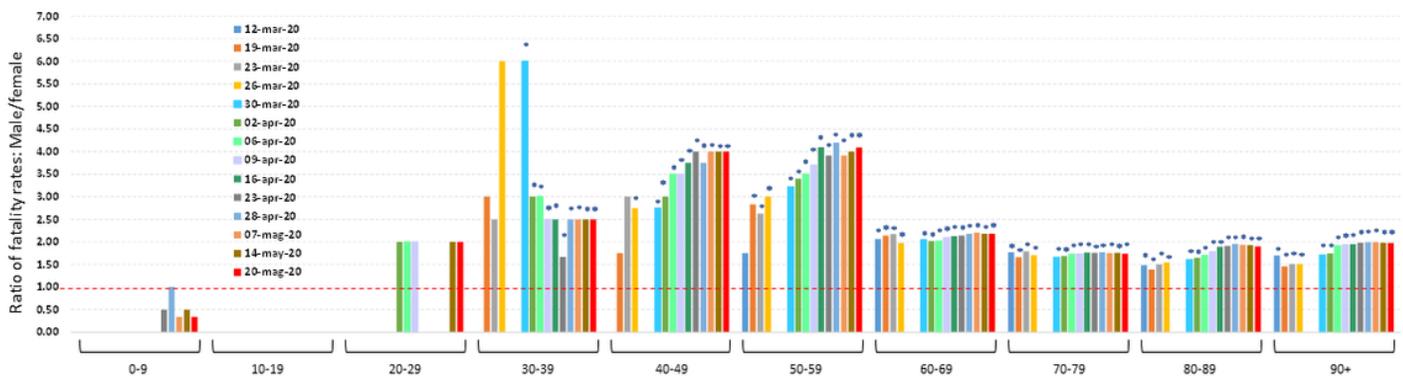


Figure 3

Ratio of fatality rates Male/Female by COVID-19, stratified by age. The ratio $Male_fatality/Female_fatality$ was computed at all the dates (14 time points) for which official statistics were available. Fatality at each time point is based on the cumulative positive cases and deaths counts. The blanks between bars on Mar 26 and Mar 30, 2020, separate the first exponential phase of infection spread, and the post lockdown phase, whose effects start to be visible about two weeks after the official lockdown date (Mar 11, 2020). The red horizontal line highlights the value $ratio=1$ (no difference between Male and Female). The statistical significance of OR (indicated by “*”) was assessed by Z-test for proportions; p-values were adjusted for multiple testing by FDR correction. The red horizontal line highlights the value

ratio=1.0 (Null hypothesis of no fatality rate difference between Male and Female). Details of statistical testing and significance are reported in Supplementary Table S3.

Fig. 4

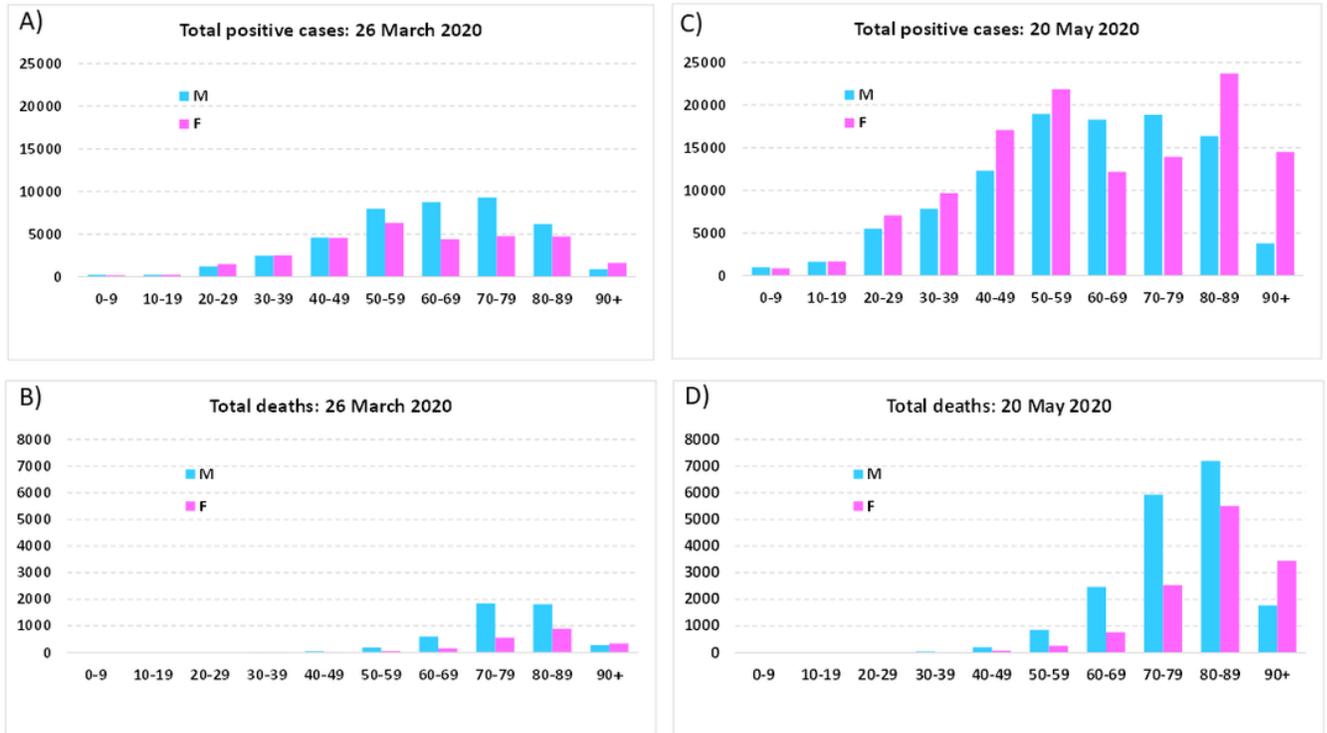


Figure 4

Barplots of data analyzed by logistic regression. The graphs correspond to the models A,B (Mar 26, 2020, positive cases and deaths) and C,D (May 20, 2020, positive cases and deaths) in Table 2, as indicated in the top left corner of each graph.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryLegends.pdf](#)
- [SupplFigures.pdf](#)
- [Supptables.pdf](#)