nature research

Peer Review File

Manuscript Title: Age-specific mortality and immunity patterns of SARS-CoV-2

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referees' comments:

Referee #1 (Remarks to the Author):

The authors dig into the relationship between age and COVID-19 mortality using data from 45 countries. They make the case that nursing home deaths can be decoupled from deaths in the general population, providing a simple explanation for the discrepancies in the data that holds true across countries and continents. They also attempt to use seroprevalence data to infer the infected proportion and the IFR in different countries.

I found the first part of this analysis to be very convincing. If anything I'd like to see some more discussion around the operational utility of these results, i.e. what do we do with the estimated log-linear relationship calculated in the absence of nursing home deaths, given our experience that it has been very difficult to keep SARS-CoV-2 out of nursing homes entirely.

I have more concerns about the second part of the analysis, estimating the infected proportion and the IFR. In particular, I am concerned that test sensitivity and specificity were only taken into account in the minority of studies where this legwork had already been done (Table S1). Observed prevalences are very low in some of these studies, making them very sensitive to assumed test specificity. A related issue is that credible intervals are unrealistically tight in many cases. Ideally not only sensitivity and specificity should be taken into account, but also uncertainty in these parameters. This would lead to more realistic propagation of uncertainty into infected proportions and IFR estimates. Without taking these factors into account the values presented here are in danger of being both biased and overly precise. An excellent statistical bar to aim for was presented here (http://www.stat.columbia.edu/~gelman/research/published/specificity.pdf).

Aside from this main issue I have a few smaller comments:

Line 101: It is stated that “The reporting of COVID-19 deaths for older individuals can also be subject to inconsistencies across settings due to variable prevalence of comorbidities with which a COVID-19-associated death could be mistakenly attributed”. This seems speculative, do you have evidence to back up this claim?

Line 157: The authors assume equal attack rates in <65 years, and 70% of this value for >=65. Why not estimate attack rates directly from the seroprevalence data?

Line 189: The authors claim that studies conducted with blood bank sera give similar results to studies in the general population. This appears to be true for the studies presented in Figure S5, but it is not true in general. In particular, I’m thinking of the UK seroprevalence estimates from NHS Blood and Transplant (NSH BT) donor samples vs. those in the REACT2 study. Perhaps this needs further qualification.

Line 204: It is stated that the estimated infected proportion indicates that the majority of countries are likely a long way from standard herd immunity thresholds. I’m not sure the idea of "standard" thresholds is accurate, as there is still wide debate and uncertainty around appropriate thresholds, i.e. there is not yet scientific consensus.

Lines 231-234: These lines suggest to me that in fact there is no information in the data as to the correct relative attack rate vs. frailty parameter, as the two are confounded. Presenting the example of a relative attack rate that is twice as small for a frailty parameter that is twice as large is not, in my opinion, a useful sensitivity analysis. If this relationship is uninformed then it should simply be stated so.

Line 464: It is stated that the assumed delay from onset to death comes from Verity et al. (2020). This reference is now out of date, being as it was a rapid assessment based on a very small sample size. Given the number of deaths that have now occurred the authors should find a more recent reference or a larger dataset on which to derive their own updated estimate.

Figure 2D: Am I right that credible intervals on this plot are too small to see? (If so see uncertainty comment above).

Figure 3A: Perhaps this information would be better in a table?

Figure 3B: I'm concerned that in some places the model fit is a long way from the data (England, France, Slovenia). Perhaps this relates to too-tight credible intervals? (If so see uncertainty comment above).

Referee #2 (Remarks to the Author):

I have a number of concerns that would need to be addressed:
(1) Sero-prevalence and immunity. An important component of the work is based on estimation of sero-prevalence, i.e., antibody based. However, we have known for a while that there are important issues surrounding sero-prevalence: (a) it takes a while before antibodies become detectable; (b) the sensitivity is considerably less than 100%; (c) waning of antibodies. So, uncertainty surrounding the nature and extent of immunity is considerable, because humoral immunity seems to wane over time and the role of T-cell immunity is yet to be studied in more detail. This issue is profound and discussion of it is lacking. Of course, immunity is complex and a comprehensive discussion is outside of the paper's scope, but it should be addressed in as far as it may impact the paper's goals: estimating IFR and its translation to total number of infection. Relevant papers to this effect:

a. Braun, Loyal, Frensch, et al. (2020). SARS-CoV-2-reactive T cells in healthy donors and patients with COVID-19. Nature, Accelerate Article Preview.
b. Danchin and Turinici (2020). Immunity after COVID-19 protection or sensitization? medRxiv.
c. Huang, Garcia-Carreras, Hitchings, et al. (2020). A systematic review of antibody mediated immunity to coronaviruses: antibody kinetics, correlates of protection, and association of antibody responses with severity of diseases. medRxiv.
d. Grzelak, Temmam, Planchais et al. (2020). SARS-CoV-2 serological analysis of COVID-19 hospitalized patients, pauci-symptomatic individuals and blood donors. medRxiv.
In this sense, it might be interesting to consider alternative routes for estimating the total number of cases in a country, e.g., based on mathematical modeling, where confirmed cases, hospitalizations, ICU occupation,… are used as input:
a. Faes, Abrams, Van Beckhoven, et al. (2020). Time between symptom onset, hospitalization and recovery or death: a statistical analysis of different time-delay distributions in Belgian COVID-19 patients. medRxiv.
b. Abrams, Wambua, Santermans, et al. (2020). Modeling the early phase of the Belgian COVID-19 epidemic using a stochastic compartmental model and studying its implied future trajectories. medRxiv.
c. Willem, Abrams, Petrof, et al. (2020). The impact of contact tracing and household bubbles on deconfinement strategies for COVID-19: an individual-based modelling study. medRxiv.

(2) Underreporting of COVID-19 mortality relative to excess mortality. It is known that the extent of reporting of COVID-19 mortality varies across countries. While Belgium may report roughly 100% of its excess mortality as COVID-19 related mortality, in other countries this fraction shrinks to much lower levels. This might invalidate the results found if not properly accommodated. This is one of the issues discussed in:
a. Aron, J., Giattino, C., Muellbauer, J., and Ritchie, H. (2020). A pandemic primer on excess mortality statistics and their comparability across countries. https://ourworldindata.org/covid-excess-mortality.
This really is a fundamental point, as examining IFR from COVID-19 confirmed deaths versus total deaths (confirmed and suspected; in hospitals as well as in NH and other settings) makes a big difference, would lead to very different rankings of countries, etc.

(3) Related work. There is some related work regarding global IFR. For example, the paper by Grewelle and De Leo (2020) is relevant in this regard, and arguably should be cited.
a. Grewelle RE, De Leo GA (2020). Estimating the global infection fatality rate of COVID-19. medRxiv.

(4) Reference age category. I can see the rationale for choosing a well-argued reference age category (e.g., 60-65). Given the relatively low rates at younger ages, and the complexity at older ages because of the convolution of NH and non-NH populations. But still, it is a narrow age band, and a sensitivity analysis should be done. Also, it should be clear that associated statistical uncertainty is properly taken into account. Generally, there are a number of estimates (in text, in tables), where either a point estimate only is given (e.g., Table S2) or where the method used to estimate precision is not explained. So, throughout the methodological appendix, precision estimation should be clearly discussed, and backed up with the appropriate formulas or numerical algorithm.

(5) Nursing home population. The authors are well aware of the specificity of the NH population. Other authors have indicated very strong differences between NH and non-NH IFR's and related quantities:

a. Molenberghs, G., Faes, C., Aerts, J., et al (2020). Belgian COVID-19 mortality, excess deaths, number of deaths per million, and infection fatality rates (8 March -- 9 May 2020). medRxiv.

The authors also refer to different frailty levels in the NH and non-NH populations, but I am lacking a clear, logical buildup. Ideally, the NH and non-NH populations should be separated completely, so as to create age-sex-NH strata, rather than merely age-sex strata. The sensitivity analysis done in lines 222-243 is a bit rudimentary; it is not plausible to assume equal frailty, except as a 'most extreme' scenario. A detailed discussion on frailty is given in:

a. Declercq, A., de Stampa, M., Geffen, L., et al. (2020). Why, in almost all countries, was residential care for older people so badly affected by COVID-19? OSE Working Paper Series, Opinion Paper No. 23. Brussels: European Social Observatory.

Detail: when used the first time in text 'CrI' should be defined.

Author Rebuttals to Initial Comments:

Note: Referee comments in blue

Referee #1 (Remarks to the Author):

Comment 1.1. The authors dig into the relationship between age and COVID-19 mortality using data from 45 countries. They make the case that nursing home deaths can be decoupled from

deaths in the general population, providing a simple explanation for the discrepancies in the data that holds true across countries and continents. They also attempt to use seroprevalence data to infer the infected proportion and the IFR in different countries.

I found the first part of this analysis to be very convincing. If anything I'd like to see some more discussion around the operational utility of these results, i.e. what do we do with the estimated log-linear relationship calculated in the absence of nursing home deaths, given our experience that it has been very difficult to keep SARS-CoV-2 out of nursing homes entirely.

We thank the reviewer for their positive feedback. It has indeed proven very difficult to keep SARS-CoV-2 out of nursing homes in many higher-income settings, whilst in many lower-income settings this is less likely to be an issue due to different population demographics. It is very clear that these hyper-vulnerable populations require the highest levels of shielding in order to reduce the mortality rates of SARS-CoV-2. While it is not per se of strict operational utility, the clear log-linear relationship between age and IFR can be of use in aiding our understanding of the proportion of the population infected, transmission patterns across settings, including at a subnational level. When looking at the relative risk of reported COVID-19 deaths by age, major deviations from a log-linear relationship may be indicative of extensive transmission amongst vulnerable populations, or where shielding of elderly populations has been successful or where under-reporting of deaths in certain age-groups may be happening.

Specific changes to document:

- *We now include further discussion to this effect at line 272: "By providing a benchmark of the expected number of deaths by age in older individuals, our approach allows us to identify countries where excess transmission in nursing home populations is likely to have occurred, far exceeding that of the general population. Our findings demonstrate how infections in nursing homes can drive population-level IFRs, through both increased attack rates and increased vulnerability; further highlighting the need to provide the highest level of protection from infection for these populations. In addition, we have been able to identify locations where deaths in the elderly population are likely to be under-reported."*

Comment 1.2. I have more concerns about the second part of the analysis, estimating the infected proportion and the IFR. In particular, I am concerned that test sensitivity and specificity were only taken into account in the minority of studies where this legwork had already been done (Table S1). Observed prevalences are very low in some of these studies, making them very sensitive to assumed test specificity. A related issue is that credible intervals are unrealistically tight in many cases. Ideally not only sensitivity and specificity should be taken into account, but also uncertainty in these parameters. This would lead to more realistic propagation of uncertainty into infected proportions and IFR estimates. Without taking these factors into account the values presented here are in danger of being both biased and overly precise. An excellent statistical bar to aim for was presented here (http://www.stat.columbia.edu/~gelman/research/published/specificity.pdf).

We thank the reviewer for raising this important point. We agree that our analysis would benefit from a more careful assessment of assay performance in each of the individual seroprevalence

studies. We have now revisited each of the seroprevalence studies included in our analysis and identified the assay sensitivity and specificity for the majority of studies (Revised Table S1). Where estimates of seroprevalence adjusted for assay performance were not provided by authors of the individual serostudies, we have applied this correction ourselves and now use these adjusted values in our main analyses (available for 24/25 studies). For completeness, we show a comparison of model estimates when considering seroprevalence values that are unadjusted and adjusted for assay performance and find only minor differences in the results (Figure S5, below). Further, in response to additional comments raised by reviewer 2 we have now conducted sensitivity analyses around the mean delay from infection-to-seroconversion and the potential for decay of seropositivity over time (Figure S4, below).

We understand the reviewer's concerns about unrealistic model uncertainty and read the suggested Gelman & Carpenter analysis with much interest. We agree that this presents a very nice statistical standard to which seroprevalence estimates should be derived. The approach presented by Gelman & Carpenter requires information on the experiments used to determine assay sensitivity and specificity, which are, unfortunately, not commonly reported in seroprevalence studies. It is therefore not possible for us to accurately propagate uncertainty in assay sensitivity and specificity to our model estimates without knowledge of these experiments from each of the 25 seroprevalence studies. However, to address the general issue of unrealistic uncertainty in our ensemble model, we now model the expected seroprevalence using a Beta distribution. Previously, the confidence of our ensemble model arose from the large number of seroprevalence samples across the combined studies, which the Binomial distribution accounts for. Using a Beta distribution allows us to model seroprevalence with unknown prior expectations of the probabilities of success or failure through the estimation of a variance parameter, providing a more accurate reflection of uncertainty in our estimates. In addition, we now present the range of model estimates derived from individual seroprevalence studies alongside the ensemble model, to specifically highlight the uncertainty as suggested by individual studies.

Specific changes to document:

- *We have ensured the adjustment of seroprevalence values for assay performance in 24/25 studies included in our analysis (Table S1) and now base our main model estimates on these adjusted values (updated throughout the manuscript).*
- *We now model expected seroprevalence assuming a Beta distribution, allowing us to more accurately capture uncertainty in the data through the estimation of a variance parameter.*
- *Throughout the text and figures we present the range of estimates as suggested by individual studies alongside the ensemble model estimates, so as to be interpreted as an external measure of uncertainty.*

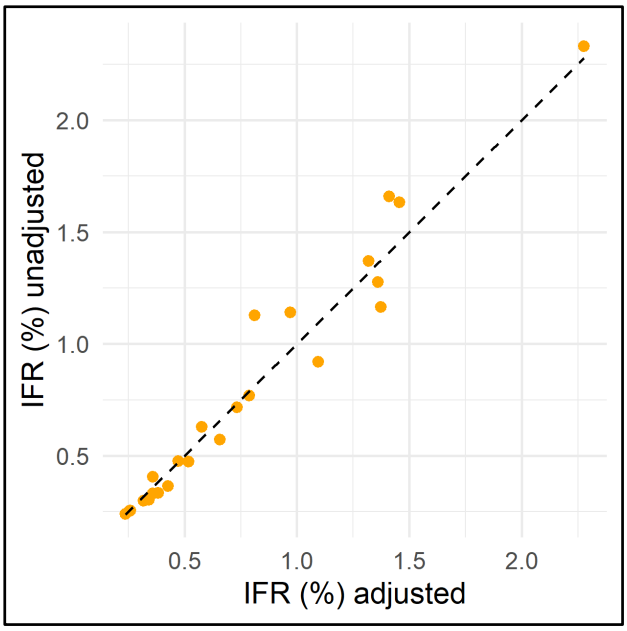


Figure S5. A comparison of population IFR estimates, using seroprevalence values that are adjusted and unadjusted for assay sensitivity and specificity. Orange points represent median model estimates of the population IFR in France derived from separately fitting individual seroprevalence studies in the model likelihood. Black dashed line represents the ideal scenario of x=y.

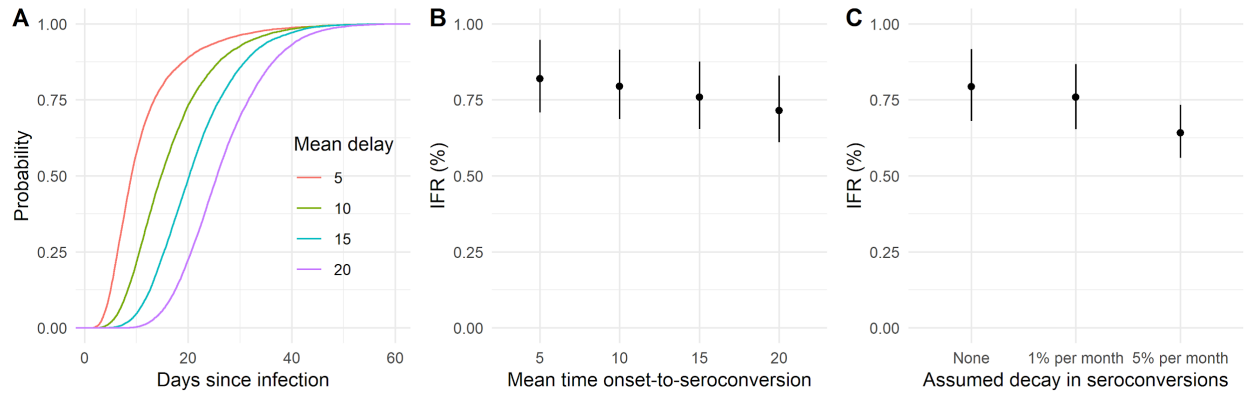


Figure S4. (A) Cumulative probability density functions for delay between infection-to-seroconversion, obtained by assuming mean times from onset-to-seroconversion of 5, 10, 15 and 20 days. (B) Model estimates of the population IFR for France under different assumptions regarding the mean time from onset-to-seroconversion. (C) Model estimates of the population IFR for France under different assumed exponential decays in seroconversions over time (no decay, 1% decay per month and 5% decay per month).

Comment 1.3. Line 101: It is stated that "The reporting of COVID-19 deaths for older individuals can also be subject to inconsistencies across settings due to variable prevalence of comorbidities with which a COVID-19-associated death could be mistakenly attributed". This seems speculative, do you have evidence to back up this claim?

In this sentence we were attempting to highlight that a greater prevalence of comorbidities in elderly individuals can result in difficulties in defining a COVID-19 death and that this definition is likely to vary across settings. We now realise that this is not clear from what we have written and thank the reviewer for highlighting this.

Specific changes to document:

- *We have removed this sentence.*

Comment 1.4. Line 157: The authors assume equal attack rates in <65 years, and 70% of this value for >=65. Why not estimate attack rates directly from the seroprevalence data?

We agree with the reviewer that estimating age-specific attack rates directly from seroprevalence data would be ideal. However, age-specific results are not available for many studies, making inference difficult. We have now made note of this limitation in our discussion.

Specific changes to document:

- *We have noted this limitation in our discussion at line 243: 'Broad availability of age-specific seroprevalence data could provide a future alternative means of directly measuring relative attack rates across ages.'*

Comment 1.5. Line 189: The authors claim that studies conducted with blood bank sera give similar results to studies in the general population. This appears to be true for the studies presented in Figure S5, but it is not true in general. In particular, I'm thinking of the UK seroprevalence estimates from NHS Blood and Transplant (NSH BT) donor samples vs. those in the REACT2 study. Perhaps this needs further qualification.

We agree with the reviewer's point and have now corrected this sentence to clarify that this observation is based only on a limited number of comparable studies and that further comparisons are required to understand the representativeness of studies conducted with blood bank sera.

Specific changes to document:

- *We have clarified at line 154: 'Of the studies included in our analysis, we find that those conducted with blood bank sera (which do not include children and require individuals to be asymptomatic at the time of sample collection) do not give significantly different results to those conducted amongst the general population (Figure S6). However, further head-to-head comparisons within the same populations are needed to fully understand the representativeness of different serological study designs.'.*

Comment 1.6. Line 204: It is stated that the estimated infected proportion indicates that the majority of countries are likely a long way from standard herd immunity thresholds. I'm not sure the idea of "standard" thresholds is accurate, as there is still wide debate and uncertainty around appropriate thresholds, i.e. there is not yet scientific consensus.

We agree with the reviewer that there is a lack of scientific consensus regarding SARS-CoV-2 herd immunity thresholds. Our intention with the use of the term 'standard' was to make the point that these values were far from the 60-70% thresholds specifically, given that lower values are still being debated. To avoid any assumptions about herd-immunity thresholds we have removed this sentence. Further, we have now updated our estimates with up-to-date mortality data and find that very high transmission is likely to have now occurred in many South American countries, consistent with recent subnational seroprevalence reports. The implications that these estimates have on herd-immunity thresholds is not yet clear as it is important to distinguish between herd-immunity thresholds and final epidemic size. However, trends in fatalities in these settings in the coming months are likely to shed some light on the issue.

Specific changes to document:

- *We removed the sentence regarding standard herd immunity thresholds and now focus our discussion on the variable attack rates and their uncertainties at line 175: "These results indicate large heterogeneity in the level of transmission across countries, with particularly high attack rates estimated in many South American countries. Given the underlying heterogeneity in IFR that could not be captured by the ensemble model, it is important to consider the full range of uncertainty in these estimates as suggested by individual seroprevalence studies (grey points in Figure 3B)."*

Comment 1.7. Lines 231-234: These lines suggest to me that in fact there is no information in the data as to the correct relative attack rate vs. frailty parameter, as the two are confounded. Presenting the example of a relative attack rate that is twice as small for a frailty parameter that is twice as large is not, in my opinion, a useful sensitivity analysis. If this relationship is uninformed then it should simply be stated so.

We agree that the relative attack rate and relative frailty parameters are confounded and apologize that it was not clear that their relationship was uninformed. Our intention with this simplistic analysis was to highlight the uncertainties in these parameters and their potential effects on IFR estimates. Reviewer 2 has also brought to our attention literature on the relative frailty of nursing home populations which has allowed us to make data-driven estimates of the IFR in nursing homes, the relative attack rates in nursing homes versus general population in France and the resulting overall population-level IFR. We also still include a sensitivity analysis where the relative frailty between the two communities is varied.

Specific changes to document:

- *We have updated the analysis to reflect recent estimates of a 3.8 times relative frailty of nursing home residents. At line 205 the text now reads "Using France as a reference population, we use the age and sex distribution of nursing home residents to derive a population-weighted IFR of 22.25% (95%CrI: 19.06-25.74%) among French nursing home residents, assuming individuals in nursing homes are 3.8 times more frail than individuals in the general population of the same age and sex, as estimated in Belgian nursing home residents (Figure 4B). Using this estimate of the IFR would suggest that 7.28% of the nursing home population had been infected by the 1st of September 2020 (95%CrI: 6.29-*

8.49%), a 1.70 fold higher infection attack rate than the general population (Supplementary Methods S3).”

Comment 1.8. Line 464: It is stated that the assumed delay from onset to death comes from Verity et al. (2020). This reference is now out of date, being as it was a rapid assessment based on a very small sample size. Given the number of deaths that have now occurred the authors should find a more recent reference or a larger dataset on which to derive their own updated estimate.

Thank you for highlighting this. We now base our assumptions of the delay between onset and death on a more recent analysis with a greater sample size (Wu et al. 2020). Wu et al., report a gamma distributed delay with mean 20 days and standard deviation of 10 days. This estimate is similar to that of Verity et al., causing minimal changes to our model estimates.

Specific changes to document:

- *We now base our assumptions of the delay between onset and death on analysis by Wu et al., which is based on a larger sample size than that of Verity et al.,.*

Comment 1.9. Figure 2D: Am I right that credible intervals on this plot are too small to see? (If so see uncertainty comment above).

The reviewer is correct that the credible intervals are too small to see on this plot where the y-axis has a large range and we agree with the reviewer’s comment that the uncertainty estimated by our ensemble model is unrealistically small. As mentioned above, the narrow credible intervals of our ensemble model are the result of the combined sample size of the 15 national-level seroprevalence studies previously included in our ensemble model (N=147,725). In our updated analysis we now include 22 national-level seroprevalence studies with a combined sample size of N=356,316, which, when using a Binomial distribution would exacerbate the issue of narrow confidence intervals in our ensemble model. We now model seroprevalence with a Beta distribution, allowing us to account for unknown prior expectations of the probabilities of success or failure through the estimation of a variance parameter, providing a more accurate reflection of uncertainty in our estimates. Using a Beta distribution in the ensemble model has the additional benefit that the contribution of different studies to the model likelihood will not be determined by their respective sample sizes, as is the case with the Binomial model. We believe that this provides an additional improvement to the robustness of our model estimates as the sample size of studies is unlikely to be the sole driver of accuracy in seroprevalence estimates. In addition, we present the range of model estimates derived from individual seroprevalence studies alongside the ensemble model, to specifically highlight the uncertainty as suggested by each individual study.

Specific changes to document:

- *We now use a Beta distribution in our ensemble model which provides a more realistic reflection of uncertainty in our model estimates.*

- *Throughout the text and figures we present the range of estimates as suggested by individual studies alongside the ensemble model estimates, so as to be interpreted as an additional measure of uncertainty external to our ensemble estimates.*

Comment 1.10. Figure 3A: Perhaps this information would be better in a table?

We agree that these estimates would be useful in a table and have added this to the Supplementary Information (Table S4). However, we also believe this figure allows a simple visual comparison between countries and continents, which would not be possible in a tabular form and therefore prefer to keep the figure in the main document.

Specific changes to document:

- *We have added a supplementary table of model estimates of the proportion infected with SARS-CoV-2 by country as of the 01/09/2020 (Table S4).*

Comment 1.11. Figure 3B: I'm concerned that in some places the model fit is a long way from the data (England, France, Slovenia). Perhaps this relates to too-tight credible intervals? (If so see uncertainty comment above).

Indeed, the ensemble model is a long way from the data in some places. This is synonymous with our finding that different seroprevalence studies suggest somewhat variable values of IFR, that cannot be explained by population demographics (Figure 2C). This means that our ensemble model is unable to reconcile the observed data in settings where the combined seroprevalence and death data suggest 'outlier' IFR values and it is for this reason that we also separately fit our model to the individual seroprevalence studies so as to reflect the full uncertainty in setting-specific IFR values. We have attempted to make these points clearer in the text and now show the fit of models to individual studies in Figure 3B.

Specific changes to document:

- *We have added into the text at line 166: 'Our ensemble model reproduces the reported seroprevalence values for the majority of studies including the dynamics of reported seroprevalence over time. However, consistent with a substantial underlying heterogeneity in IFR across countries, the ensemble model cannot fully reconcile the relationship between reported seroprevalence and age-specific death data in some locations (Figure 3B).'*
- *We also include fits to the individual serostudy models in Figure 3B to better highlight where the range of values from individual serostudies are derived from.*

Referee #2 (Remarks to the Author):

I have a number of concerns that would need to be addressed:

Comment 2.1. Sero-prevalence and immunity. An important component of the work is based on estimation of sero-prevalence, i.e., antibody based. However, we have known for a while that there are important issues surrounding sero-prevalence: (a) it takes a while before antibodies become detectable; (b) the sensitivity is considerably less than 100%; (c) waning of antibodies. So, uncertainty surrounding the nature and extent of immunity is considerable, because humoral immunity seems to wane over time and the role of T-cell immunity is yet to be studied in more detail. This issue is profound and discussion of it is lacking. Of course, immunity is complex and a comprehensive discussion is outside of the paper's scope, but it should be addressed in as far as it may impact the paper's goals: estimating IFR and its translation to total number of infection. Relevant papers to this effect:

a. Braun, Loyal, Frensch, et al. (2020). SARS-CoV-2-reactive T cells in healthy donors and patients with COVID-19. Nature, Accelerate Article Preview.

b. Danchin and Turinici (2020). Immunity after COVID-19 protection or sensitization? medRxiv.

c. Huang, Garcia-Carreras, Hitchings, et al. (2020). A systematic review of antibody mediated immunity to coronaviruses: antibody kinetics, correlates of protection, and association of antibody responses with severity of diseases. medRxiv.

d. Grzelak, Temmam, Planchais et al. (2020). SARS-CoV-2 serological analysis of COVID-19 hospitalized patients, pauci-symptomatic individuals and blood donors. medRxiv.

We agree that the role of serological assay performance was not thoroughly considered in the original manuscript and we thank the reviewer for raising this important issue. In response to point (a) we agree that it is known to take some time before antibodies become detectable and our model explicitly accounts for uncertainty in the delay between infection and seroconversion. However, we agree that our analysis would benefit from further assessment of the robustness of our estimates to the assumed mean delay to seroconversion. Our model assumes a gamma distributed delay between infection and onset with mean 6.5 days, standard deviation 2.6 days and between onset and seroconversion with mean 10 days and standard deviation 8 days. In Figure S4 (below) we show the resulting cumulative probability density functions for the infection-to-seroconversion delay when varying the mean time from onset-to-seroconversion (5, 10, 15 and 20 days). We find that our model estimates of IFR are largely robust to different assumptions regarding the mean times from onset-to-seroconversion, with only minor differences in median model estimates (Figure S4 and below).

With regards to point (b) we have revisited each seroprevalence study included in our analysis and identified the assay specificity and sensitivity, allowing us to adjust seroprevalence estimates for assay performance where this had not already been conducted by the study authors. We were able to ensure the adjustment of seroprevalence values for assay performance in 24/25 seroprevalence studies now included in our analysis (further details on this is provided in response to comment 1.2, above) and have updated our model estimates using these adjusted values. For completeness, we conducted a comparison of model estimates derived from unadjusted and adjusted seroprevalence values and find only minor discrepancies in the results (Figure S5 and below).

In point (c) the reviewer raises the key point regarding the potential for antibody decay. We use the time-series of reported deaths to reconstruct the timing of infections in each population to evaluate the timing of infections as compared to the timing of the seroprevalence studies in each setting. On average we find that 60% of all infections happened in the month prior to the seroprevalence studies and 89% in the 2 months prior (individual study range: 51-100%). Given that long-term studies of SARS-CoV and MERS-CoV have typically found levels of IgG to remain detectable for at least a year after infection (Huang et al., 2020), we do not expect the waning of seroprevalence over time to be a significant issue for our analysis at this stage of the pandemic. However, we have now conducted sensitivity analyses around the duration of seropositivity and find that in an extreme scenario of exponential decay of seroconversions at 5% per month our ensemble model estimate of the French population IFR would be only slightly reduced to 0.65% (95%CrI: 0.56-0.76%) as compared to when assuming no decay within the first 9 months (0.79%, 95%CrI: 0.68-0.92%).

We agree with the reviewer as to the uncertainties surrounding SARS-CoV-2 immunity and have taken care in presenting our estimates of infection attack rates so as to avoid any misinterpretations regarding what these mean for immunity. We completely agree that the role of T-cell immunity is a potentially important factor that requires further evaluation and have now incorporated a discussion point to this effect. We thank the reviewer for these references and include them in the revised paper.

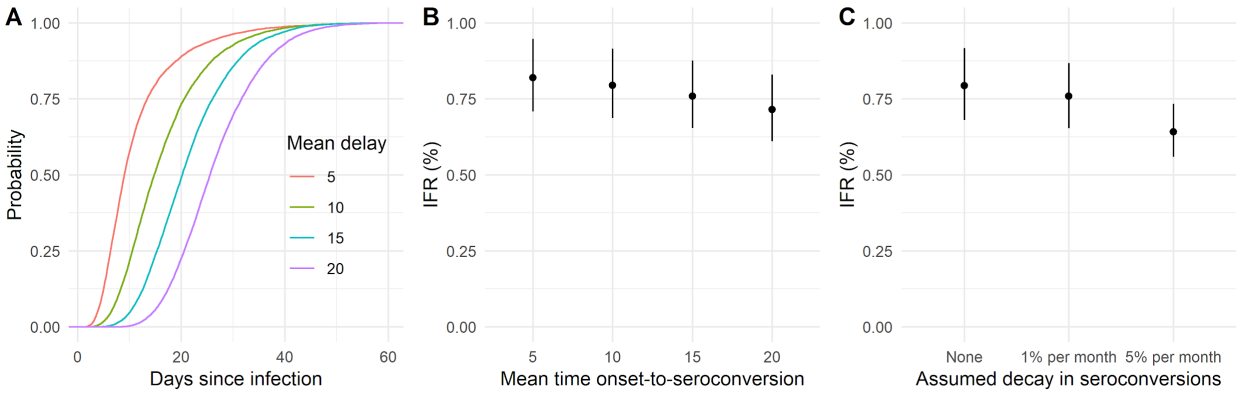


Figure S4. (A) Cumulative probability density functions for delay between infection-to-seroconversion, obtained by assuming mean times from onset-to-seroconversion of 5, 10, 15 and 20 days. (B) Model estimates of the population IFR for France under different assumptions regarding the mean time from onset-to-seroconversion. (C) Model estimates of the population

IFR for France under different assumed exponential decays in seroconversions over time (no decay, 1% decay per month and 5% decay per month).

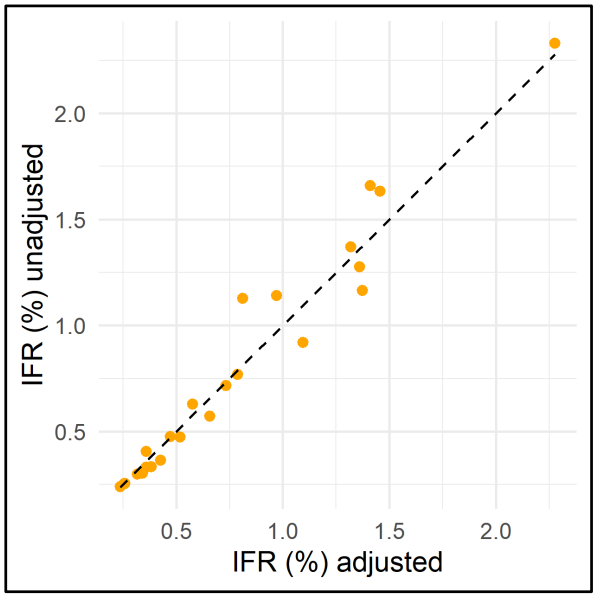


Figure S5. A comparison of population IFR estimates, using seroprevalence values that are adjusted and unadjusted for assay sensitivity and specificity. Orange points represent median model estimates of the population IFR in France derived from separately fitting individual seroprevalence studies in the model likelihood. Black dashed line represents the ideal scenario of x=y.

Specific changes to document:

- *The main model is now conducted using seroprevalence estimates adjusted for assay specificity and sensitivity.*
- *We have conducted sensitivity analyses surrounding the potential for antibody decay and varying delays to seroconversion which we show in a new figure (Figure S4).*
- *We have added a note on this in our main manuscript at line 147 "As the duration of SARS-CoV-2 seropositivity amongst infected individuals is as-yet unclear[14], in sensitivity analyses we explore the potential effect of waning antibodies over time. In an extreme scenario with assumed exponential decay of SARS-CoV-2 seroconversions at 5% per month the ensemble model estimates a population IFR of 0.65% in France (95%CrI: 0.56-0.73%) (Figure S4). Further, we demonstrate that our results are robust to different assumptions regarding the mean delay between infection and seroconversion (Figure S4)."*
- *We have also noted at line 151 "It is important to note that there may be individuals who never seroconvert and instead only develop a T-cell response, and would therefore be missed in these studies[15].*

Comment 2.2. In this sense, it might be interesting to consider alternative routes for estimating the total number of cases in a country, e.g., based on mathematical modeling, where confirmed cases, hospitalizations, ICU occupation,… are used as input:

a. Faes, Abrams, Van Beckhoven, et al. (2020). Time between symptom onset, hospitalization and recovery or death: a statistical analysis of different time-delay distributions in Belgian COVID-19 patients. medRxiv.

b. Abrams, Wambua, Santermans, et al. (2020). Modeling the early phase of the Belgian COVID-19 epidemic using a stochastic compartmental model and studying its implied future trajectories. medRxiv.

c. Willem, Abrams, Petrof, et al. (2020). The impact of contact tracing and household bubbles on deconfinement strategies for COVID-19: an individual-based modelling study. medRxiv.

We agree that other approaches that utilise case, hospitalization and ICU admission data are of immense value to understanding epidemic dynamics. We believe however, that the sole use of age-specific death data underpins the novelty and usefulness of our approach as hospitalization and ICU admission data are not frequently available and case data are subject to much dependence on surveillance system capacities. However, we agree that comparisons to alternative approaches would be very useful and now incorporate this in our analysis. We compare our estimates of the infected population proportion to those derived from other modelling efforts (Abrams et al., use hospital admission and death data; Flaxman et al., use case and death data; and Salje et al., use hospital admission, ICU admission and death data) and find them to be highly consistent (Figure S7 and below). We thank the reviewer for suggesting this valuable external validation.

Specific changes to document:

- *We have added a point in the text at line 181: 'Our estimates are also consistent with mathematical modelling efforts for individual countries, where additional metrics of epidemic size (e.g. numbers of cases, hospitalizations and/or ICU admissions) have been considered12,19,20 (Figure S7).'*
- *We show the comparison of our estimates to those of external modelling efforts in Figure S7, below.*

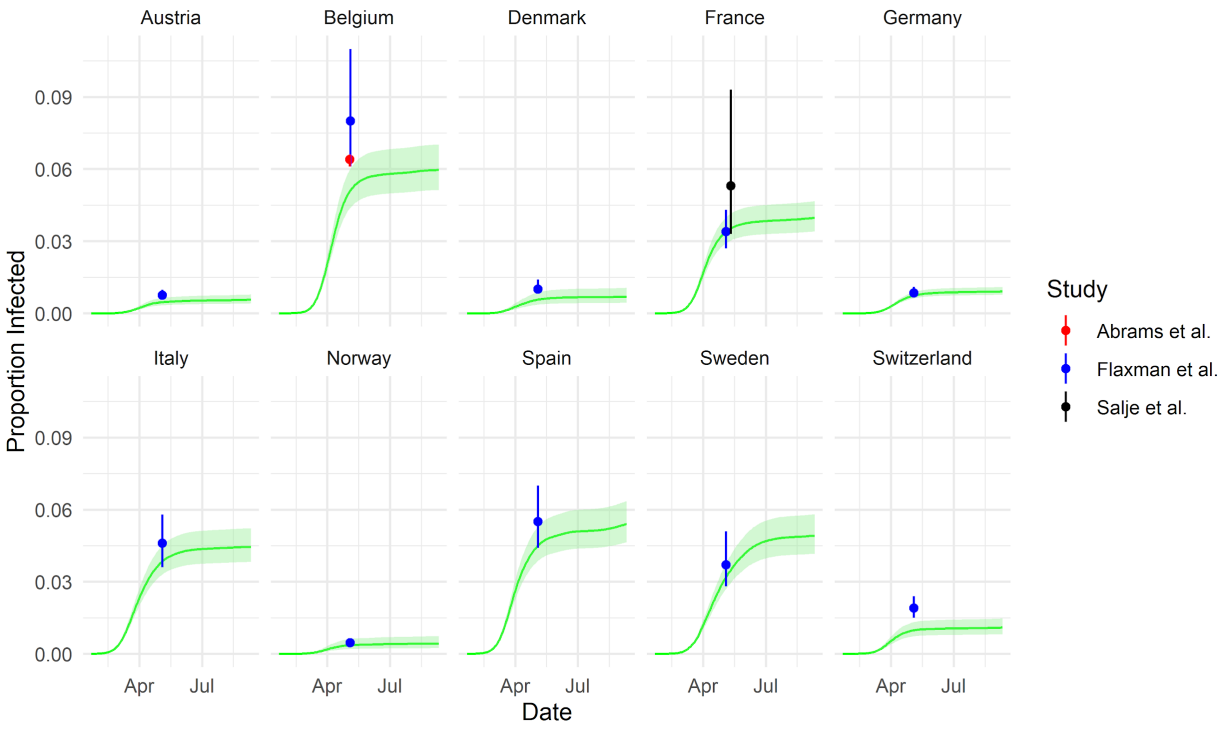


Figure S7. Comparison of ensemble model estimates of the proportion infected to those of other modelling efforts that use additional metrics of epidemic size. Green lines and ribbons indicate the median and 95% credible interval estimates of the ensemble model. Coloured points and lines represent the central estimates and their associated uncertainty reported by external analyses.

Comment 2.3. Underreporting of COVID-19 mortality relative to excess mortality. It is known that the extent of reporting of COVID-19 mortality varies across countries. While Belgium may report roughly 100% of its excess mortality as COVID-19 related mortality, in other countries this fraction shrinks to much lower levels. This might invalidate the results found if not properly accommodated. This is one of the issues discussed in:

a. Aron, J., Giattino, C., Muellbauer, J., and Ritchie, H. (2020). A pandemic primer on excess mortality statistics and their comparability across countries. https://ourworldindata.org/covid-excess-mortality.

This really is a fundamental point, as examining IFR from COVID-19 confirmed deaths versus total deaths (confirmed and suspected; in hospitals as well as in NH and other settings) makes a big difference, would lead to very different rankings of countries, etc.

We agree that there are question marks as to the completeness of COVID-19 deaths. A central feature of our paper is using death data from younger individuals (<65y olds) in driving our IFR estimates, rather than relying on deaths in older individuals, which can be heavily impacted by deaths in nursing home populations or be missing due to causes of death being less likely to be investigated in older individuals.

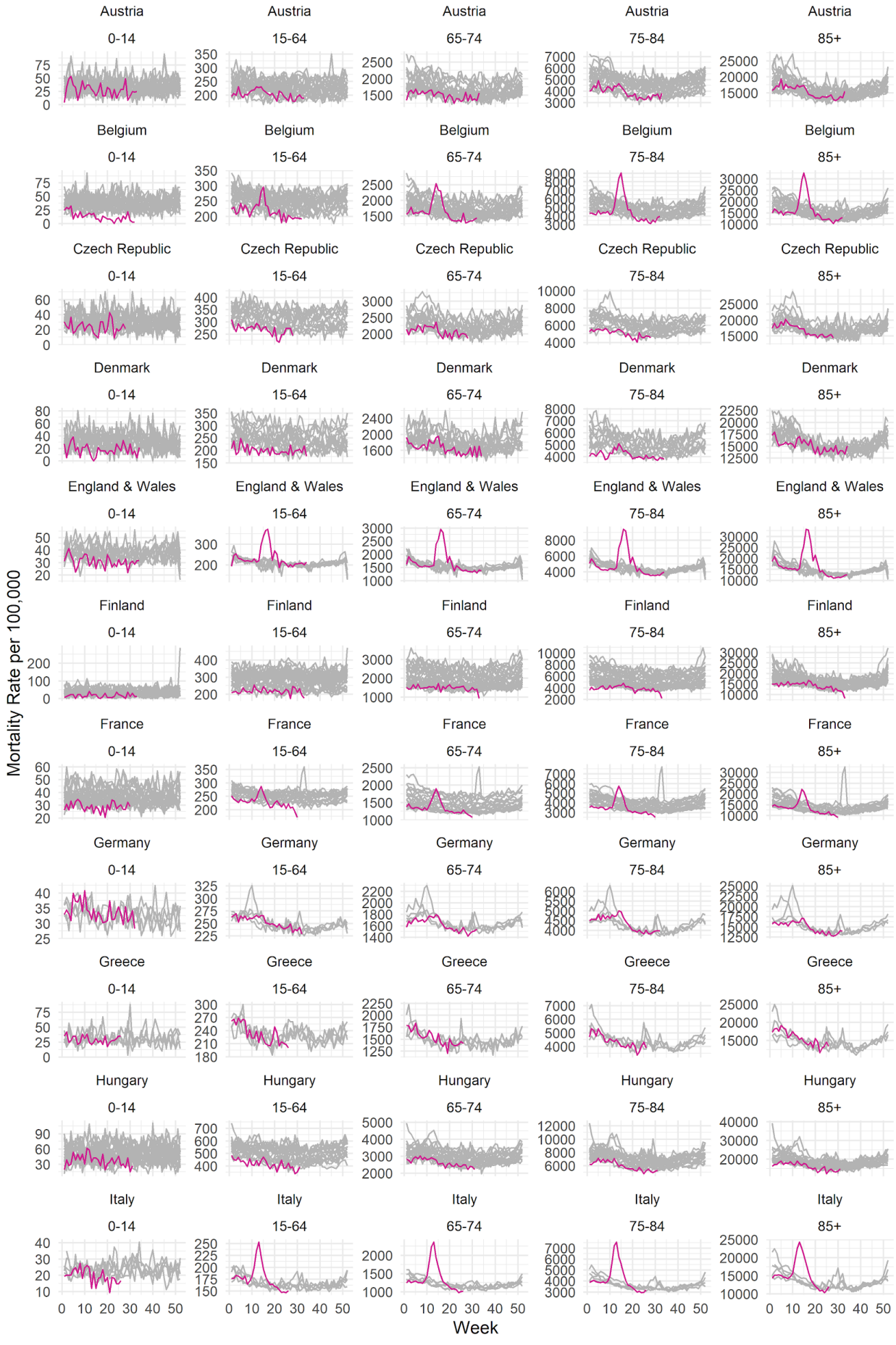
In our initial submission we investigated the consistency of COVID-19 death data with excess mortality data and found them to be broadly consistent. However, we agree that a deeper comparison could be done. We have therefore extracted age-specific excess death data from the Human Mortality Database (the underlying main data source in the reference suggested by the reviewer) for settings where this is available (22/45 countries). We find that calculating age-specific IFRs using these data is complicated as (a) the excess death calculation is highly sensitive to the reference time period used to calculate expected deaths (Figures S11-S12 and below), (b) the frequent presence of negative excess deaths, especially in younger age groups (Figure S12 and below), (c) age-bins are wide (e.g., a single bin for 15-64y olds), and (d) excess death data are rarely available in resource-poor countries. Of particular note is that, outside the clear epidemic peak, the number of observed deaths in 2020 appears to be at the bottom of the range of deaths from prior years for many countries (Figure S11). This suggests there are a multitude of complex (and difficult/impossible to measure) factors that impact the excess death calculation - e.g., changes in behaviour in lockdowns affecting risk of death from other causes (such as accidents), or from natural underlying variability in risk of death across years. However, without a detailed understanding of causes of deaths between years, it is impossible to explore this further. These observations show that it is unlikely that excess death data will be able to form the basis of COVID-19 mortality estimates in most circumstances. Instead, reported COVID-19 deaths will remain the key data source for understanding mortality from the SARS-CoV-2 epidemic going forward.

Despite these limitations, we agree that it is useful to have an understanding of what missing deaths could do to our estimates. In the main reference provided by the reviewer (https://ourworldindata.org/covid-excess-mortality) the authors provide estimates of the proportion of missing deaths for a small subset of countries using excess deaths (range from 40% undercounted to 10% overcounted). If we take the extreme example that deaths across all ages are undercounted by these amounts, this only produces a slight shift in our IFR estimates (original mean IFR of 0.66% across these 6 countries to a mean of 0.87% when adjusting for these potentially missing deaths). Note that, as mentioned in the reference, this is a very extreme example as many unaccounted deaths are likely to be in nursing home communities (e.g. 45% of deaths are attributed to nursing homes in England), which do not affect our main estimates.

We provide a more detailed discussion of the comparison with excess death data in the revised manuscript.

Specific changes to document:

- *We have added discussion on this at line 244: "Here we have used data from national reporting systems of COVID-19 associated deaths. However, in some settings these may not capture all deaths associated with COVID-19. It has been estimated for a subset of countries (N=6/45) that reported COVID-19 deaths were between 40% undercounted to 10% overcounted as compared to excess death estimates[25]. Assuming that these differences occur equally across all age groups would result in a change of mean IFR for these countries of 0.66% to 0.87%. Note that this represents an extreme scenario, as most unaccounted for deaths are likely to be in the oldest age groups, which would not affect our estimates[25]. We note that there are a number of complexities in the interpretation of excess death data that can inhibit their direct use in the assessment of IFR. Specifically, excess death estimates are highly sensitive to the reference time period used to calculate expected deaths (Figures S11-S12), there are frequently negative excess deaths, especially in younger age groups (Figure S12), the data is only available with wide age-bins and rarely available for low and middle income countries. For example, outside the clear epidemic peak, the number of observed deaths in 2020 is at the bottom of the range of deaths from prior years for many countries (Figure S11). This suggests there are a multitude of complex factors that impact the excess death calculation, potentially including changes in behaviour affecting risk of death from other causes (such as accidents), or from natural underlying variability in risk of death across years. However, without a detailed understanding of causes of deaths between years, it is impossible to explore this further. While both seroprevalence data and COVID-19 death data can be subject to potential limitations, considering these data across multiple settings in a harmonized framework allows us to robustly assess trends in the transmission and fatality rates of SARS-CoV-2 and derive global ensemble estimates.*


Austria
Belgium
Czech Republic
Denmark
England & Wales
Finland
France
Germany
Greece
Hungary
Italy
0-14
15-64
65-74
75-84
85+
Mortality Rate per 100,000
Week
0 10 20 30 40 50

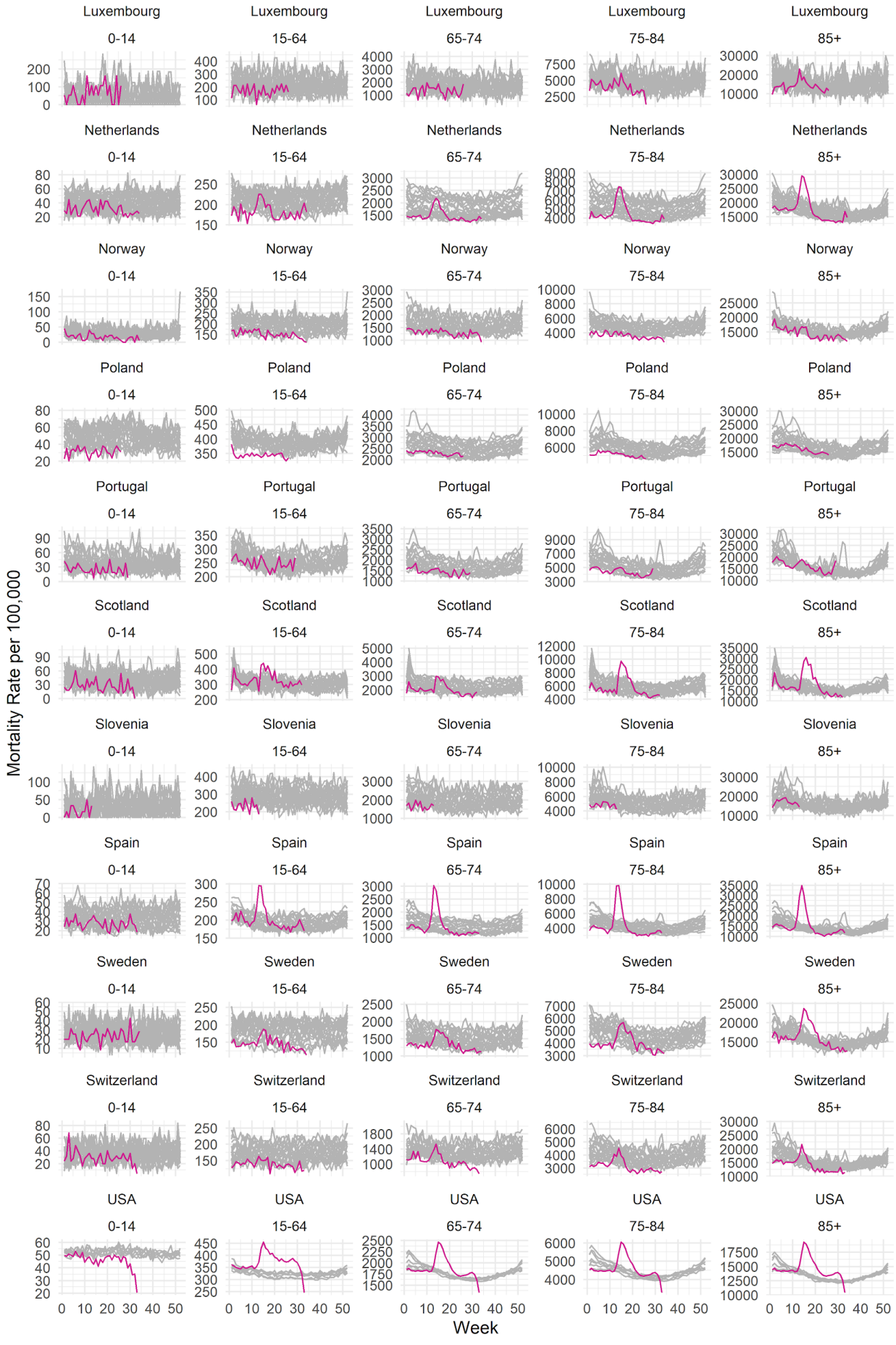

Luxembourg
Netherlands
Norway
Poland
Portugal
Scotland
Slovenia
Spain
Sweden
Switzerland
USA
0-14
15-64
65-74
75-84
85+
Mortality Rate per 100,000
Week

Figure S11. Age-specific mortality rates per 100,000 population in 22 countries from the Human Mortality Database. Grey lines indicate weekly mortality rates for all available years prior to 2020 and the red line represents weekly mortality rates in 2020.

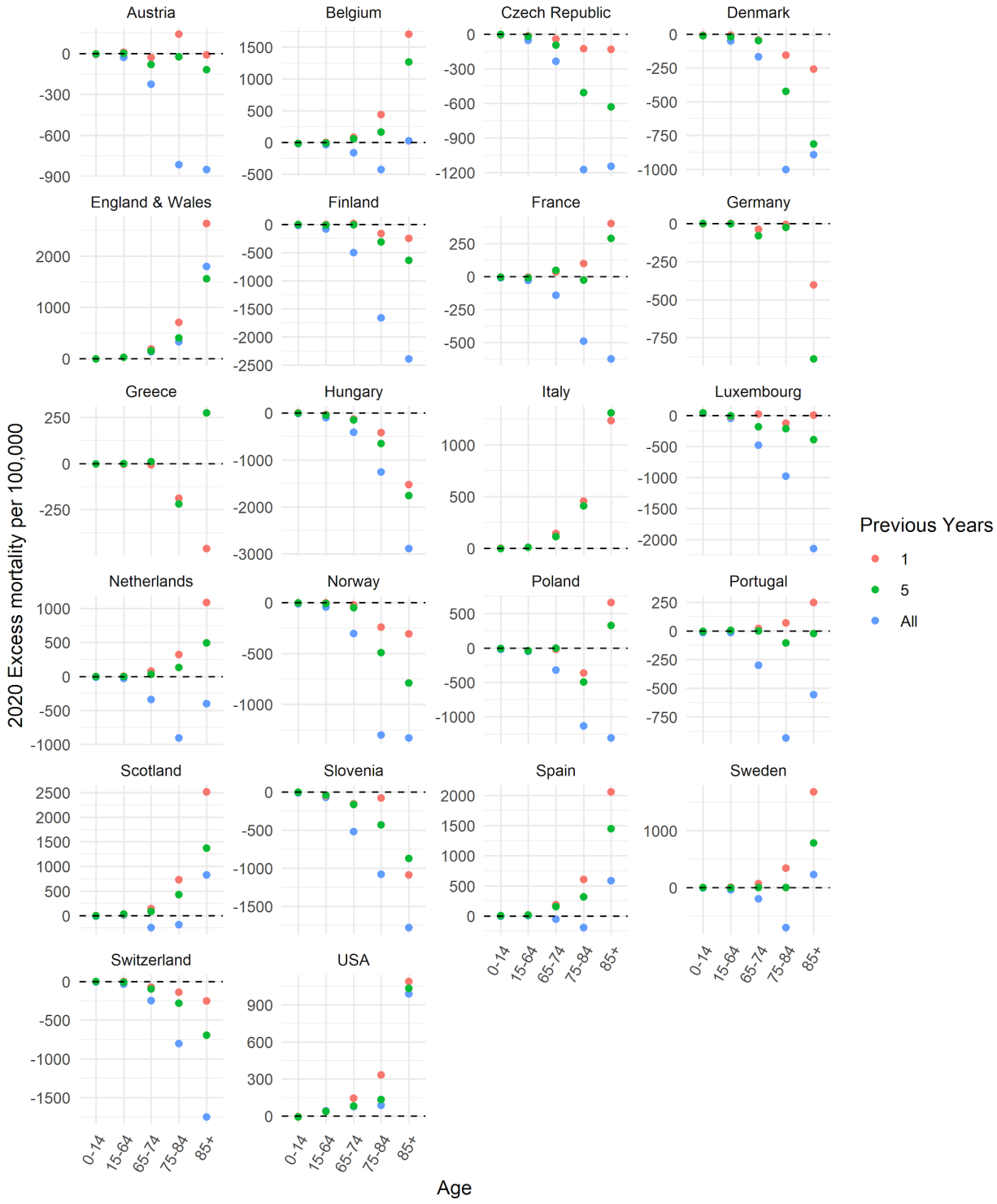


Figure S12. Age-specific excess mortality in 2020 as compared to the average of previous years for the same time period. Coloured points show mortality rates as compared to the same time

period in 2019 (coral), 2015-2019 (green) and all previous years (blue). Black dashed lines indicate values of zero excess deaths.

Comment 2.4. Related work. There is some related work regarding global IFR. For example, the paper by Grewelle and De Leo (2020) is relevant in this regard, and arguably should be cited.

a. Grewelle RE, De Leo GA (2020). Estimating the global infection fatality rate of COVID-19. medRxiv.

We thank the reviewer for their suggestion. The analysis by Grewelle & De Leo takes quite a different approach to that of our analysis, deriving their IFR estimates from case fatality data and do not use age in their estimation. We therefore have some reservations about this approach given the particular limitations of case surveillance data and the estimation of a single IFR for all countries when it is now clear that countries with younger populations will have very different IFRs to countries with older populations. However, we agree that comparisons to other independent estimates are important. As suggested in comment 2.2 above, we make comparisons with external estimates of the proportion of the population infected for a range of different countries.

Comment 2.5. Reference age category. I can see the rationale for choosing a well-argued reference age category (e.g., 60-65). Given the relatively low rates at younger ages, and the complexity at older ages because of the convolution of NH and non-NH populations. But still, it is a narrow age band, and a sensitivity analysis should be done. Also, it should be clear that associated statistical uncertainty is properly taken into account. Generally, there are a number of estimates (in text, in tables), where either a point estimate only is given (e.g., Table S2) or where the method used to estimate precision is not explained. So, throughout the methodological appendix, precision estimation should be clearly discussed, and backed up with the appropriate formulas or numerical algorithm.

We thank the reviewer for raising this point. We use a reference age-group as a simple means to describe the overall trends in risk of mortality by age across settings, as observed from our data, in Figure 1. Note that our model does not rely on any reference age-group however, so this choice has no bearing on our main analysis. The choice of a reference age-category to describe the data is further complicated by the variable age-groupings used by different countries to report their data. We show a sensitivity analysis below where we use different age categories as the reference (those with an upper bound of 39 and 49) and obtain a consistent log-linear pattern (Figure below). The reviewer is correct that use of younger age-groups as the reference can be problematic, as in instances where zero deaths have occurred in the reference group, the relative risks in other age-groups become infinite (Figure below). However, our general observations of a log-linear trend amongst <65s and heterogenous risks observed in those 65+ remains. In table S2 we report the proportion of all COVID-19 deaths that have been attributed to nursing home residents in multiple countries, as reported by each of the national surveillance systems. This is an observed proportion (i.e. not from a sample) and therefore there is no sensible measure of

uncertainty that can be estimated around this. We have now clarified how model uncertainty is calculated in our methods section and thank the reviewer for noticing this omission.

Specific changes to document:

- *We have conducted a sensitivity analysis on the reference age category used to describe the relative risk of death observed in the data (below).*
- *In the methods section we have clarified how we obtain our estimates of uncertainty "95% credible intervals (CrI) are calculated by taking the 0.025 and 0.975 quantiles of the posterior distribution.".*

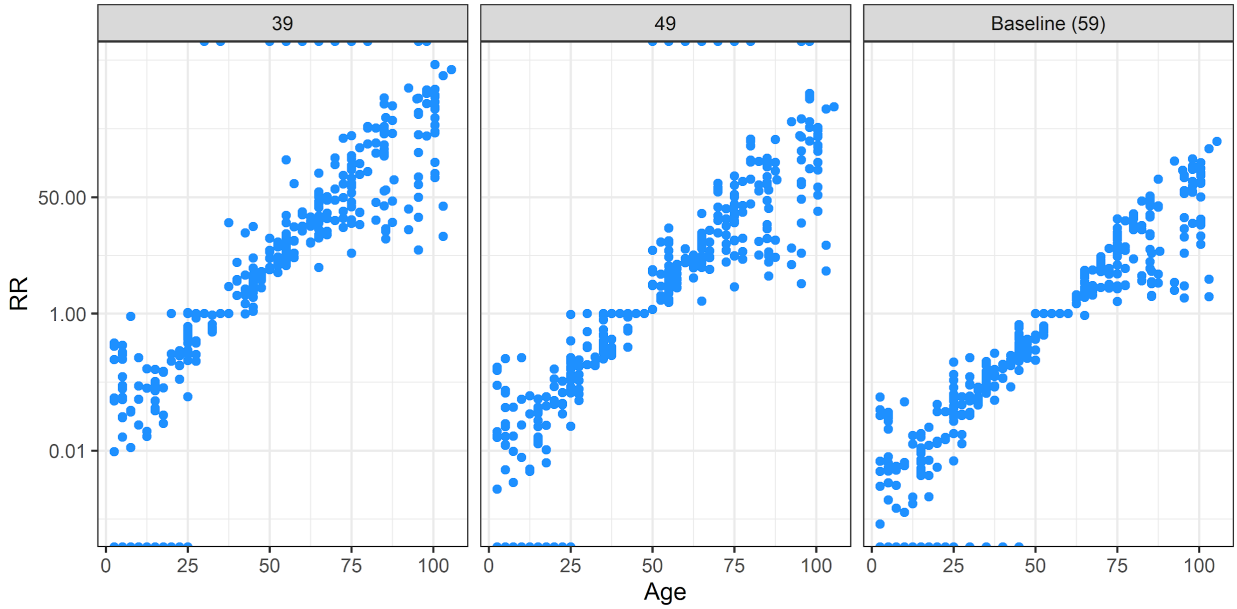


Figure. Relative risk of COVID-19 death by age across the 45 countries, with varying reference age categories (those with an upper bound of 39, 49, or 59-64). Points are plotted at the mid-point of the reported age-groups.

Comment 2.6. Nursing home population. The authors are well aware of the specificity of the NH population. Other authors have indicated very strong differences between NH and non-NH IFR's and related quantities:

a. Molenberghs, G., Faes, C., Aerts, J., et al (2020). Belgian COVID-19 mortality, excess deaths, number of deaths per million, and infection fatality rates (8 March -- 9 May 2020). medRxiv.

Thank you for pointing us to this very useful analysis that directly assesses the issue of higher frailties in Belgian nursing home residents. In addition to this paper, we have identified another analysis from Belgium that has specifically estimated the relative frailty between nursing homes and the general population (see response to Comment 2.7) that we now use to help guide the estimates. We now refer to both these analyses in our manuscript.

Specific changes to document:

- *We now reference this paper in our manuscript.*

Comment 2.7. The authors also refer to different frailty levels in the NH and non-NH populations, but I am lacking a clear, logical buildup. Ideally, the NH and non-NH populations should be separated completely, so as to create age-sex-NH strata, rather than merely age-sex strata. The sensitivity analysis done in lines 222-243 is a bit rudimentary; it is not plausible to assume equal frailty, except as a 'most extreme' scenario. A detailed discussion on frailty is given in:

a. Declercq, A., de Stampa, M., Geffen, L., et al. (2020). Why, in almost all countries, was residential care for older people so badly affected by COVID-19? OSE Working Paper Series, Opinion Paper No. 23. Brussels: European Social Observatory.

We apologise for the confusion. The analysis was conducted exactly as the reviewer sets out - with age-sex-NH strata. In the main part of the paper, we concentrate just on the non-NH population - this allows us to have comparable estimates across locations. We then specifically explore the impact of infections and deaths in the NH population on overall IFR - demonstrating how outbreaks this fragile population can drive overall IFR estimates.

We agree that assuming equal frailty is not a realistic assumption. We have now identified an analysis (https://www.medrxiv.org/content/10.1101/2020.08.29.20183210v1) that specifically estimates the relative frailty between nursing home residents and the general population with adjustments for age. It finds that individuals in nursing homes are 3.8 more likely to die following infection than people of the same age in the general population. We use this number to guide our estimates of the IFR when incorporating nursing home deaths. We also thank the reviewer for highlighting this paper which we now cite in our revised manuscript.

Specific changes to document:

- *We have updated the analysis and the text at line 212: 'In our baseline model we derive IFR estimates amongst the general population (i.e. excluding nursing home deaths) so as to facilitate robust comparisons of IFR and general population transmission across settings. However, we demonstrate that where high rates of infection have occurred amongst nursing home residents, overall IFR estimates will be significantly greater than in scenarios where these populations have been successfully shielded or experienced little exposure (Figure 4C). For example, in France, including deaths in nursing homes, increases the IFR from 0.74% for the general population (95%CrI: 0.64-0.86%) to 1.10% overall (95%CrI: 0.95-1.28%). This highlights the complexity in comparing headline IFR estimates across populations where very different levels of transmission may have occurred in these hyper-vulnerable communities.'*
- *We now use higher estimates of the relative frailty of nursing home residents (2, 3.8 and 6) in our analysis of the contribution of nursing home transmission to IFR estimates. Figures 4B and 4C have been updated accordingly and the main text now reads "Using France as a reference population, we use the age and sex distribution of nursing home residents to derive a population-weighted IFR of 22.25% (95%CrI: 19.06-25.74%) among*

French nursing home residents, assuming individuals in nursing homes are 3.8 times more frail than individuals in the general population of the same age and sex as estimated in Belgian nursing home residents[24] (Figure 4B).".

Comment 2.8. Detail: when used the first time in text 'CrI' should be defined.

We thank the reviewer for noticing this discrepancy and have now corrected this in the text.

Specific changes to document:

- *We now define CrI when first appearing in the text.*

Reviewer Reports on the First Revision:

Referees' comments:

Referee #1 (Remarks to the Author):

The authors have done a good job dealing with my initial concerns, particularly with regards to frailty and relative attack rates in nursing home populations. Overall I feel this is a valuable contribution to the discussion on COVID-19 fatality rates, and I congratulate the authors on a good piece of work.

Referee #2 (Remarks to the Author):

I have had the chance to review the revised version of the manuscript. The authors have conducted a very thorough and, in fact, impressive review. The comments, many of which admittedly tricky, have been dealt with adequately.

Referee #3 (Remarks to the Author):

This is a most elegant demonstration of how a systematic, careful series of inferential analyses, as a harmonised ensemble model, based on known distributions and relativities can yield robust insights into population COVID-19 burden. It is the most thoughtful set of findings synthesised to date.

Specific comments:
1. Lines 115-6 – How reasonable is the assumption of equal infection attack rates across age groups <65 years? Fig S9 shows 2 more scenarios but how about other different permutations of age-dependent susceptibility and infectiousness?
2. Achilles heel of the ensemble model – The frailest findings concern the population-weighted IFR estimates by the ensemble model illustrated in Fig 3. Model predictions vs empirical seroprevalence in Fig 3B require more detailed and justified dissection.
3. Did the authors consider time-varying IFR due to improved knowledge thus management regimens as the pandemic has progressed? For that matter health care resources, in terms of surge capacity of the health system as well as medications and equipment, are also differentially distributed across countries.

I have also reviewed the point-by-point response to the two original reviewers' comments. I find the authors' subsequent revision satisfactory.

Author Rebuttals to First Revision:

Referee #1.

Comment 1.1. The authors have done a good job dealing with my initial concerns, particularly with regards to frailty and relative attack rates in nursing home populations. Overall I feel this is a valuable contribution to the discussion on COVID-19 fatality rates, and I congratulate the authors on a good piece of work.

We thank Referee #1 for the kind words and for their very helpful review of our analysis.

Referee #2.

Comment 2.1. I have had the chance to review the revised version of the manuscript. The authors have conducted a very thorough and, in fact, impressive review. The comments, many of which admittedly tricky, have been dealt with adequately.

We are very glad that Referee #2 has found our revision satisfactory and extend our sincere thanks to them for their feedback.

Referee #3.

Comment 3.1. This is a most elegant demonstration of how a systematic, careful series of inferential analyses, as a harmonised ensemble model, based on known distributions and relativities can yield robust insights into population COVID-19 burden. It is the most thoughtful set of findings synthesised to date.

We thank Referee #3 for their extremely generous praise of our analysis.

Comment 3.2. Lines 115-6 – How reasonable is the assumption of equal infection attack rates across age groups <65 years? Fig S9 shows 2 more scenarios but how about other different permutations of age-dependent susceptibility and infectiousness?

We agree with the reviewer that assuming equal infection attack rates by age for those <65 years is a simplistic assumption, made necessary by complex and variable contact patterns that are likely to have occurred throughout the SARS-CoV-2 pandemic. Large heterogeneities in the magnitude and timing of implemented non-pharmaceutical interventions across countries has made systematic quantification of age-specific contact rates essentially impossible. We therefore agree with the reviewer that additional permutations of different age-dependent infection attack-rates would be beneficial. The previous version of the model contained three different scenarios: (a) the main model of reduced attack rates in those >65y compared to the rest of the population reflecting reduced contacts/increased shielding in older people; (b) additional increased attack rate in 20-40 year olds reflecting increased contacts in young adults; (c) equal attack rates across all ages. In the revised manuscript, we now include additional sensitivity analyses where we consider reduced attack rates in the youngest age-groups (<20 year olds) to

reflect reduced susceptibility or infectiousness in this age group. We consider this reduced susceptibility alongside reduced attack rates in oldest individuals and separately where we include increased attack rates in young adults too. The sensitivity analyses only result in small changes in overall IFR (Figure S9).

Comment 3.3. Achilles heel of the ensemble model – The frailest findings concern the population-weighted IFR estimates by the ensemble model illustrated in Fig 3. Model predictions vs empirical seroprevalence in Fig 3B require more detailed and justified dissection.

We believe that our finding that the ensemble model cannot capture the observed seropositivity of strong interest. In our analysis, we demonstrate that population demographics, variable reporting of deaths in elderly populations, antibody dynamics, delays from infection to death, and reported assay sensitivity/specificity cannot explain all of the differences between what is estimated by the ensemble model and what is directly measured. It is possible that differences in the representativeness of the seroprevalence studies are driving some of the heterogeneity and we have noted this throughout the results. Given what is known about risk-factors for severe COVID-19 disease, there remains the strong probability that the underlying health of populations are playing a large role. Ultimately, given the currently available data, it is impossible to disentangle the drivers of these differences further and we have tried to make this clear in our results: "Potential explanations for the variable IFR estimates observed across settings include different prevalences of high-risk populations (e.g. individuals with comorbidities), differences in methodology and representativeness of seroprevalence studies, heterogeneities in availability and quality of care or variations in reporting of COVID-19 deaths. We have fit our model to seroprevalence data adjusted for reported assay sensitivity and specificity but find that using unadjusted estimates provides similar results (Figure S5). As the duration of SARS-CoV-2 seropositivity amongst infected individuals is as-yet unclear14, in sensitivity analyses we explore the potential effect of waning antibodies over time. In an extreme scenario with assumed 5% exponential decay of seroconversions per month the ensemble model estimates a population IFR of 0.65% in France (95%CrI: 0.56-0.73%) (Figure S4). Further, we demonstrate that our results are robust to different assumptions regarding the mean delay between infection and seroconversion (Figure S4)."

In light of the fact that the data does not allow us to disentangle these factors, we also strongly highlight the need to consider the full range of uncertainty derived from the models we fit separately to each individual seroprevalence survey: "Given the underlying heterogeneity in IFR that could not be captured by the ensemble model, it is important to consider the full range of uncertainty in these estimates as suggested by individual seroprevalence studies (grey points in Figure 3B)."

Comment 3.4. Did the authors consider time-varying IFR due to improved knowledge thus management regimens as the pandemic has progressed? For that matter health care resources, in terms of surge capacity of the health system as well as medications and equipment, are also differentially distributed across countries.

We thank the reviewer for raising this important point. It is possible that there have been improvements in outcomes, however, we cannot explore this without more detailed understanding of the different practices across the different countries and their likely impact. The results from our paper will represent an average IFR across the pandemic

to-date. We now note this important point in the methods section of the revised manuscript: “This assumes that age- and sex-specific IFRs are constant over the course of the pandemic. Where improvements in COVID-19 outcomes have occurred over time, our estimates would represent the average probabilities to-date.”

Comment 3.5. I have also reviewed the point-by-point response to the two original reviewers’ comments. I find the authors’ subsequent revision satisfactory.

Thank you