Statistical comments
In general, the methodology and the statistical analysis used appears appropriate. However, there are a few issues relating to the serosurvey and the estimates of institutional deaths which should be addressed:

1. It would be useful to include more details of the serosurvey in the text, rather than simply referring to the Lancet paper (Reference 6 - which covers the period April 27 to May 11, whereas this study uses data from an extended period April 27 to June 22). For example, the response rate of the survey should be given as well as the number of non-valid results.

Response: The revised manuscript includes additional details and a new Figure 1 that describes the response rate and the number of non-valid results. The revised Methods section reads: "Of 98,891 individuals who were eligible for the ENE-COVID serosurvey, 10,238 could not be contacted, 14,926 declined to participate, 15 had missing age, and 5,421 did not have valid results from the point-of-care test. Of the remaining 68,291 study participants, 61,098 received the CMIA test in at least one round (61.8% of eligible individuals and 68.9% of contacted individuals), with 43,212 participants receiving the CMIA test in all three rounds, 11,618 in two rounds, and 6,268 in one round (figure 1). Response to the CMIA test was lower in individuals younger than 10 years (22.3% of eligible individuals) and older than 80 years (51.7%), and in men aged 20–59 years compared with women (62.6% versus 70.1%)."

Note that the seroprevalence data were collected between April 27 and June 22 but deaths were included through July 15 to account for both the lag between infection and death and potential delays in reporting.

2. A sensitivity analysis which corrects for the estimated sensitivity and specificity of the antibody detection test is carried out, using point estimates derived from a meta-analysis of diagnostic accuracy studies, and appropriately incorporates the associated variance of the estimates. Although full details are provided in the Supplementary material, it would be helpful to expand a description of this sensitivity/specificity information in the text to include the range of sensitivity/specificity values obtained from these studies.

Response: Done. The text now reads: "In sensitivity analyses, we corrected the SARS-CoV-2 seroprevalence estimates for the CMIA’s sensitivity and specificity, which were estimated to be 90.6% (95% confidence interval 88.1% to 92.6%) and 99.3% (99.0% to 99.5%), respectively, from a meta-analysis of 23 diagnostic accuracy studies. In these studies, sensitivity ranged from 75% to 100% in 1,494 samples from confirmed COVID-19 cases of different severity, and specificity ranged from 97.5% to 100% in 7,696 samples obtained before the pandemic, with moderate between-study heterogeneity in sensitivity (I2=46%) and no heterogeneity in specificity (I2=0%) (see supplementary methods and supplementary figure 1 for details)."

3. An estimate of the number of institutional deaths (which are then subtracted from all deaths) is derived from a variety of Regional sources (Supplementary Table 1). How accurate are these figures? It is slightly concerning that the confirmed to suspected COVID-19 death ratios are very different between regions (eg. Andalucia records 558
confirmed and 0 suspected COVID-19 deaths in nursing homes, whereas Madrid records 1,253 confirmed and 4,734 suspected COVID-19 deaths in nursing homes). Is it possible that for some regions the number of confirmed plus suspected COVID-19 death reported is a gross underestimate? How much do the IFR estimates vary between regions?

Response: In the absence of a national surveillance system for deaths in nursing homes, we painstakingly collected region-specific information to create the summary shown in Supplementary Table 2. To the best of our knowledge, this is the most accurate summary that is currently available for deaths in nursing homes in Spain. While there was some between-region heterogeneity in the reporting of these deaths, it is unlikely that this variability resulted in an underestimate of the IFRs. On the other hand, it is possible that underreporting in some regions may have resulted in a slight overestimation of IFRs when using excess deaths. Note that the between-region differences in confirmed to suspected COVID-19 death ratios are largely explained by the intensity of the pandemic wave. Whereas in areas with low seroprevalence like Andalucía (8.4 million people with seroprevalence <3%) most COVID-19 cases in nursing homes were confirmed, in areas with high seroprevalence like Madrid (6.6 million people with seroprevalence >11%) a larger proportion of them were labeled as “suspected” due to lack of PCRs. Though the IFR estimates cannot be meaningfully compared across Spanish administrative regions because of the large statistical variability, the revised paper includes a Supplementary Figure 2 with age- and sex-standardized IFR ratios for each NUT1 region (first-level territorial units developed by the European Union) using the whole country as reference.

Editor’s comments
Please report the response rate for the serosurvey. That is what proportion of households approached consented to take part and provided samples?

Response: Done. Please see response to the first comment above.

Please report the results of the serosurvey in a little more detail. What were the proportions of positive, negative and indeterminate results for each age and were there any missing data?

Response: Done. This information is shown in the new Supplementary Table 1.

Please explain and discuss the large differences in the ratio of confirmed vs suspected COVID deaths between regions reported in Supplementary Table 1.

Response: Please see our response to Statistical Comments above. Also, the revised Discussion reads: “A potential limitation of the study is that neither of the two mortality surveillance systems differentiated between deaths in institutionalized and non-institutionalized individuals. Hence, we had to calculate deaths with confirmed and suspected COVID-19 in nursing homes as reported from different Spanish Regional Authorities, disaggregate these deaths by sex and age group (as described in supplementary methods), and subtract them from the total number of deaths. Because of insufficient diagnostic capacity in regions heavily hit by the first wave of the pandemic, COVID-19 cases could not be confirmed in many instances while deaths among suspected
COVID-19 cases living in nursing homes might be underreported. If this were the case, the IFR based on excess deaths might be somewhat overestimated.

COMMENTS FROM REVIEWERS

Reviewer 1

The authors provide a comprehensive and concise overview of the infection fatality risk of SARS-CoV-2 in their nation's population. The article is well constructed and experimental methodologies well conducted, analyzed, and explained. There are a few grammatical errors that can be improved prior to publication.

Response: Thank you.

Reviewer 2

Many thanks to the authors for this very interesting piece of work covering a large study. The language used in the discussion is clear and approachable, making it accessible to patients and carers with an interest in the subject matter. The paper is relatively clear in presenting the results of the research findings, the limitations of the review and suggestions for further research. However, it does pose questions around the IFR rates for different population groups other than age and gender which are not part of the analysis. Conclusions could be more rounded on the continued need for control measures.

Response: We have revised the text concerning the need for control measures. The revised conclusions read: "The high IFR in the older age groups supports existing measures (e.g., social distancing, face masks, and educational campaigns) to shield these groups from infection. However, relying exclusively on attempting to shield the elderly might be a high-risk strategy for pandemic management. Given the high transmissibility of the disease and the high proportion of susceptible population, even a substantial reduction of transmission in the elderly may result in a large number of deaths."

The topic is very relevant and pertinent to patients and carers. Anxiety around being infected with COVID-19, especially for patients with underlying health conditions is high as local outbreaks continue and the infection rate increases. This study, which takes place exclusively in non-institutional settings provides a vital aspect on the academic response to COVID-19. It will offer important messages for older groups who may also have underlying conditions who are out and about in the community or those caring for them and could help their decision making about their choices as a second wave looks likely. It could also be used in the public debate around how we can best live with COVID-19 and maintain economic viability.

Response: We agree and hope that these findings continue to be used in the public debate about pandemic management.
The section on P3 ‘What is already known about this topic’ is useful in setting the scene for the study. The paper clearly describes its additionality to the body of research on the risk of COVID-19 in ‘What this study adds’. The Introduction on P4 is clear and offers the context and rationale for the study and the intention of the study to use excess mortality as a research tool to complement the IFR and the reason for doing this.

The data estimating excess deaths is of particular relevance to patients and carers. The indirect deaths and ill health have been alluded to since the pandemic began. P8 offers some explanation of the morbidity due to the rationing of services we have seen during the Pandemic, which patients and carers will be able to relate to and which is very useful.

The need the study identified for further and specific work in care homes was welcome as comparing IFR in the community and in institutional settings would be very helpful to patients and carers.

The Table on P18 was very helpful in understanding the narrative and comparing the IFR for the different age/gender groups. Whilst it is widely reported that older groups and males are more susceptible to mortality with COVID-19, i.e. this isn’t in itself new, the data it provided on specific age groups and the rate of overall risk compared with seasonal flu was very useful.

Response: We agree with the reviewer.

The research topic took a binary approach and focused on infection rates and death rates, by age and gender, in order to obtain the Infection Fatality Risk (IFR) For me, there were unanswered questions about the study participants. It would have been interesting to see some disaggregation on:
• Any differences between urban and rural areas of IFR
• Occupational breakdown of IFR
The paper also made me want to know more about:
• Behavioural differences between age groups and genders helping to drive these different IFR levels? E.g. the level of adherence to government restrictions or their ability to adhere to restrictions?
• Any socio-economic drivers influencing the IFR rate?
• The role of PPE in the community to manage the IFR among vulnerable groups
• The impact of asymptomatic COVID-19 in the community and the IFR.
• Non-mortality aspects of COVID, such as ‘long COVID’

Response: These are certainly very interesting and relevant questions, but they cannot be answered with the information available. Even though the ENE-COVID study included a questionnaire with some information regarding occupation, behaviour, and several risk factors, the two surveillance systems providing information on deaths, namely RENAVE and MoMo, do not provide data on these factors. In fact, this is the case in all studies estimating IFR in different countries.

The paper highlights the dilemmas faced by countries that have ageing populations
The methodology was explained at length, however, I still found it a little difficult to understand how data sets were utilized to reach the numbers cited. This was mitigated to some extent when the paper stated that findings were in-keeping with other studies.

Response: We expect that the additional detail included in the revised text (see responses to the Editorial Committee above) will facilitate the understanding of the methodology.

There is a growing feeling among some in our communities to eschew the COVID-19 precautions, after months of social and economic disruption, which could ultimately affect the acuity of any second wave. This paper reminds us of the magnitude of risk that older/male groups face, which could support patients and carers in their conversations with their doctors about their treatment plans and their risk appetite from patients and from health providers around this - so support realistic medicine conversations. Similarly, it may offer grist to the mill for sceptics and the majority of adults who’s IFR is relatively low. Therefore, possibly the outcomes could have attended more to the complexity of the virus and how it uses human behaviour to reach more vulnerable groups.

Response: That is certainly our hope.

I had to read the paper a number of times before I felt I understood the study and its outcomes. Perhaps more plain English in the Results section on P2 would be helpful, with the data presented in a small table. Although I understand the statistical notation is the norm for research papers, lay people may not understand the significance – or otherwise of the percentages, hence my suggestion for a textual summary followed by a table with the numbers cited in this section.

The paper introduces the study by reminding us about the debate over the need and usefulness of lockdowns and social restrictions on P4. In the conclusion on Page 10, the authors consider that the evidence does “provide support for strong control measures”. I think it would have been useful here to make more of this at the conclusion – as it provides resolution to the research hypothesis. COVID-19 is afterall, ten times the risk of seasonal flu (a fact that could usefully be made clear in the Results section on P2) – and enable patients and carers to see the evidenced for continued caution.

Response: In addition to the changes referred to above, we have revised the conclusion regarding control measures for improved clarity (see above).

There was no participant involvement in this study as it was using existing datasets and not embarking on any primary research. The write up of the study is relatively clear and approachable, but also technical in places, particularly in the results section for those without a statistical background to understand, which possibly some patient/carer involvement may have influenced to better explain the statistical notations in accessible terms and conveyed the results in ways which were more meaningful for non-researchers.

Many thanks to the authors for this study.

Response: Thank you.
Reviewer 3

This is an important and interesting paper utilising large-scale serological data from Spain as the basis for calculating the infection fatality risks. I have only specific points.

In the methods section on page 6 it says "Epidemiologic questionnaires and serology tests were administered to 68,292 individuals who participated in at least one round" this suggests some people participated in more than one round – could that be made clear.

Response: The paper now clarifies this: “Of the remaining 68,291 study participants, 61,098 received the CMIA test in at least one round (61.8% of eligible individuals and 68.9% of contacted individuals), with 43,212 participants receiving the CMIA test in all three rounds, 11,618 in two rounds, and 6,268 in one round (figure 1).” The flowchart with this information is included as the new Figure 1.

Two categories of deaths – those with laboratory-confirmed COVID-19 and the estimated excess all-cause deaths were used. Detailed methods for calculating the latter are not provided, yet it is a key statistic. Reference to a study in Italy and reference to the EuroMOMO data for Spain are given, but not adequate information on estimation methods. Whilst it is reasonable to assume the two categories these are the lower and upper estimates, examining the historical series shows that several non-pandemic years have increases above the historical series for the ten prior years, so the excess could contain what would have been a non-pandemic “excess” component, and deaths with laboratory – confirmed COVID could have been deaths with, rather than from, the condition. Therefore there is some additional uncertainty (in both directions) that should be acknowledged. The fact that the excess mortality estimates are different to those given for Spain recently https://www.nature.com/articles/s41591-020-1112-0 (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 (2020) when looking at the same period within the figures presented demonstrates this uncertainty.

Response: The revised paper describes the MoMo methodology in more detail: “MoMo data are used to quantify excess deaths for a particular period, as the difference between the observed daily deaths corrected for reporting delay and those expected based on historical seasonal variation (centered 7-day moving averages) and a non-linear secular trend (annual median daily deaths ) from the last 10 years […] MoMo estimates are of similar magnitude to those reported using a different method (Kontis et al.).” Also, we agree that it is conceivable that the excess deaths could contain an unidentified non-pandemic “excess” component (e.g., a more severe than expected flu season) even if we cannot find evidence of that being the case (e.g., at the beginning of March, when the excess mortality started, the seasonal influenza was decreasing and the rates of hospitalization due to influenza were low). Regardless, this hypothetical non-pandemic excess would not affect the fact that our IFR estimates based on excess deaths should be viewed as an upper bound of the IFR.

There is clear evidence that some seropositive individuals become seronegative, and not just due to measurement error. This should be discussed.
Response: The reviewer is right. The revised Discussion section acknowledges this issue and explains why it is unlikely to affect our estimates: “Thus, the study was designed to detect all individuals with antibodies during the first epidemic wave because most participants would have been infected one month before their first participation and IgG antibodies are detected 2–3 weeks after symptom onset in more than 90% of infections. Though IgG antibodies in some infected participants could have decreased over time, particularly in mild infections, a recent study shows that this phenomenon occurs at least 3 months after infection, so a substantial underestimation of infected individuals is unlikely.”

The IFR reported relates to the age distribution of the population who became infected. It would be useful to see this adjusted to the age distribution of the total population, since this value is the one that was used in many of the initial projections which influenced the policy of countries with regard to the pandemic.

Response: The adjustment suggested by the reviewer does not materially change the estimates because seroprevalence varied little by age and sex: The standardized IFR (95% confidence interval) to the sex and age distribution of the entire non-institutionalized Spanish population was 0.86% (0.75% to 0.98%) for confirmed COVID-19 deaths and 1.11% (0.96% to 1.28%) for excess deaths.

It is misleading to refer to COVID as having an IFR ten times larger than that for pandemic influenza. The authors cite the 2009 pandemic, known to be associated with either slightly lower or a little higher than usual seasonal flu mortality. They cannot generalise from this, and it would be very valuable if they would add a comparison with the IFR in previous pandemics, 1957-9, 1968-9, for which good data are available for Spain, and in particular to present these by age group.

Response: We agree. To avoid distractions, we have removed the reference to pandemic influenza. The debate about the relative IFR of SARS-CoV-2 and influenza revolves around seasonal (not pandemic) influenza.