



OPEN

## Evolution of the lethality due to SARS-CoV-2 in Spain according to age group and sex

Pedro Saavedra<sup>1,3</sup>, Angelo Santana<sup>1,3</sup>✉, Alejandro Peñate<sup>2,3</sup>, Carol Herrera<sup>2,3</sup> & José-Miguel Pacheco<sup>1,3</sup>

The emergence of SARS-CoV-2 in China in December 2019 has posed a major challenge to health systems in all countries around the world. One of the most relevant epidemiological measures to consider during the course of a pandemic is the proportion of cases that eventually die from the disease (case fatality ratio, CFR). Monitoring the evolution of this indicator is of paramount importance because it allows for the assessment of both variations in the lethality of the virus and the effectiveness of the control measures implemented by health authorities. One of the problems with estimating the CFR in practice is that the available data only show daily or weekly counts of new cases and deaths; there is no information on when each deceased patient was infected and therefore it is not possible to know exactly how many cases there were at the time the patient became infected. Various approaches have been proposed for calculating the CFR by correcting for the time lag between infection and death. In this paper, we present a novel methodology to perform a non-parametric estimation of the evolution of the CFR by initially identifying an optimal time lag between infections and deaths. The goodness of this procedure is assessed by means of a simulation study and the method is applied to the estimation of the CFR in Spain in the period from July 2020 to March 2022.

### Abbreviations

AIC	Akaike information criteria
CFR	Case fatality ratio
IFR	Infection fatality ratio
ISCIII	Instituto de Salud Carlos III (Carlos III Health Institute)
RR	Rate ratio

At the beginning of March 2020, the severe acute respiratory syndrome coronavirus 2 (SARS-2-CoV-2) infection appeared in Spain and after a few weeks, in almost all of Europe. After a decrease in the number of cases detected in July of that year, there occurred successive waves of infections and deaths attributable to the virus. One question that is still not clear is how the lethality of the virus has evolved in the successive waves. There are many factors influencing this evolution<sup>1,2</sup>, especially vaccination campaigns and mutations in the virus genome that have changed its virulence and lethality over time. Although it now appears that the end of the pandemic is closer<sup>3-5</sup>, the immunity provided by vaccines decreases over time, while the risk of new variants evading immunity remains still as an uncertain factor. Therefore, even when the pandemic is declared over, health systems must remain vigilant<sup>6</sup>. In this sense, many countries have developed disease monitoring systems by constructing databases in which multiple variables are collected that provide information on patient characteristics—sex, age, comorbidities, date of onset, known exposures, etc. Efficient methods have been developed for the selection of variables<sup>7-9</sup> that allow models to be developed to assess the risk of death and guide preventive policies.

During the course of a pandemic two main measures are used to assess the lethality of the virus<sup>10</sup>. The first is infection fatality ratio (IFR), which estimates the proportion of deaths among all infected individuals. The second is case fatality rate (CFR), which estimates the proportion of deaths among identified confirmed cases. The IFR is hard to estimate as it requires knowledge of the prevalence of the disease in the population, as well as the actual number of deaths that can be attributed to the disease<sup>11</sup>. While this number is readily available (at

<sup>1</sup>Department of Mathematics, University of Las Palmas de Gran Canaria, 35017 Las Palmas de Gran Canaria, Spain. <sup>2</sup>Faculty of Health Sciences, University of Las Palmas de Gran Canaria, 35016 Las Palmas de Gran Canaria, Spain. <sup>3</sup>These authors contributed equally: Pedro Saavedra, Angelo Santana, Alejandro Peñate, Carol Herrera and José-Miguel Pacheco. ✉email: angelo.santana@ulpgc.es

least in those countries with advanced health care systems), estimation of prevalence usually requires serological surveys, which are expensive and time-intensive<sup>12</sup> and are therefore not suitable for rapid response to variations in virus lethality.

The case fatality ratio (CFR) is easier to obtain, but the fact that many asymptomatic or mildly symptomatic people are not identified as cases means that the CFR tends to overestimate the severity of the disease. However Luo et al.<sup>13</sup> show that in locations that meet certain conditions (large-scale community transmission, extensive testing and without medical breakdown, as is the case of Spain) the CFR computed from cases detected by PCR (polymerase chain reaction) tests can be considered as a reliable indicator of the lethality of COVID-19. Traditionally, the fatality rate is assumed to be constant throughout the epidemic<sup>14</sup>, but this may not be true in rapidly evolving contexts<sup>15–17</sup> such as the current SARS-CoV-2 pandemic. The simplest estimate of real time CFR is to divide the cumulative number of deaths by the cumulative number of cases over a period of time (usually 1 day or 1 week). This estimate is known as the crude (or naïve) CFR<sup>18</sup>. There are a number of reasons why this estimator may be biased<sup>19,20</sup>, but one of the most obvious is that the persons who die on a given day were infected much earlier, and thus the denominator of the CFR should be the total number of patients infected at the same time as those who died<sup>21</sup>. But usually the statistical records available show only daily or weekly numbers of new cases and deaths; there is no information on when each deceased patient was infected and therefore it is not possible to know exactly how many cases there were at the time the patient became infected.

A number of procedures have been proposed to circumvent this problem by correcting the CFR value to take into account the time elapsed between infection and death. Wilson et al.<sup>22</sup> and Baud et al.<sup>21</sup> estimated corrected CFR values for China during February 2020 by dividing the number of deaths on a given day by the number of patients with confirmed COVID-19 infection 13 or 14 days before. These values were chosen based on previous studies showing that the median time from infection to death was in this range. Shim et al.<sup>23</sup> and Newall et al.<sup>24</sup> compute a time delay adjusted CFR for COVID-19 in South Korea up to June 2020 fitting a gamma or lognormal distribution to the survival time of patients using previous case studies. Manisha et al.<sup>25</sup> estimate the CFR for China and for countries outside China before April 2020 also using a 14-day delay between death and cases; but instead of simply dividing deaths by delayed cases, they also fit a regression line to these values and estimates the CFR as the slope of the regression, in an attempt to reduce the bias still present. Rothman et al.<sup>26</sup> calculated age specific time corrected crude symptomatic CFR values from 7 countries using two independent time to fatality correction methods; Thomas et al.<sup>18</sup> use what they call *time-shifted distribution (TSD) analysis method*. The CFR is obtained as the number of deaths in day  $t$  divided by the number of cases in day  $t - t_d$  where the value  $t_d$  is obtained as the one that minimizes the root mean squared error in the linear regression of deaths versus delayed cases for  $t_d = 1, 2, \dots, 25$ .

In this paper a novel method is proposed to estimate the weekly progression of SARS-CoV-2 lethality (CFR) by age group and sex when available data are only daily total counts of cases and deaths. Like Thomas et al.<sup>18</sup> we propose to approximate the number of deaths occurring among the cases registered in the week  $w$  by the number of deaths that occur  $\delta$  weeks later. The CFR is then estimated by using a log-additive model<sup>27</sup> selecting for  $\delta$  the value which minimizes the Akaike Information Criterion<sup>28</sup>. The goodness of this procedure is assessed by means of a simulation study based on actual data. The method is applied to the estimation of the CFR in Spain in the period from July 2020 to March 2022.

## Material and methods

**Data source.** All data has been obtained from the Carlos III Health Institute (ISCIII)<sup>29</sup>. This institute is the Public Research Organization of the Spanish Government responsible for funding and executing national biomedical research and depends on the Ministry of Science, Innovation and Universities of Spain, and it is also attached to the Ministry of Health, Consumption and Social Welfare. This institution maintains a panel of open data obtained through the epidemiological surveillance network of the National Epidemiology Centre, which can be downloaded from its website<sup>30</sup>. The data available are daily counts of reported COVID-19 cases, number of deaths, number of people hospitalised and number admitted to ICU, broken down by age group and sex.

Data on the number of infections in the first wave of the pandemic are not reliable as only highly symptomatic cases were detected at that time<sup>31</sup>. For this reason, in this study only data from week 20–26 July 2020 to week 14–20 March 2022 (87 weeks) have been considered. The study was carried out separately for three age groups, namely: “less than 50 years”, “from 50 to 69 years” and “70 or over years”. For each of these groups, the data used in the analysis are the weekly counts of infections and deaths by sex group. More specifically, they are of the form:

$$\{(Cases_{Sex,w}, Deaths_{Sex,w}) : Sex = Male, Female; w = 1, \dots, 87\}$$

where  $Cases_{Sex,w}$  and  $Deaths_{Sex,w}$  denote the count of registered SARS-CoV-2 cases and deaths by sex group and week.

**Statistical analysis.** For each age group, we denote by  $d_{Sex,w}$  the count of deaths among those subjects registered as SARS-CoV-2 cases at week  $w$ . These deaths will therefore occur over the next few weeks. We assume that:

$$d_{Sex,w} \sim \text{binomial}(Cases_{Sex,w}; p_{Sex}(w)) \quad (1)$$

where  $p_{Sex}(w)$  is the *actual Case Fatality Rate* at  $w$ , i.e. is the probability of dying for those subjects registered as SARS-CoV-2 cases in week  $w$ .

In addition, we assume for  $p_{Sex}(w)$  the additive log-model<sup>27</sup>:

$$\log(p_{\text{Sex}}(w)) = \alpha + \beta_{\text{Sex}} + s(w) \quad (2)$$

Here, “log” denotes the natural logarithm,  $\beta_{\text{Female}} = 0$  (Female sex is taken as reference) and  $s(w)$  is a *smooth* function of the week, which will be estimated nonparametrically using cubic splines.

The ISCIII data do not provide the values of  $d_{\text{Sex},w}$  and, therefore, the CFR  $p_{\text{Sex}}(w)$  cannot be directly estimated. However, Fig. 1 suggests that for each sex group there is a proportionality between the case count at week  $w$  ( $\text{Cases}_{\text{Sex},w}$ ) and that of deaths  $\delta$  weeks later ( $\text{Deaths}_{\text{Sex},w+\delta}$ ) for some delay  $\delta$ , leading to the fact that:

$$\text{Death}_{\text{Sex},w+\delta} \sim \text{binomial}(\text{Cases}_{\text{Sex},w}; \tilde{p}_{\text{Sex}}(w)) \quad (3)$$

Here we assume that  $\delta$  depends on sex but not on  $w$ . This parameter could be considered as the expected survival time of those cases who die.

In order to determine the delay  $\delta$  by sex group, we fit the model (3) for values  $\delta \in \{0, 1, 2, 3, 4, 5\}$  estimating  $\tilde{p}_{\text{Sex}}(w)$  nonparametrically by means of a cubic spline. For each  $\delta$  the *Akaike Information Criterion*<sup>28</sup>  $AIC(\delta)$  is computed which measures the lack of fit of the model. Finally, we consider as optimum the value of  $\delta$  that minimizes  $AIC(\delta)$ . Note that the value of  $\delta$  obtained depends on the sex group.

We can expect that most of the deaths among the cases registered in week  $w$  will occur on the days of week  $w + \delta$ . There will certainly be deaths among such cases occurring outside that period. Similarly, some of the deaths corresponding to cases registered in weeks close to  $w$  will occur in week  $w + \delta$ . If we assume that  $p_{\text{Sex}}(w)$  varies smoothly from 1 week to the next, then cases registered in  $w$  that die outside  $w + \delta$  are roughly offset by deaths in  $w + \delta$  corresponding to cases registered outside but close to  $w$ . Thus the smooth variation assumption of  $p_{\text{Sex}}(w)$  leads us to the approximation:  $d_{\text{Sex},w} \approx \text{Death}_{\text{Sex},w+\delta}$ , and therefore the CFR  $p_{\text{Sex}}(w)$  in the log-additive model (2) can be estimated by  $\tilde{p}_{\text{Sex}}(w)$  in model (3). We examine the goodness-of-fit of this method by a simulation study in the next section.

Once the models given in (1) are estimated for each age group, the rate ratios *Males:Females* (RR) are obtained as  $\exp(\beta_M)$  and estimated by means of 95% confidence intervals (CI). Progression of the CFR is expressed as *expected deaths by 10,000 infections-week* ( $10^4 \times p_{\text{Sex},w}$ ).

Statistical significance was set at  $p < 0.05$ . Data were analyzed using the R statistical language and environment, version 4.2.1<sup>32</sup>.

## Simulation study

As input for this study, the actual data of the number of cases corresponding to the cohort of males aged 70 years or older have been used. The probability of death among those registered at week  $w$  (*theoretical CFR*) has been modelled using a function  $p(w)$  with two peaks of lethality. This function is shown in Fig. 3.

The simulation proceeds as follows:

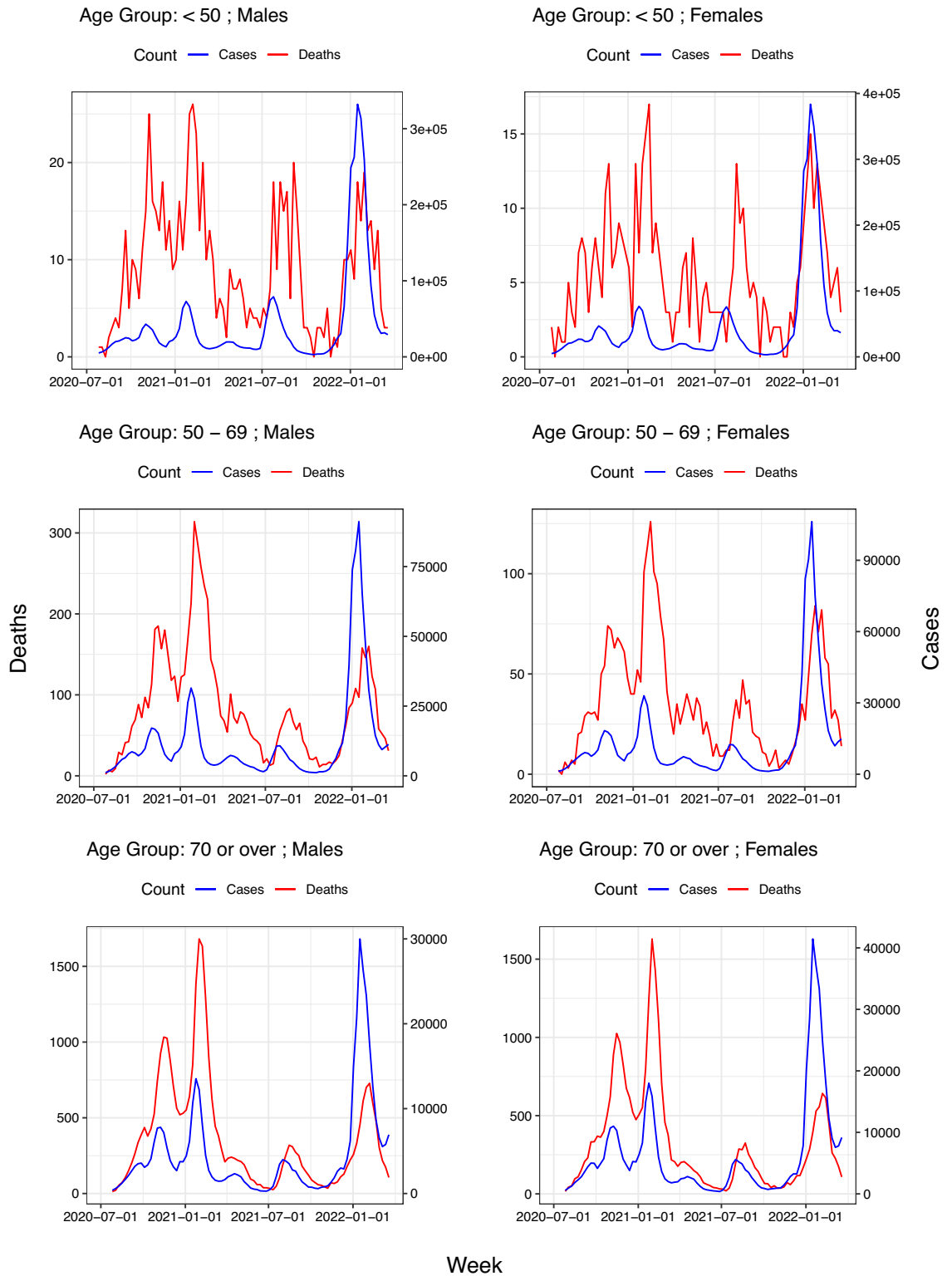
1. As the number of new cases in week  $w$  ( $\text{Cases}_w$ ) we use the number that actually occurred in Spain between 20–26 July 2020 and the week 14–20 March 2022 (87 weeks).
2. For week  $w$  in which  $\text{Cases}_w$  infections have been recorded, the number of those who will die,  $d_w$  is generated from a distribution  $\text{binomial}(\text{Cases}_w, p(w))$ .
3. Survival times for each patient are generated independently with a common Weibull probability distribution with parameters  $\lambda = 2$  and  $\kappa = 2.257$ . Thus, the expected survival time among those who are going to die is 2 weeks (the estimate for actual data is 1 week, but here we use 2 weeks because this increases the variance in survival time and spreads the number of deaths over longer periods, making the CFR more difficult to estimate).
4. Figure 2a shows the weekly counts of infections ( $\text{Cases}_w$ ) and the count of deaths ( $\text{Deaths}_w$ ) obtained in the simulation. Figure 2b show the same data but with deaths delayed by 2 weeks.
5. For values of  $\delta \in \{0, 1, 2, 3, 4, 5\}$ , we estimate the model  $\text{Deaths}_{w+\delta} \sim \text{binomial}(\text{Cases}_w, p(w))$ , being  $p(w) = \alpha + s(w)$  (we are considering only males).  $AIC(\delta)$  denotes the corresponding AIC value. Note that the optimal  $\delta$  coincides with the expected survival time (Weibull distribution).

Figure 3 plots the CFR estimated from these data in 100 simulations of the process, as well as the theoretical CFR, and shows that the procedure described produces a good estimate of the CFR used to simulate the data.

## Results

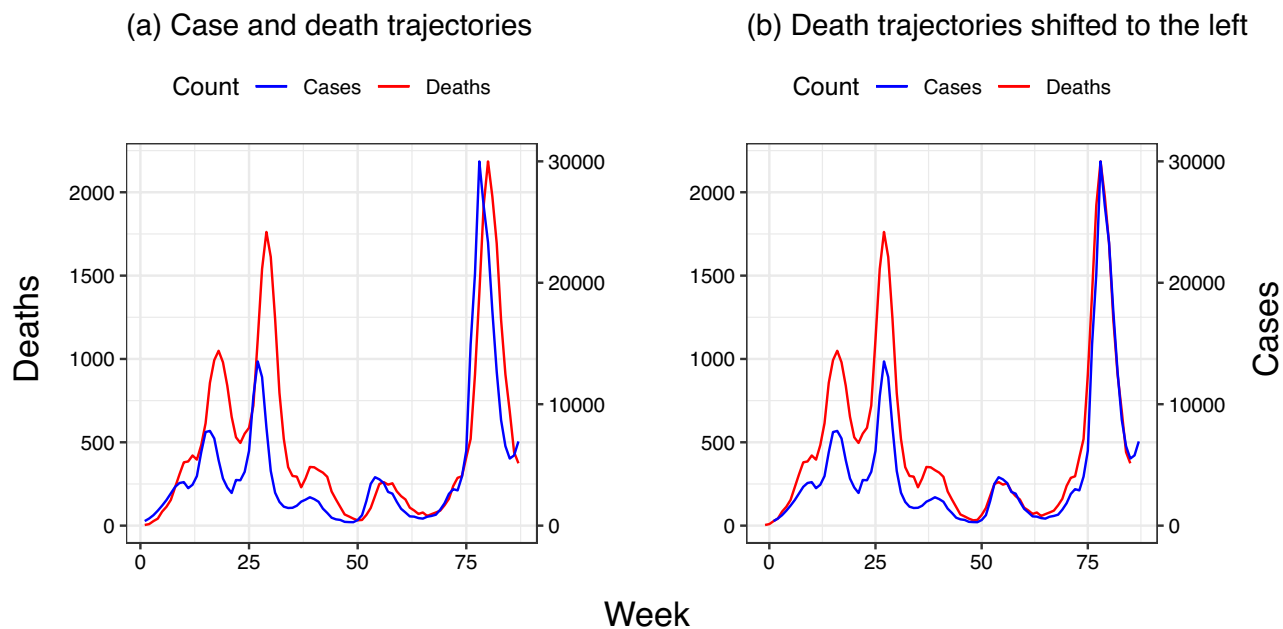
Table 1 summarizes the six cohorts of study (three age groups for each sex). For the entire follow-up period and the three age groups, the infection rates were similar by sex, but mortality rates were higher in men than in women. Figure 1 displays jointly the weekly counts of *cases* and *deaths* (note the different scales on the y-axis) for each age group and sex. Note that the trajectories of deaths exhibit a shift to the right relative to that of infections, which appears to be week independent. Such displacements would correspond to the *expected times of survival* among those patients who die. Figure 4 shows the values of AIC versus the different delays corresponding to model (3) and their corresponding optima  $\delta$ .

Note that the older the patients, the lower the  $\delta$  values. When the trajectories  $\text{Cases}_{\text{Sex},w}$  and the approximations of  $d_{\text{Sex},w}$  by  $\text{Death}_{\text{Sex},w+\delta}$  are plotted together (Fig. 5), it can be observed that they show to be *in phase*. Table 2 summarizes the estimation of the three binomial models, one for each age group. The estimated  $s(w)$  curves in the additive models (2) are shown in Fig. 6. It can be observed that for each of the three age groups three peaks were reached during the follow-up period. In addition, Fig. 7 displays the progress in the CFR rate

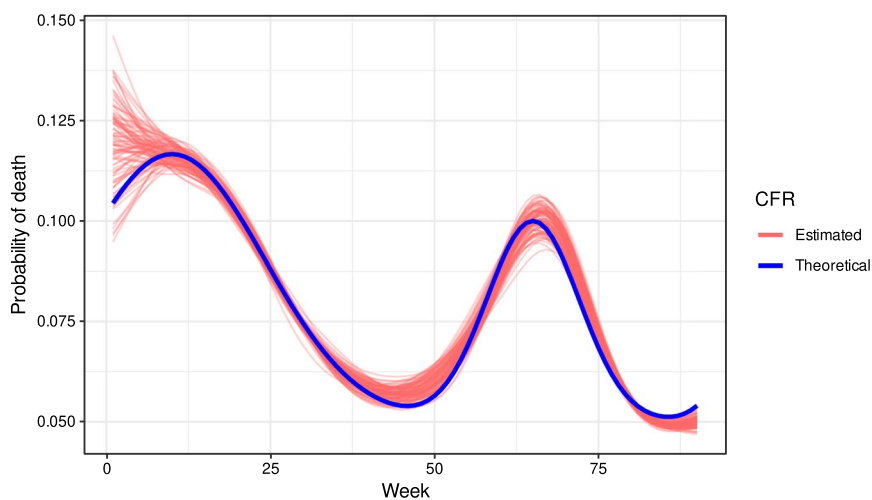


**Figure 1.** Joint plot of weekly cases and deaths counts by age group and sex. Note that the trajectories of the counts of deaths appear shifted to the right with respect to that of the cases and that such displacements depend on age group. The strong oscillation of the death counts in the group under 50 years of age is attributable to the low lethality rate since it increases the coefficient of variation of deaths.

(expressed as expected deaths by 10,000 persons-week) for each age group showing also the differences by sex as estimated by the model. In all groups, the CFR was significantly higher in men. For the group of patients under



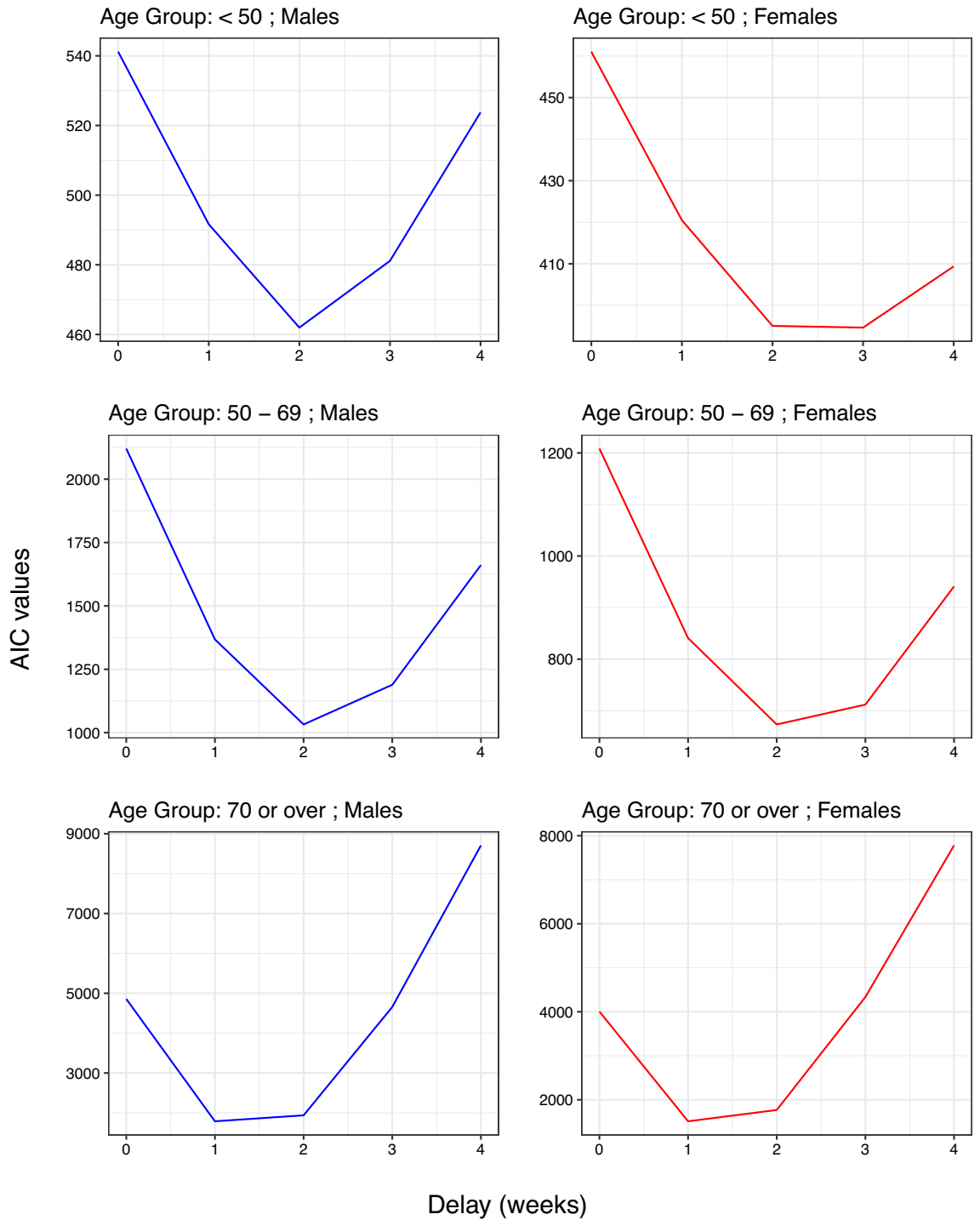
**Figure 2.** Simulated trajectories of cases and deaths. The expected time of survival among the death was 2 weeks and thus, the trajectory of death was delayed 2 weeks.



**Figure 3.** Simulation study: lethality rate (theoretical and estimated in 100 simulations).

Age (years)	Gender	Population	Infections		Deaths	
			Count	Rate	Count	Rate
< 50	Male	14,213,918	3,748,502	263.7	799	56.21
	Female	13,791,776	4,102,042	297.4	489	35.46
50–69	Male	6,124,088	1,053,728	172.1	7288	1190
	Female	6,384,636	1,168,291	183	3126	489.6
70 or over	Male	2,870,781	401,664	139.9	31,437	10,951
	Female	3,970,486	548,416	138.1	28,257	7117

**Table 1.** Summary of the study cohorts for the period July 2020–March 2022. Infection rate is shown as number of infected per 1000 people and mortality rate is calculated as number of deaths per million people in the population.



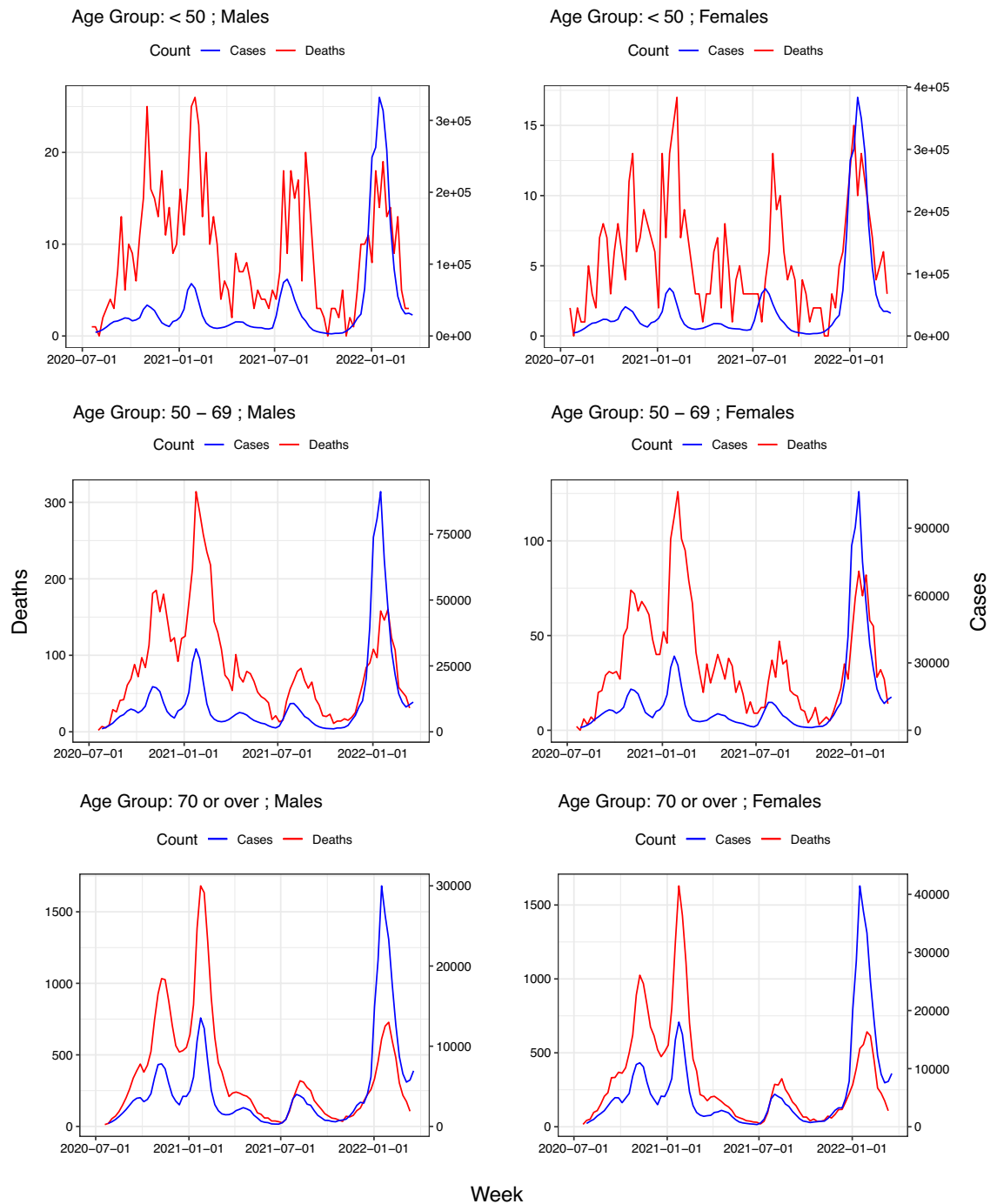
**Figure 4.** Values of AIC according to delay by age group and sex.

50 years of age, the CFR rate ratio of men vs. women was  $RR = 1.73$  (95% CI 1.55; 1.94); for those between the ages of 50 and 69, the rate ratio was the highest ( $RR = 2.49$ ; 95% CI 2.39; 2.60). Finally, for the group of 70 years or older, it was found that the CFR rate was 50% higher in men than in women ( $RR = 1.52$ ; 95% CI 1.49; 1.54).

**Discussion**

The results obtained show how SARS-CoV-2 case fatality ratio has evolved during the course of the pandemic. Once Spain reached a stage of stable extensive testing, this progression of lethality rates could be explained by new virus variants that emerged during the observation period (see Table 3), as well as by the gradual introduction of vaccination against SARS-CoV-2 from the beginning of 2021 (see Fig. 8).

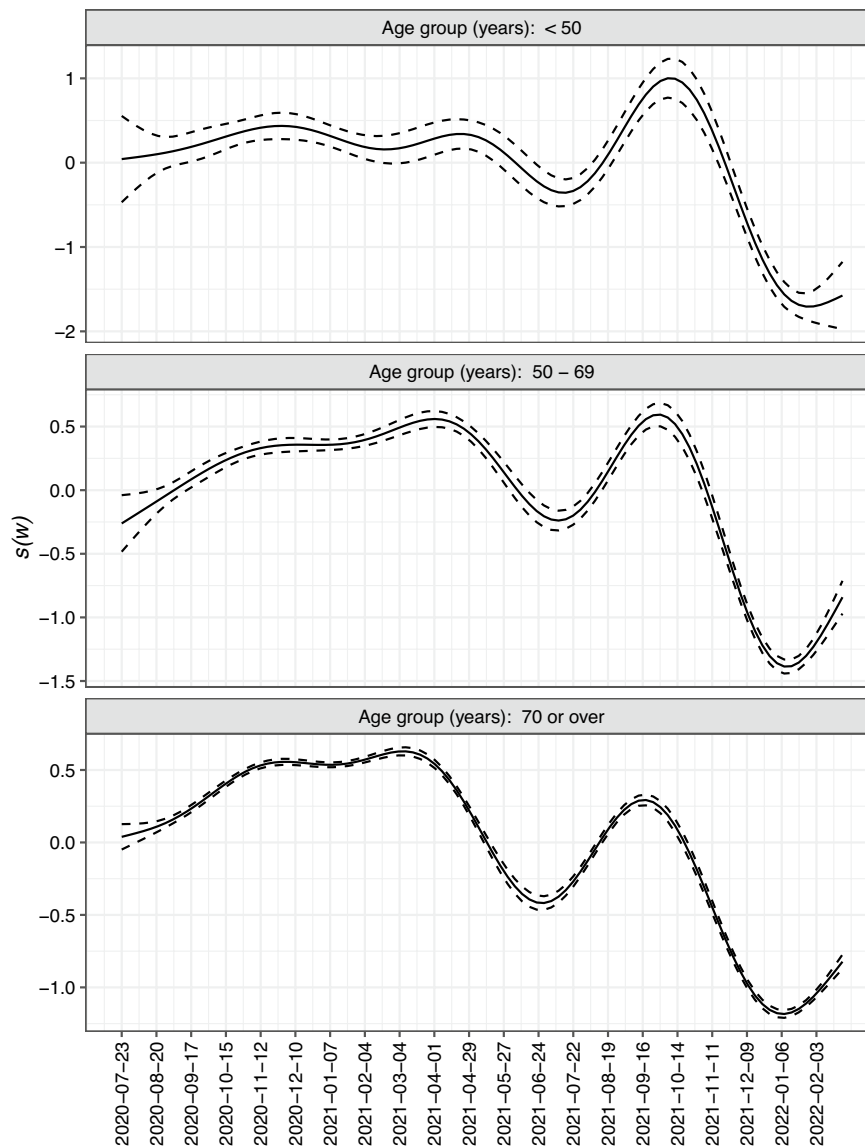
Figure 7 shows that CFR increases with age and is higher in men than in women for all age groups. These results are in line with those of<sup>33</sup>, who studies the lethality in the first wave of SARS-CoV-2 in Spain. This sex



**Figure 5.** Trajectories of the weekly counts of cases and deaths. When the death curves are shifted to the left the corresponding number of weeks, there tends to be a proportionality between infections and deaths; the divergences correspond to variations in the lethality rates.

disparity is likely explained by a combination of biological sex differences (differences in chromosomes and related sex steroids) and gender-specific factors such as differential behaviors. Men are more likely to engage in poor health behaviors such as smoking and alcohol consumption resulting in higher rates of pre-existing co-morbidities (hypertension, cardiovascular disease, COPD) associated with a poor COVID 19 prognosis<sup>34</sup>. Concerning the smoking habit, current smokers compared with never smokers have significantly upregulated ACE2 expression in the oral and lung epithelium. Given that smoking is more prevalent among males, higher expression of ACE2 due to this risk factor could explain the worse outcomes of SARS-CoV-2 in males<sup>35,36</sup>. Additionally, women have a higher antibody production and more efficacious vaccine responses overall. Furthermore, healthy females are known to have higher numbers of CD4+ T cells, greater CD4+/CD8+ ratios and an increased number of activated T cells, cytotoxic T cells, and B cells compared with males, resulting in a prompt response to the presence of infection<sup>34,37</sup>. Finally, low levels of testosterone in elderly men have been





**Figure 6.** Splines corresponding to the progress of the lethality rate (95% CI) according to age group.

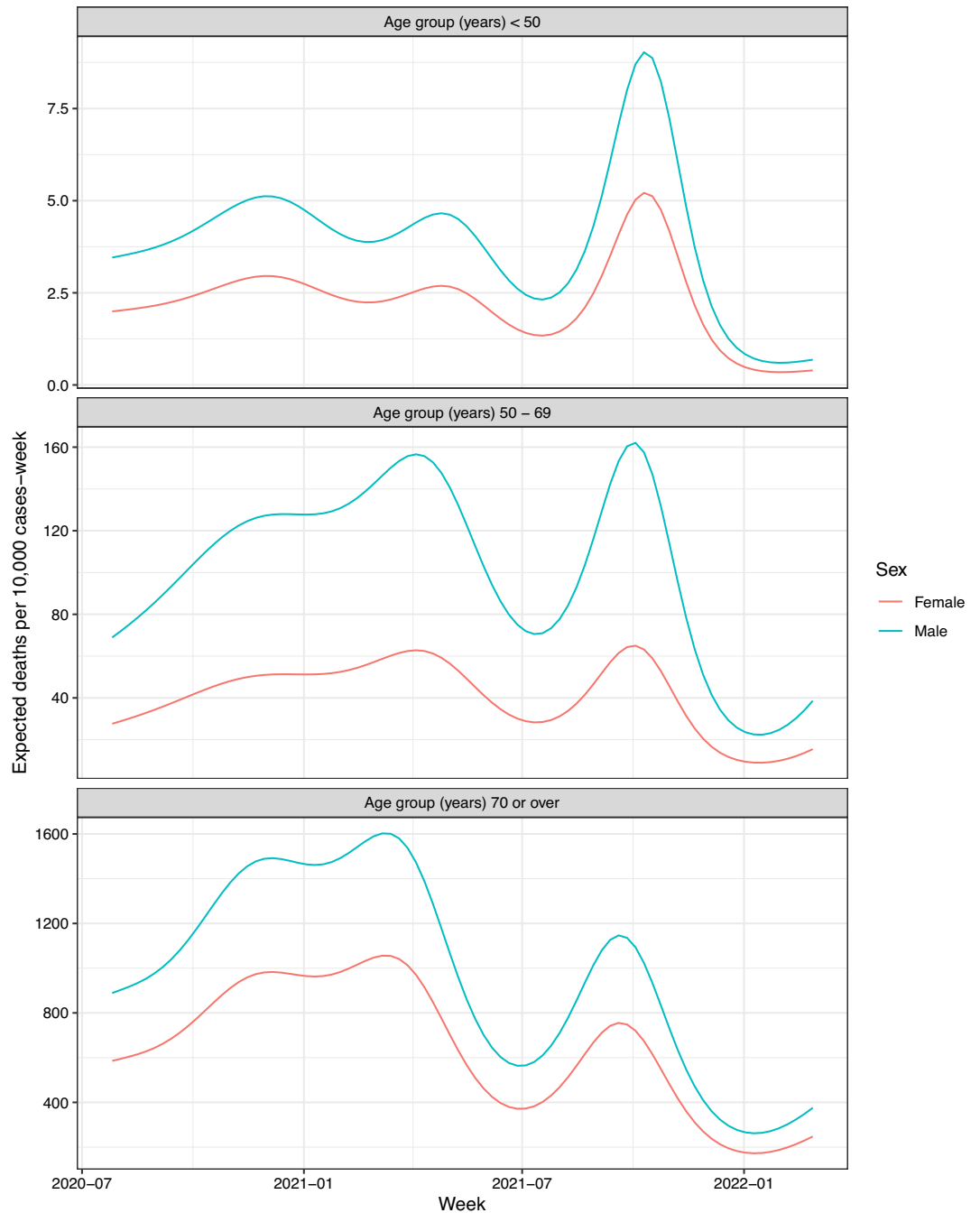
associated with upregulation of inflammatory markers and possible risk of lung damage, as well as respiratory muscle catabolism and increased need for assisted ventilation<sup>38</sup>. All of these factors could contribute to a poorer response to SARS-CoV-2 infection in men, but this requires further research.

**Limitations and future work.** The procedure described in this article estimates the CFR at the week  $w$  by assuming that for each sex and age group the number of deaths at  $w + \delta$  follows a *binomial*( $Cases_w; \tilde{p}(w)$ ) distribution with  $\tilde{p}(w)$  similar to  $p(w)$ , the true CFR. This approximation is valid when both the variation in the number of cases and the evolution of virus lethality occur smoothly. When a peak occurs in the number of cases the approximation does not work well. It can be seen in the simulation (Fig. 3) how the estimation of the function  $p(w)$  is worse around week 70, when there was a peak in the number of cases caused by the irruption of the omicron variant. Further refinements of this work are related with the fact that the CFR may have other biases<sup>19,20</sup> that we have not corrected for. Future work is required to improve the approximation at times when there is a rapid increase in the number of cases or mortality and to introduce corrections for other biases. It is also interesting to study how this CFR estimate can be used to improve estimates of the IFR, the other important lethality index, which is extensively used by health systems for pandemic management.

## Conclusion

The procedure we have developed in this paper to estimate the CFR during a pandemic aims to correct for the bias due to the time lag between the time of infection and the time of death. Previous works make this correction by calculating the CFR as the number of deaths at  $t + \delta$  divided by the number of cases at  $t$ . In those papers, several strategies are followed to estimate the appropriate value of  $\delta$ : using  $\delta$  values chosen subjectively by the researcher,





**Figure 7.** Progress in the lethality rate by age group and sex.

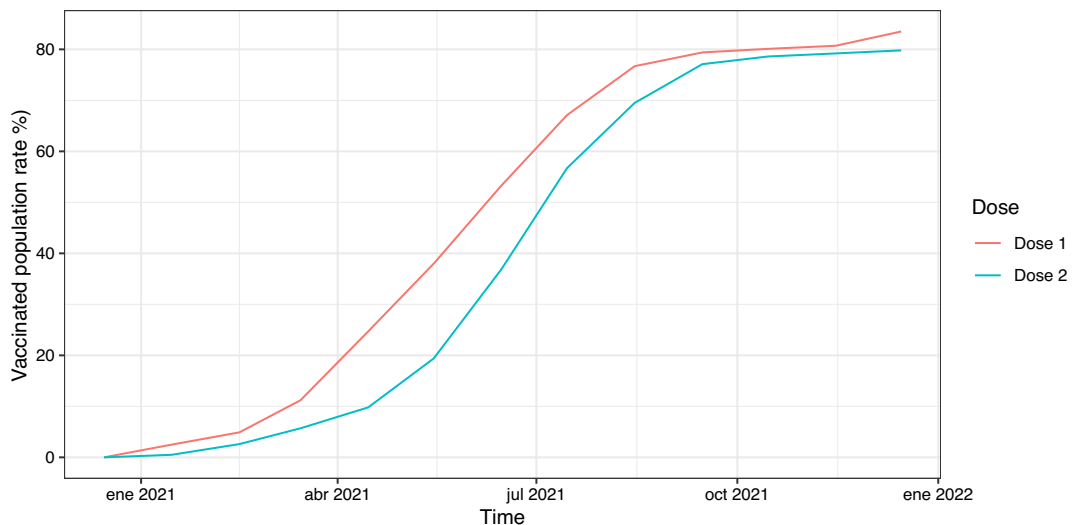
using clinical data to estimate the distribution of survival time, or using the similarity of the curves of cases  $C(t)$  and deaths  $D(t)$  to determine the value of  $\delta$  for which a better linear fit is achieved  $D(t + \delta) = \lambda C(t)$ . Our procedure improves this approximation by considering that  $D(t + \delta)$  follows for each sex and age group a binomial distribution of parameters  $C(t)$  and  $p(t)$ , where  $p(t)$  is estimated non-parametrically by means of a spline that allows us to directly obtain the evolution of the CFR. This procedure could be promising for decision-making in public health policies in a scenario in which the vaccination rate remains constant in a high percentage of the population, and changes in the CFR could basically be attributed to changes in virus lethality.

Age group (years)	Factor	Coefficient (SE)	p value	Rate ratio (95% CI)
< 50	Intercept ( $\alpha$ )	- 8.562 (0.049)	< 0.001	
	Sex			
	Female (Ref.)	0	-	1
	Male ( $\beta_M$ )	0.549 (0.058)	< 0.001	1.73 (1.55; 1.94)
	Week	Cubic spline	< 0.001	
50–69	Intercept ( $\alpha$ )	- 5.630 (0.020)	< 0.001	
	Sex			
	Female (Ref.)	0	-	1
	Male ( $\beta_M$ )	0.914 (0.021)	< 0.001	2.49 (2.39; 2.60)
	Week	Cubic spline	< 0.001	
70 or over	Intercept ( $\alpha$ )	- 2.876 (0.007)	< 0.001	
	Sex			
	Female (Ref.)	0	-	1
	Male ( $\beta_M$ )	0.417 (0.008)	< 0.001	1.52 (1.49; 1.54)
	Week	Cubic spline	< 0.001	

**Table 2.** Estimation of the binomial models by age group.

WHO level	PANGOLIN lineage	Detection	First case sequenced in Spain
Alpha	B.1.1.7	September 2020	December 2020
		United Kingdom	
Beta	B.1.351	May 2020	December 2020
		South Africa	
Gamma	P.1	November 2020	January 2021
		Brazil	
Delta	B.1.617.2	October 2020	April 2021
		India	
Omicron	B.1.1.529	November 2021	November 2021
		South Africa	

**Table 3.** Dates and places of first detection of new virus variants.



**Figure 8.** Progression of vaccination Rates in Spain.

## Data availability

The dataset supporting the conclusions of this article is publicly available through the web page of the Carlos III Health Institute (Spain), <https://cncovid.isciii.es/covid19/#documentación-y-datos>. R source code for the analysis is available at [www.github.com/angeloSdP/covidWeek](http://www.github.com/angeloSdP/covidWeek).

Received: 22 August 2022; Accepted: 2 December 2022

Published online: 21 December 2022

## References

- Clementi, N. *et al.* Lower nasopharyngeal viral load during the latest phase of COVID-19 pandemic in a northern Italy university hospital. *Clin. Chem. Lab. Med.* **58**, 1573–1577. <https://doi.org/10.1515/cclm-2020-0815> (2020).
- Balasco, N., D'alessandro, V., Ferrara, P., Smaldone, G. & Vitagliano, L. Analysis of the time evolution of COVID-19 lethality during the first epidemic wave in Italy. *Acta Bio Med. Atenei Parmensis* **92**, 20. <https://doi.org/10.23750/ABM.V92I2.11149> (2021).
- Murray, C. J. Covid-19 will continue but the end of the pandemic is near. *Lancet* **399**, 417–419. [https://doi.org/10.1016/S0140-6736\(22\)00100-3](https://doi.org/10.1016/S0140-6736(22)00100-3) (2022).
- Lazarus, J. V. *et al.* A multinational Delphi consensus to end the COVID-19 public health threat. *Nature* **2022**(15), 1–14. <https://doi.org/10.1038/s41586-022-05398-2> (2022).
- Daria, S. & Islam, M. R. The SARS-COV-2 omicron wave is indicating the end of the pandemic phase but the COVID-19 will continue. *J. Med. Virol.* **94**, 2343. <https://doi.org/10.1002/JMV.27635> (2022).
- World Health Organization. *Strategic Preparedness, Readiness and Response Plan to End the Global COVID-19 Emergency in 2022* (World Health Organization, \*\*, 2022).
- Piri, J. *et al.* Feature selection using artificial gorilla troop optimization for biomedical data: A case analysis with COVID-19 data. *Mathematics* **10**, 2742. <https://doi.org/10.3390/MATH10152742> (2022).
- Sorkhabi, L. B., Gharehchopogh, F. S. & Shahamfar, J. A systematic approach for pre-processing electronic health records for mining: Case study of heart disease. *Int. J. Data Min. Bioinform.* **24**, 97. <https://doi.org/10.1504/IJDMB.2020.10032420> (2020).
- Ayar, M., Isazadeh, A., Gharehchopogh, F. S. & Seyedi, M. H. Chaotic-based divide-and-conquer feature selection method and its application in cardiac arrhythmia classification. *J. Supercomput.* **78**, 5856–5882. <https://doi.org/10.1007/S11227-021-04108-5/FIGURES/7> (2022).
- World Health Organization. *Estimating Mortality from COVID-19* (World Health Organization, 2019).
- Staerk, C., Wistuba, T. & Mayr, A. Estimating effective infection fatality rates during the course of the COVID-19 pandemic in Germany. *BMC Public Health* **21**, 1–9. <https://doi.org/10.1186/S12889-021-11127-7/FIGURES/3> (2021).
- Sánchez-Romero, M., di Lego, V., Prskawetz, A. & Queiroz, B. L. An indirect method to monitor the fraction of people ever infected with COVID-19: An application to the united states. *PLoS One* <https://doi.org/10.1371/JOURNAL.PONE.0245845> (2021).
- Luo, G., Zhang, X., Zheng, H. & He, D. Infection fatality ratio and case fatality ratio of COVID-19. *Int. J. Infect. Dis.* **113**, 43–46. <https://doi.org/10.1016/J.IJID.2021.10.004> (2021).
- Lam, K. F. *et al.* A test for constant fatality rate of an emerging epidemic: With applications to severe acute respiratory syndrome in Hong Kong and Beijing. *Biometrics* **64**, 869–876. <https://doi.org/10.1111/J.1541-0420.2007.00935.X> (2008).
- Ghani, A. C. *et al.* Methods for estimating the case fatality ratio for a novel, emerging infectious disease. *Am. J. Epidemiol.* **162**, 479–486. <https://doi.org/10.1093/AJE/KW1230> (2005).
- Lu, Z. & Chen, Z. Timely case-fatality risk estimation. *Epidemiology* **26**, e26–e27. <https://doi.org/10.1097/EDE.000000000000258> (2015).
- Yip, P. S. *et al.* A comparison study of realtime fatality rates: Severe acute respiratory syndrome in Hong Kong, Singapore, Taiwan, Toronto and Beijing, China. *J. R. Stat. Soc. A. Stat. Soc.* **168**, 233–243. <https://doi.org/10.1111/J.1467-985X.2004.00345.X> (2005).
- Thomas, B. S. & Marks, N. A. Epidemiology and infection estimating the case fatality ratio for COVID-19 using a time-shifted distribution analysis. *Epidemiol. Infect.* **149**, e197. <https://doi.org/10.1017/S0950268821001436> (2021).
- Angelopoulos, A. N., Pathak, R., Varma, R. & Jordan, M. I. On identifying and mitigating bias in the estimation of the COVID-19 case fatality rate. *Harvard Data Sci. Rev.* <https://doi.org/10.1162/99608f92.f01ee285> (2020).
- Lipsitch, M. *et al.* Potential biases in estimating absolute and relative case-fatality risks during outbreaks. *PLoS Negl. Trop. Dis.* **9**(7), e0003846. <https://doi.org/10.1371/journal.pntd.0003846> (2015).
- Baud, D. *et al.* Real estimates of mortality following COVID-19 infection. *Lancet Infect. Dis.* **20**, 773. [https://doi.org/10.1016/S1473-3099\(20\)30195-X](https://doi.org/10.1016/S1473-3099(20)30195-X) (2020).
- Wilson, N., Kvalsvig, A., Barnard, L. T. & Baker, M. G. Case-fatality risk estimates for COVID-19 calculated by using a lag time for fatality. *Emerg. Infect. Dis.* **26**, 1339. <https://doi.org/10.3201/EID2606.200320> (2020).
- Shim, E., Mizumoto, K., Choi, W. & Chowell, G. Estimating the risk of COVID-19 death during the course of the outbreak in Korea, February–May 2020. *J. Clin. Med.* <https://doi.org/10.3390/JCM9061641> (2020).
- Newall, A. T. *et al.* Delay-adjusted age- and sex-specific case fatality rates for COVID-19 in South Korea: Evolution in the estimated risk of mortality throughout the epidemic. *Int. J. Infect. Dis.* **101**, 306. <https://doi.org/10.1016/J.IJID.2020.09.1478> (2020).
- Manisha, M. & Shyamapada, M. Estimating global case fatality rate of coronavirus disease 2019 (COVID-19) pandemic. *Int. J. Clin. Microbiol. Biochem. Technol.* **3**, 036–039. <https://doi.org/10.29328/JOURNAL.IJCMBT.1001014> (2020).
- Rothman, J. E., Eidelberg, D., Rothman, S. L., Holford, T. R. & Rothman, D. L. Analysis of the time course of COVID-19 cases and deaths from countries with extensive testing allows accurate early estimates of the age specific symptomatic cfr values. *PLoS One* **16**(8), e0253843. <https://doi.org/10.1371/journal.pone.0253843> (2021).
- Hastie, T. & Tibshirani, R. Generalized additive models for medical research. *Stat. Methods Med. Res.* **4**, 187–196. <https://doi.org/10.1177/096228029500400302> (1995).
- Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **19**, 716–723. <https://doi.org/10.1109/TAC.1974.1100705> (1974).
- Ministry of Science, Innovation and Universities, Spain. What is the ISCIII and what does it do? These are its functions and objectives (2022). <https://eng.isciii.es/eng.isciii.es/InformacionCiudadanos/DivulgacionCulturaCientifica/DivulgacionISCIII/Paginas/Divulgacion/Que-es-el-ISCIII.html>. Accessed 8 Nov 2022.
- Ministry of Science, Innovation and Universities, Spain. COVID-19, documentación y datos (2022). <https://cncovid.isciii.es/covid19/#documentación-y-datos>. Accessed 8 Nov 2022.
- Martín-Sánchez, V. *et al.* Análisis de las tasas de letalidad de la infección por sars-cov-2 en las comunidades autónomas de España. *Med. Fam. SEMERGEN* **46**, 12–19. <https://doi.org/10.1016/J.SEMERG.2020.06.013> (2020).
- R Core Team. *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing (R Foundation for Statistical Computing, 2022).
- Moraga, P., Ketcheson, D. I., Ombao, H. C. & Duarte, C. M. Assessing the age- and gender-dependence of the severity and case fatality rates of COVID-19 disease in Spain. *Wellcome Open Res.* <https://doi.org/10.12688/WELLCOMEOPENRES.15996.1> (2020).
- Haitao, T. *et al.* Covid-19 and sex differences: Mechanisms and biomarkers. *Mayo Clin. Proc.* **95**, 2189–2203. <https://doi.org/10.1016/J.MAYOCP.2020.07.024> (2020).

35. Bienvenu, L. A., Noonan, J., Wang, X. & Peter, K. Higher mortality of COVID-19 in males: Sex differences in immune response and cardiovascular comorbidities. *Cardiovasc. Res.* **116**, 2197–2206. <https://doi.org/10.1093/cvr/cvaa284> (2020).
36. Leung, J. M. *et al.* Ace-2 expression in the small airway epithelia of smokers and copd patients: Implications for COVID-19. *Eur. Respir. J.* <https://doi.org/10.1183/13993003.00688-2020> (2020).
37. Klein, S. L. & Flanagan, K. L. Sex differences in immune responses. *Nat. Rev. Immunol.* **16**, 626–638. <https://doi.org/10.1038/nri.2016.90> (2016).
38. Al-Lami, R. A., Urban, R. J., Volpi, E., Algburi, A. M. & Baillargeon, J. Sex hormones and novel corona virus infectious disease (COVID-19). *Mayo Clin. Proc.* **95**, 1710–1714. <https://doi.org/10.1016/J.MAYOCP.2020.05.013> (2020).

### Author contributions

P.S. wrote the first version of the manuscript. P.S. and A.S. carried out statistical analysis and prepared figures and tables. C.H., A.P. and J.M.P. participated in the writing of the results and discussion sections. All authors reviewed the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to A.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://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

Scientific Reports is a copyright of Springer, 2022. All Rights Reserved.