INFERRING THE COVID-19 INFECTION FATALITY RATE IN THE COMMUNITY-DWELLING POPULATION: A SIMPLE BAYESIAN EVIDENCE SYNTHESIS OF SEROPREVALENCE STUDY DATA AND IMPRECISE MORTALITY DATA

Harlan Campbell
Department of Statistics, University of British Columbia
Vancouver, British Columbia, Canada
harlan.campbell@stat.ubc.ca
ORCID ID: 0000-0002-0959-594

Paul Gustafson
Department of Statistics, University of British Columbia
Vancouver, British Columbia, Canada
ORCID ID: 0000-0002-2375-5006

November 2, 2021

Acknowledgments - This work was supported by the European Union’s Horizon 2020 research and innovation programme under ReCoDID grant agreement No 825746 (■) and by the Canadian Institutes of Health Research, Institute of Genetics (CIHR-IG) under Grant Agreement No 01886-00-0. We wish to thank Andrew Levin, Thomas Debray, Valentijn de Jong, and Thomas Jaenisch for their valuable feedback. We also wish to thank Ariel Karlinsky, Nana Owusu-Boaitey, Lauren Maxwell, and Sayali Arvind Chavan for help with data collection. Finally, thank you to the many fellow researchers who helped by providing information about various data including, amongst others, Maurizio Napolitano, Shiwani Mahajan, Miguel Quartin, Stefano Lombardo, and Timon Gaertner.

This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.
ABSTRACT - Estimating the COVID-19 infection fatality rate (IFR) has proven to be particularly challenging—and rather controversial—due to the fact that both the data on deaths and the data on the number of individuals infected are subject to many different biases. We consider a Bayesian evidence synthesis approach which, while simple enough for researchers to understand and use, accounts for many important sources of uncertainty inherent in both the seroprevalence and mortality data. With the understanding that the results of one’s evidence synthesis analysis may be largely driven by which studies are included and which are excluded, we conduct two separate parallel analyses based on two lists of eligible studies obtained from two different research teams. The results from both analyses are rather similar. With the first analysis, we estimate the COVID-19 IFR to be 0.31% (95% credible interval of (0.16%, 0.53%)) for a typical community-dwelling population where 9% of the population is aged over 65 years and where the gross domestic product at purchasing-power parity (GDP at PPP) per capita is $17.8k (the approximate worldwide average). With the second analysis, we obtain 0.32% (95% credible interval of (0.19%, 0.47%)). Our results suggest that, as one might expect, lower IFRs are associated with younger populations (and may also be associated with wealthier populations). For a typical community-dwelling population with the age and wealth of the United States we obtain IFR estimates of 0.43% and 0.41%; and with the age and wealth of the European Union, we obtain IFR estimates of 0.67% and 0.51%. Keywords: COVID-19; evidence synthesis; Bayesian inference; infection fatality rate.

Above all, what’s needed is humility in the face of an intricately evolving body of evidence. The pandemic could well drift or shift into something that defies our best efforts to model and characterize it.

Siddhartha Mukherjee, *The New Yorker*
February 22, 2021

1 Introduction

The infection fatality ratio (IFR), defined as the proportion of individuals infected who will go on to die as a result of their infection, is a crucial statistic for understanding SARS-CoV-2 and the ongoing COVID-19 pandemic. Estimating the COVID-19 IFR has proven to be particularly challenging—and rather controversial—due to the fact that both the data on deaths and the data on the number of individuals infected are subject to many different biases.

SARS-CoV-2 seroprevalence studies can help provide a better understanding of the true number of infections in a given population and for this reason several researchers have sought to leverage seroprevalence study data to infer the COVID-19 IFR [21]. In particular, Ioannidis [33], Levin et al.
Brazeau et al. [13], and O’Driscoll et al. [59] have all undertaken analyses, of varying degrees of complexity, in which they combine data from multiple seroprevalence studies with available mortality statistics to derive IFR estimates.

The analyses of both Brazeau et al. [13] and O’Driscoll et al. [59] are done using rather complex Bayesian models which rely on numerous detailed assumptions. For instance, Brazeau et al. [13] use a Bayesian “statistical age-based model that incorporates delays from onset of infection to seroconversion and onset of infection to death, differences in IFR and infection rates by age, and the uncertainty in the serosample collection time and the sensitivity and specificity of serological tests.” O’Driscoll et al. [59] employ a Bayesian ensemble model which assumes “a gamma-distributed delay between onset [of infection] and death” and assumes different risks of infection for “individuals aged 65 years and older, relative to those under 65.” While these analyses go to great lengths to account for the various sources of uncertainty in the data, the complexity of the models will no doubt make it challenging for other researchers to fit these models to different data in a constantly evolving pandemic.

In contrast, the analyses of Ioannidis [33] and Levin et al. [43] are decidedly more simple. For each seroprevalence study under consideration, Ioannidis [33] counts the cumulative number of deaths (from the beginning of the pandemic) until 7 days after the study mid-point (or until the date the study authors suggest), and divides this number of deaths by the estimated number of (previous or current) infections to obtain a study-specific IFR estimate. A “location specific” IFR estimate is then obtained by taking a weighted (by the study’s sample size) average of the study-specific IFR estimates for a given location (i.e., for a given country or state). Ioannidis [33] then calculates the median of all the location specific IFR estimates. No uncertainty interval for this estimate is provided. As such, it is impossible to determine what level of confidence one should place in Ioannidis [33]’s estimates.

The analysis of Levin et al. [43] is based on a standard frequentist random-effects meta-analysis model. For each age-group and seroprevalence study under consideration, Levin et al. [43] calculate a 95% confidence interval (CI) for a study-specific IFR by counting the cumulative number of deaths up until 4 weeks after the study mid-point and dividing this number of deaths by the estimated upper and lower bounds of the number of infected individuals. The meta-analysis model then combines each of these study-specific IFRs. While this analysis provides standard confidence intervals and is relatively straightforward, it does not take into account certain important sources of uncertainty (to be discussed in Section 2).

The analysis method we propose is simple enough for researchers to easily understand and use, while accounting for important sources of uncertainty inherent in both the seroprevalence data and the mortality data. Similar Bayesian models have been used previously for evidence synthesis of seroprevalence data for other infectious diseases (e.g., Brody-Moore [14]). We will apply the method
in an analysis with the objective of estimating the average COVID-19 IFR in a community-dwelling population with a certain approximate age-composition and wealth.

A major part in any evidence synthesis is determining which studies to consider within the analysis. Determining appropriate inclusion and exclusion criteria for seroprevalence studies is a rather complicated and delicate issue when it comes to estimating the COVID-19 IFR \[8, 34\]. Reviewing and evaluating the merits of the hundreds of available seroprevalence studies also involves a tremendous amount of review work and time. Fortunately, both Chen et al. [20] and Arora et al. [1] have done comprehensive and thorough reviews to ascertain study quality (i.e., risk of bias). We will work from these two lists to conduct two separate parallel analyses. This approach—conducting two analyses based on two distinct and independent literature reviews—will allow us to better understand the impact of different inclusion and exclusion criteria [26]. We will review the data and how it was obtained in Section 3, following a review of the methods in Section 2. In Section 4, we summarize the results of our analysis and conclude in Section 5.

2 Methods

2.1 Bayesian model for evidence synthesis

Suppose we have data from \(K\) different seroprevalence studies. Then, for \(k = 1, \ldots, K\), let:

- \(T_k\) be the total number of individuals tested in the \(k\)-th study;
- \(CC_k\) be the total number of confirmed cases (of past or current infection) resulting from those tested in the \(k\)-th study;
- \(P_k\) be the number of individuals at risk of infection in the population of interest for the \(k\)-th study; and
- \(D_k\) be the total number of observed deaths (cumulative since pandemic onset) in the population of interest that are attributed to infection.

We do not observe the following latent (i.e., unknown) variables; for \(k = 1, \ldots, K\), let:

- \(C_k\) be the total number of infected people (cases) in the \(k\)-th population;
- \(IR_k\) be the true infection rate (proportion of the \(k\)-th population which has been infected), which is the expected value of \(C_k/P_k\); and
- \(IFR_k\) be the true underlying infection fatality rate, which is the expected value of \(D_k/C_k\) (given \(C_k\)).
We will make a series of simple binomial assumptions such that, for \( k = 1, \ldots, K \):

\[
CC_k \sim \text{Binom}(T_k, C_k/P_k),
\]
\[
C_k \sim \text{Binom}(P_k, IR_k), \quad \text{and}
\]
\[
D_k|C_k \sim \text{Binom}(C_k, IFR_k).
\]

We wish to emphasize the importance of the third “\( D|C \)” binomial distribution above. Failing to account for the conditional distribution of the deaths given the cases may lead to inappropriately precise estimates of the IFR. For example, Streeck et al. [86] (in their original preprint (medRxiv, May 8, 2020)) calculate an uncertainty interval for the IFR by dividing the number of deaths \( (D = 7) \) by the upper and lower bounds of the 95% CI for the number of infections \( (95\% \text{ CI for } C = [1.551, 2.389]) \). Doing so, they obtain a relatively narrow 95% CI for the IFR: \( [0.17\%, 0.45\%] = [7/1.551, 7/2.389] \). In the published version of their article (Nature Communications, November 17, 2020), an alternative interval “accounting for uncertainty in the number of recorded deaths” is provided. This alternative interval, which essentially takes into account the \( D|C \) binomial distribution, is substantially wider: \( [0.17\%; 0.77\%] \). In a very similar way, Levin et al. [43] also fail to take into account the \( D|C \) binomial distribution when estimating study-specific IFRs resulting in spuriously precise study-specific IFR estimates.

Having established simple binomial distributions for the study-specific IRs and IFRs, we define a simple random-effects model such that, for \( k = 1, \ldots, K \):

\[
g(IR_k) \sim \mathcal{N}(\beta, \sigma^2),
\]
\[
g(IFR_k) \sim \mathcal{N}(\theta_0 + \theta_1 Z_{1k} + \theta_2 Z_{2k}, \tau^2),
\]

where \( \theta_0 \) represents the mean \( \text{g(infection fatality rate)} \), \( \tau^2 \) represents between group infection fatality rate heterogeneity, \( \beta \) represents the mean \( \text{g(infection rate)} \), \( \sigma^2 \) describes the variability in infection rates across the \( K \) groups, \( Z_{1k} \) and \( Z_{2k} \) are covariates of interest that may be related to the infection fatality rate by means of the \( \theta_1 \) and \( \theta_2 \) parameters, and \( g() \) is a given link function. In our analysis, we define \( g() \) as the complimentary log-log link function (cloglog), though there are other sensible choices including the logit and probit functions. As for the two covariates, \( Z_{1k} \) and \( Z_{2k} \), we will define them as the centered and scaled logarithm of the proportion of the population aged over 65 years \( (65_{yo_k}) \), and of the GDP (at purchasing power parity (PPP)) per capita \( (GDP_{pk}) \), respectively.

The model is considered within a Bayesian framework requiring the specification of priors for the unknown parameters. Our strategy for priors is to assume weakly informative priors. Beta, Normal, and half-Normal priors (following the recommendations of Gelman et al. [27] and Kümerer et al. [40]) are set accordingly: \( g^{-1}(\theta) \sim \text{Beta}(0.3, 3); g^{-1}(\beta) \sim \text{Beta}(1, 30); \theta_1 \sim \mathcal{N}(0, 10); \theta_2 \sim \mathcal{N}(0, 10); \sigma \sim \text{half-}\mathcal{N}(0, 10); \) and \( \tau \sim \text{half-}\mathcal{N}(0, 10) \). Note that the performance of any
Bayesian estimator will depend on the choice of priors and that this choice can substantially influence the posterior when few data are available [2, 41]. In the Supplementary Material, we show results obtained with an alternative set of priors as a sensitivity analysis.

2.2 Uncertainty in infection rates

While some seroprevalence studies report the exact number of individuals tested and the exact number of confirmed cases amongst those tested, to obtain estimates for the infection rate there are typically numerous adjustments made (e.g., adjusting for imperfect diagnostic test accuracy, adjusting for clustering of individuals within a household). For this reason, the sample size of a given study might not be a reliable indicator of its precision and weighting a study’s contribution in an evidence synthesis based solely on its sample size (as in e.g., Ioannidis [33]) may not be appropriate.

Rather than work with the raw testing numbers published in the seroprevalence studies, we calculate effective data values for $T_k$ and $CC_k$ based on a binomial distribution that corresponds to the reported 95% CI for the IR. By “inverting uncertainty intervals” in this way, we are able to properly use the adjusted numbers provided. (This is a similar approach to the strategy employed by Küpper et al. [40].) Tables 1 and 2 list the 95% uncertainty intervals obtained from each of the seroprevalence studies in our two parallel analyses and Tables 3 and 4 list the corresponding values for $T_k$ and $CC_k$.

It must be noted that, as Ioannidis [33] cautions, it is possible that under our “inverting uncertainty intervals” approach, poorly conducted seroprevalence studies which fail to make proper adjustments (and thereby have spuriously narrower uncertainty intervals) receive more weight in our analysis, while high-quality studies, which make proper adjustments, are unfairly penalized. Ioannidis [33] notes that the strategy of “weighting the study-specific infection fatality rates by the sample size of each study” avoids giving more weight to studies “with seemingly narrower confidence intervals because of poor or no adjustments, while still giving more weight to larger studies.” Since we are restricting our analysis to only those supposedly high quality seroprevalence studies, we hope to largely avoid this issue. Weighting studies based on their true precision is obviously the goal in any evidence synthesis, and we recognize that this is particularly difficult when so many studies may misrepresent the precision of their estimates [12, 15].

2.3 Uncertainty in mortality

Matching prevalence estimates with a relevant number of fatalities is a difficult task. Prevalence estimates obtained from a seroprevalence study do not typically correspond to a specific date. Instead, these estimates will correspond to a window of time during which testing occurred. This period may be only a few days for some studies (e.g., 4 days for Petersen et al. [64]), but can also
be several weeks or months for others (e.g., 135 days for Ward et al. [93]). Tables 1 and 2 list the sampling window start and end dates for each of the studies in our two parallel analyses.

Evidently, a longer sampling window will lead to greater uncertainty when it comes to establishing the relevant number of deaths. It can be difficult to account for this uncertainty and analyses will often simply select a specific date at which to count deaths based on some simple rule of thumb. For example, Ioannidis [33] considers the number of deaths at 7 days after the mid-point of the sampling window (or as the relevant number of deaths discussed by the seroprevalence study’s authors). As another example, Meyerowitz-Katz and Merone [51] take the number of deaths as recorded at 10 days after the end of the sampling window. While these two particular analytical choices are not all that different, each may lead to a substantially different number of deaths for a given study if the study was conducted during a period of time during which the number of deaths was rapidly accelerating. Levin et al. [43], who consider the number of deaths up until 4 weeks after the sampling window mid-point, acknowledge this limitation noting that “matching prevalence estimates with subsequent fatalities is not feasible if a seroprevalence study was conducted in the midst of an accelerating outbreak.”

In order to account for the uncertainty in selecting the relevant number of deaths for a given seroprevalence study, we propose considering the number of deaths as interval censored data. Tables 3 and 4 list numbers for an interval corresponding to the number of deaths recorded 14 days after the start of the sampling window and 14 days after the end of sampling window for each seroprevalence study. While we might not know exactly what number of deaths is most appropriate, we can be fairly confident that the appropriate number lies somewhere within this interval. (Note that some intervals in Tables 3 and 4 have also been widened to account for other sources of uncertainty in the number of deaths; see details in the Supplementary Material). The 14 day offset allows for the known delay between the onset of infection and death, taking into consideration the delay between the onset of infection and the development of detectable antibodies [45, 97].

3 The data

3.1 Seroprevalence data

As the COVID-19 pandemic has progressed, a rapidly increasing number of SARS-CoV-2 seroprevalence studies have been conducted worldwide [3]. However, many of these studies have produced biased estimates or are otherwise unreliable due to a variety of different issues with study design, and/or with data collection, and/or with inappropriate statistical analysis [12]. We seek to restrict our analysis to high quality studies which used probability-based sampling methods. Such studies are less likely to suffer from substantial biases [81]. Based on the reviews of Chen et al. [20] and of Arora et al. [3], we compiled two separate sets of studies for analysis (these are listed in Tables 1
and 2, respectively). With the understanding that the results of an evidence synthesis may be largely driven by which studies are included/excluded, we will use these two separate sets to conduct two separate analyses. Note that the data collected for both analyses are relevant to the time before the widespread availability of COVID-19 vaccinations.

Figure 1: Flowchart of seroprevalence studies considered for Chen et al. -based analysis (based on the review of Chen et al. [20]).

Chen et al. [20] reviewed the literature for articles published between December 1, 2019, and December 22, 2020, and identified more than 400 unique seroprevalence studies. For each of these, study quality was established using a scoring system developed on the basis of a seroepidemiological protocol from the Consortium for the Standardization of Influenza Seroepidemiology [31]. In total, Chen et al. [20] identified 38 articles which considered a sample based on a “general population” and which obtained a study quality grade of A or B (see full list in Table S8 of Chen et al. [20]). We consider these 38 articles as a starting point for inclusion for our analysis. After excluding those studies which are duplicates (n=2), those that used a “convenience” or “non-probability” based sampling method (according to the classification of Arora et al. [3]) (n=8), a study no longer considered accurate based on new information about the accuracy of the antibody test used (n=1), a study that has a very narrowly defined target population (n=1), studies for which relevant death data could not be found (n=5), and studies which did not provide a 95% uncertainty interval (n=2), we were left with a set of $K=19$ studies for analysis; see Figure [1] and details in the Supplementary Material.
Arora et al. [3] conducted the Serotracker “living systematic review” of COVID-19 seroprevalence studies whereby the results of the review are continuously updated on serotracker.com/data. For each study reviewed, the risk of bias was evaluated based on an assessment using the Joanna Briggs Institute Critical Appraisal Guidelines for Prevalence studies [12, 55]. For analysis, we consider the 45 studies listed on serotracker.com/data (as of June 5, 2021), that are categorized as having a “low risk of bias” and are categorized as targeting “household and community samples.” After excluding those studies which are duplicates (n=3), one study that used a “convenience” or “non-probability” based sampling method (according to the classification of Arora et al. [3]) (n=1), those studies no longer considered accurate based on new information about the accuracy of the antibody test used (n=2), those that have very narrowly defined target populations (n=2), those for which relevant death data could not be found (n=8), and those which did not provide a 95% uncertainty interval for the estimated prevalence (n=1), we are left with a set of $K = 28$ studies for analysis; see Figure 2 and details in the Supplementary Material.

![Figure 2: Flowchart of seroprevalence studies considered for Serotracker-based analysis (based on the Serotracker review Arora et al. [3]).](https://www.cambridge.org/core/terms.https://doi.org/10.1017/S0950268821002405)
and assumptions), we selected the lowest value amongst the different lower bounds and the highest value amongst the different upper bounds. These numbers are recorded in Tables 1 and 2 under IR interval. Based on these numbers, we calculated effective data values for the number of tests \((T_k)\) and the number of confirmed cases \((CC_k)\) which are listed in Tables 3 and 4 alongside population numbers \((P_k)\) and numbers corresponding to the proportion of the population over 65 years old \((65yo_k)\) and the GDP (PPP) per capita \((GDP_k)\); see Supplemental Material for details and data sources.

### 3.2 Mortality data

Mortality data was obtained from various sources (e.g., academic, government, health authority); see details in Supplementary Material. If a seroprevalence study referenced a specific source for mortality data, we used the referenced source for our numbers whenever possible. If no source was referenced or suggested, we considered publicly available data sources.

For many populations, there were concerns that cause of death information may be very inaccurate and lead to biased COVID-19 mortality statistics. To overcome this issue, many have suggested looking to “excess deaths” by comparing aggregate data for all-cause deaths from the time during the pandemic to the years prior \([42]\). For populations with a large discrepancy between the “official” number of deaths attributed to COVID-19 and the number of excess deaths –as suggested, when possible, by a large undercount ratio (UCR) derived by Karlinsky and Kobak \([37]\)– we used the “official” number of deaths attributed to COVID-19 for the lower bound of the \(D_k\) interval and used numbers based on excess deaths for the upper bound of the \(D_k\) interval.

India, Pakistan, Palestine, Ethiopia, and China are the only countries represented in the studies that we assessed for data availability that were not included in Karlinsky and Kobak \([37]\)’s analysis. There was evidence of substantial under-reporting of COVID-19 deaths in India \([6, 68]\) while little could be gathered about the reliability of official mortality data for Pakistan, Palestine, Ethiopia, and China (but do see \([17, 89]\)). As such, we excluded the Qutob et al. \([70]\) (“Palestinian population residing in the West Bank”) and He et al. \([29]\) (“Wuhan, China”) studies from the Serotracker-based analysis and excluded the Alemu et al. \([1]\) (“Addis Ababa, Ethiopia”) and the Nisar et al. \([8]\) (“Two neighborhoods of Karachi, Pakistan”) studies from the Chen et al.-based analysis. For India, Mukherjee et al. \([54]\) and Purkayastha et al. \([69]\) estimate UCRs for the entire country as well as for each individual Indian state and union territory. We used these UCRs to adjust the death numbers for seroprevalence studies in India (see Supplementary Material for details).

---

2. Official regional death numbers for Ethiopia have been made available previously (e.g., http://web.archive.org/web/2020*/https://www.covid19.et/covid-19/ and the Twitter account: https://twitter.com/Harun_Asefa/status/12590698328779280; accessed August 4, 2021).
There are two countries represented within our data that were identified by Karlinsky and Kobak [37] as having large discrepancies between the official number of deaths attributed to COVID-19 and the number of excess deaths Iran (with UCR=2.4) and Russia (with UCR=4.5). As such, for the Barchuk et al. [7] (“Saint Petersburg, Russia”) and for the Khalagi et al. [38] (“Iran”), we used numbers based on excess deaths for the upper bound of the $D_k$ interval (see $D_k$ numbers in Tables 3 and 4 and see Supplementary Material for details).

Finally, our target of inference is the IFR for the community-dwelling population and does not apply to people living in long-term care (LTC) facilities (also known as “nursing homes” or, in France as “Établissement d’hébergement pour personnes âgées dépendantes” (EHPAD)). The spread of COVID-19 is substantially different in LTC facilities than in the general population, and residents of LTC facilities are particularly vulnerable to severe illness and death from infection; see Danis et al. [22]. With this in mind, we made adjustments (when appropriate/possible) to the mortality numbers used in our analysis in order to exclude deaths of LTC residents; see Supplementary Material for details. Modeling the spread and mortality of COVID-19 within LTC facilities will require unique approaches and should be considered in a separate analysis; see the recommendations of Pillemer et al. [66].

Table 1: Seroprevalence studies selected for the analysis based on the list compiled by Chen et al. [20] (listed in alphabetical order of authors), with geographic location of sampling, sampling dates, and 95% uncertainty interval for the infection rate (IR interval). Also noted, under “In both analyses”, is whether or not each study is included in the Serotracker-based analysis (i.e., is also in Table 2). Studies with yes* are alternate versions of studies that are included in the Serotracker-based analysis. Note that sampling for all studies took place during mid-2020, before the widespread availability of COVID-19 vaccinations.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Location</th>
<th>Sampling (mm/dd)</th>
<th>IR interval (%)</th>
<th>In both analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barchuk et al. [7]</td>
<td>Saint Petersburg, Russia</td>
<td>05/27 - 06/26</td>
<td>[5.60, 12.90]</td>
<td>yes</td>
</tr>
<tr>
<td>Biggs et al. [11]</td>
<td>DeKalb and Fulton, GA, USA (A)</td>
<td>04/28 - 05/03</td>
<td>[1.40, 4.50]</td>
<td>yes*</td>
</tr>
<tr>
<td>Bruckner et al. [16]</td>
<td>Orange County, CA, USA</td>
<td>07/10 - 08/16</td>
<td>[8.10, 15.50]</td>
<td>yes</td>
</tr>
<tr>
<td>Carrat et al. [18]</td>
<td>Ile-de-France, France</td>
<td>05/04 - 06/14</td>
<td>[8.90, 11.30]</td>
<td>yes</td>
</tr>
<tr>
<td>Mahajan et al. [48]</td>
<td>Connecticut, USA</td>
<td>06/10 - 07/29</td>
<td>[1.70, 6.30]</td>
<td>yes</td>
</tr>
<tr>
<td>Murhekar et al. (A) [56]</td>
<td>India</td>
<td>05/11 - 06/04</td>
<td>[0.34, 1.13]</td>
<td>yes</td>
</tr>
<tr>
<td>Petersen et al. [64]</td>
<td>Copenhagen, Denmark</td>
<td>04/27 - 05/01</td>
<td>[0.10, 1.20]</td>
<td>yes</td>
</tr>
<tr>
<td>Pollan et al. [67]</td>
<td>Spain</td>
<td>04/27 - 05/11</td>
<td>[3.30, 6.60]</td>
<td>yes</td>
</tr>
<tr>
<td>Samore et al. [76]</td>
<td>Four counties in UT, USA</td>
<td>05/04 - 06/30</td>
<td>[0.10, 1.60]</td>
<td>yes</td>
</tr>
<tr>
<td>Santos-Hovener et al. [81]</td>
<td>Regensburg, Germany</td>
<td>05/20 - 06/09</td>
<td>[10.40, 14.00]</td>
<td>yes</td>
</tr>
<tr>
<td>Sharma et al. [79]</td>
<td>Delhi, India</td>
<td>08/01 - 08/07</td>
<td>[27.65, 29.14]</td>
<td>yes</td>
</tr>
<tr>
<td>Snoeck et al. [82]</td>
<td>Luxembourg</td>
<td>04/15 - 05/05</td>
<td>[1.23, 2.77]</td>
<td>yes</td>
</tr>
<tr>
<td>Sood et al. [83]</td>
<td>Los Angeles County, CA, USA</td>
<td>04/10 - 04/14</td>
<td>[2.52, 7.07]</td>
<td>yes*</td>
</tr>
<tr>
<td>Statistics Jersey (A) [84]</td>
<td>Jersey, UK (A)</td>
<td>04/29 - 05/05</td>
<td>[1.80, 4.40]</td>
<td>yes*</td>
</tr>
<tr>
<td>Streeck et al. [86]</td>
<td>Gangelt, Germany</td>
<td>03/31 - 04/06</td>
<td>[12.31, 24.40]</td>
<td>yes</td>
</tr>
<tr>
<td>Stringhini et al. (A) [87]</td>
<td>Geneva, Switzerland (A)</td>
<td>04/06 - 05/09</td>
<td>[8.15, 13.95]</td>
<td>yes*</td>
</tr>
<tr>
<td>Vos et al. [91]</td>
<td>Netherlands</td>
<td>03/31 - 05/11</td>
<td>[2.10, 3.70]</td>
<td>yes</td>
</tr>
<tr>
<td>Ward et al. [93]</td>
<td>England, UK (B)</td>
<td>06/20 - 07/13</td>
<td>[5.78, 6.14]</td>
<td>yes</td>
</tr>
</tbody>
</table>

Table 1: Seroprevalence studies selected for the analysis based on the list compiled by Chen et al. [20] (listed in alphabetical order of authors), with geographic location of sampling, sampling dates, and 95% uncertainty interval for the infection rate (IR interval). Also noted, under “In both analyses”, is whether or not each study is included in the Serotracker-based analysis (i.e., is also in Table 2). Studies with yes* are alternate versions of studies that are included in the Serotracker-based analysis. Note that sampling for all studies took place during mid-2020, before the widespread availability of COVID-19 vaccinations.
Table 2: Seroprevalence studies selected for the analysis based on the list compiled by Serotracker (listed in alphabetical order of authors), with geographic location of sampling, sampling dates, and 95% uncertainty interval for the infection rate (IR interval). Also noted, under “In both analyses”, is whether or not each study is included in the Chen et al.-based analysis (i.e., is also in Table 1). Studies with yes* are alternate versions of studies that are included in the Chen et al.-based analysis. Note that sampling for all studies took place during mid-to-late 2020, before the widespread availability of COVID-19 vaccinations.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Location</th>
<th>Sampling (mm/dd)</th>
<th>IR interval (%)</th>
<th>In both analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alvarez-Antonio et al.</td>
<td>Iquitos, Peru</td>
<td>07/13 - 07/18</td>
<td>[67.00, 73.00]</td>
<td></td>
</tr>
<tr>
<td>Bajema et al.</td>
<td>DeKalb and Fulton, GA, USA (B)</td>
<td>04/28 - 05/03</td>
<td>[1.49, 6.67]</td>
<td>yes*</td>
</tr>
<tr>
<td>Barchuk et al.</td>
<td>Saint Petersburg, Russia</td>
<td>05/27 - 06/26</td>
<td>[5.60, 12.90]</td>
<td>yes</td>
</tr>
<tr>
<td>Bruckner et al.</td>
<td>Orange County, CA, USA</td>
<td>07/10 - 08/16</td>
<td>[8.10, 15.50]</td>
<td>yes</td>
</tr>
<tr>
<td>Chan et al.</td>
<td>Rhode Island, USA</td>
<td>05/05 - 05/22</td>
<td>[1.00, 6.20]</td>
<td></td>
</tr>
<tr>
<td>Gov. of Andorra</td>
<td>Andorra</td>
<td>05/04 - 05/28</td>
<td>[10.50, 11.50]</td>
<td></td>
</tr>
<tr>
<td>Kar et al.</td>
<td>Puducherry District, India</td>
<td>08/11 - 08/16</td>
<td>[3.50, 6.40]</td>
<td></td>
</tr>
<tr>
<td>Khalagi et al.</td>
<td>Iran</td>
<td>08/03 - 10/31</td>
<td>[13.30, 15.20]</td>
<td></td>
</tr>
<tr>
<td>Malani et al.</td>
<td>Tamil Nadu, India</td>
<td>10/19 - 11/30</td>
<td>[30.40, 32.80]</td>
<td></td>
</tr>
<tr>
<td>Melotti et al.</td>
<td>Gardena Valley, Italy</td>
<td>05/26 - 06/08</td>
<td>[25.20, 28.60]</td>
<td></td>
</tr>
<tr>
<td>MoHoI</td>
<td>Israel</td>
<td>06/28 - 09/17</td>
<td>[5.30, 5.60]</td>
<td></td>
</tr>
<tr>
<td>Murhekar et al. (A)</td>
<td>India</td>
<td>05/11 - 06/04</td>
<td>[0.34, 1.13]</td>
<td>yes</td>
</tr>
<tr>
<td>Nawa et al.</td>
<td>Utsunomiya City, Japan</td>
<td>06/14 - 07/05</td>
<td>[0.08, 1.20]</td>
<td></td>
</tr>
<tr>
<td>Pagani et al.</td>
<td>Castiglione d’Adda, Italy</td>
<td>05/18 - 06/07</td>
<td>[17.20, 29.10]</td>
<td></td>
</tr>
<tr>
<td>Petersen et al.</td>
<td>Faroe Islands, Denmark</td>
<td>04/27 - 05/01</td>
<td>[0.10, 0.20]</td>
<td>yes</td>
</tr>
<tr>
<td>Pollan et al.</td>
<td>Spain</td>
<td>04/27 - 05/11</td>
<td>[3.30, 6.60]</td>
<td>yes</td>
</tr>
<tr>
<td>Radon et al.</td>
<td>Munich, Germany</td>
<td>11/02 - 01/31</td>
<td>[22.20, 28.20]</td>
<td></td>
</tr>
<tr>
<td>Reyes-Vega et al.</td>
<td>Lima, Peru</td>
<td>06/28 - 07/09</td>
<td>[22.20, 28.20]</td>
<td></td>
</tr>
<tr>
<td>Richard et al.</td>
<td>Geneva, Switzerland (B)</td>
<td>04/06 - 06/30</td>
<td>[7.80, 8.90]</td>
<td>yes*</td>
</tr>
<tr>
<td>Santos-Hovener et al.</td>
<td>Kupferzell, Germany</td>
<td>05/20 - 06/04</td>
<td>[0.40, 14.00]</td>
<td>yes</td>
</tr>
<tr>
<td>Selvaraju et al.</td>
<td>Chennai, Tamil Nadu, India</td>
<td>07/12 - 08/31</td>
<td>[14.80, 22.60]</td>
<td></td>
</tr>
<tr>
<td>Sharma et al.</td>
<td>Delhi, India</td>
<td>08/11 - 08/60</td>
<td>[27.65, 29.14]</td>
<td>yes</td>
</tr>
<tr>
<td>Statistics Jersey (B)</td>
<td>Jersey, UK (B)</td>
<td>02/21 - 06/27</td>
<td>[2.80, 5.20]</td>
<td>yes*</td>
</tr>
<tr>
<td>Streeck et al.</td>
<td>Gagel, Germany</td>
<td>09/1 - 04/06</td>
<td>[12.31, 24.40]</td>
<td>yes</td>
</tr>
<tr>
<td>Vos et al.</td>
<td>Netherlands</td>
<td>03/1 - 05/11</td>
<td>[2.10, 3.70]</td>
<td>yes</td>
</tr>
<tr>
<td>Ward et al.</td>
<td>England, UK (B)</td>
<td>03/20 - 07/13</td>
<td>[5.78, 6.14]</td>
<td></td>
</tr>
<tr>
<td>Warszawski et al.</td>
<td>Metropolitan France</td>
<td>05/02 - 06/02</td>
<td>[3.90, 5.00]</td>
<td></td>
</tr>
<tr>
<td>Yoshiyama et al.</td>
<td>Tokyo, Japan</td>
<td>06/01 - 06/07</td>
<td>[0.01, 0.37]</td>
<td></td>
</tr>
</tbody>
</table>

4 Results

The model as described in Section 2, was fit to the two datasets as described in Section 3. We fit the model using JAGS (just another Gibbs sampler) [39], with 5 independent chains, each with 2 million draws (20% burn-in, thinning of 100); see Supplementary Material for details and JAGS code.

We report posterior median estimates and 95% highest probability density (HPD) credible intervals (CrI). Figure 3 (for the Chen et al.-based analysis) and Figure 4 (for the Serotracker-based analysis) plot the point estimates and credible intervals obtained for $IFR_k$, for $k$ in 1, . . . , $K$, respectively; see Figures S1 and S2 for $IR_k$ are in the Supplementary Material. In these figures the seroprevalence studies are listed in order of their “fitted” IFR values (the posterior median of $g^{-1}(\theta_0 + \theta_1 Z_{1k} + \theta_2 Z_{2k})$, $12$
Table 3: The Chen et al. based dataset required for the Bayesian evidence synthesis model.

<table>
<thead>
<tr>
<th>Location</th>
<th>$P_k$</th>
<th>$D_k^{\text{lower}}$</th>
<th>$D_k^{\text{upper}}$</th>
<th>$T_k$</th>
<th>$CC_k$</th>
<th>$65yo_k$ (%)</th>
<th>$GDP_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saint Petersburg, Russia</td>
<td>5351935</td>
<td>1596</td>
<td>5128</td>
<td>233</td>
<td>21</td>
<td>18</td>
<td>30144</td>
</tr>
<tr>
<td>DeKalb and Fulton, GA, USA (A)</td>
<td>1823234</td>
<td>122</td>
<td>136</td>
<td>419</td>
<td>11</td>
<td>12</td>
<td>58933</td>
</tr>
<tr>
<td>Orange County, CA, USA</td>
<td>3010232</td>
<td>512</td>
<td>839</td>
<td>285</td>
<td>33</td>
<td>15</td>
<td>79287</td>
</tr>
<tr>
<td>Ile-de-France, France</td>
<td>12213447</td>
<td>6816</td>
<td>7427</td>
<td>2414</td>
<td>243</td>
<td>14</td>
<td>82574</td>
</tr>
<tr>
<td>Connecticut, USA</td>
<td>2837877</td>
<td>1136</td>
<td>1179</td>
<td>251</td>
<td>9</td>
<td>22</td>
<td>80729</td>
</tr>
<tr>
<td>India</td>
<td>136641775</td>
<td>4172</td>
<td>45766</td>
<td>1632</td>
<td>11</td>
<td>6</td>
<td>43310</td>
</tr>
<tr>
<td>England, UK (A)</td>
<td>56550000</td>
<td>25975</td>
<td>30709</td>
<td>3100</td>
<td>193</td>
<td>18</td>
<td>43310</td>
</tr>
<tr>
<td>Faroe Islands, Denmark</td>
<td>52154</td>
<td>0</td>
<td>0</td>
<td>592</td>
<td>3</td>
<td>17</td>
<td>60475</td>
</tr>
<tr>
<td>Spain</td>
<td>47351567</td>
<td>17710</td>
<td>17767</td>
<td>643</td>
<td>31</td>
<td>20</td>
<td>42756</td>
</tr>
<tr>
<td>Four counties in UT, USA</td>
<td>2194298</td>
<td>40</td>
<td>97</td>
<td>393</td>
<td>2</td>
<td>10</td>
<td>50603</td>
</tr>
<tr>
<td>Kupferzell, Germany</td>
<td>6247</td>
<td>1</td>
<td>3</td>
<td>1263</td>
<td>153</td>
<td>16</td>
<td>63885</td>
</tr>
<tr>
<td>Delhi, India</td>
<td>19800000</td>
<td>4188</td>
<td>26901</td>
<td>13966</td>
<td>5</td>
<td>12867</td>
<td></td>
</tr>
<tr>
<td>Luxembourg</td>
<td>632275</td>
<td>47</td>
<td>58</td>
<td>1214</td>
<td>23</td>
<td>14</td>
<td>422166</td>
</tr>
<tr>
<td>Los Angeles County, CA, USA</td>
<td>10039107</td>
<td>529</td>
<td>606</td>
<td>316</td>
<td>14</td>
<td>14</td>
<td>92874</td>
</tr>
<tr>
<td>Jersey, UK (A)</td>
<td>107800</td>
<td>17</td>
<td>18</td>
<td>648</td>
<td>19</td>
<td>17</td>
<td>48365</td>
</tr>
<tr>
<td>Gangelt, Germany</td>
<td>12597</td>
<td>8</td>
<td>8</td>
<td>153</td>
<td>27</td>
<td>18</td>
<td>53751</td>
</tr>
<tr>
<td>Geneva, Switzerland (A)</td>
<td>504128</td>
<td>122</td>
<td>155</td>
<td>442</td>
<td>16</td>
<td>16</td>
<td>68964</td>
</tr>
<tr>
<td>Netherlands</td>
<td>17344874</td>
<td>2621</td>
<td>4186</td>
<td>1652</td>
<td>47</td>
<td>20</td>
<td>59685</td>
</tr>
<tr>
<td>England, UK (B)</td>
<td>56550000</td>
<td>29526</td>
<td>30038</td>
<td>10835</td>
<td>14</td>
<td>18</td>
<td>43310</td>
</tr>
</tbody>
</table>

for $k$ in 1, ..., $K$, marked on the plot by the × symbols. Results obtained for the other model parameters are listed in Table 5.

In general, the Chen et al.-based analysis and the Serotracker-based analysis provide mostly similar results. Notably, the Serotracker-based analysis considers a much more geographically diverse set of seroprevalence studies and several studies that appear to be prominent outliers (e.g., “Tamil Nadu”, “Castiglione d’Adda, Italy” and “Utsunomiya City, Japan”), see Figure 4. These outliers could be due to infection rates in these populations being markedly different for the elderly relative to the general population.

With regards to heterogeneity, fitting the model without any covariates, one obtains $\hat{\tau} = 0.62$ (Chen et al.-based analysis) and $\hat{\tau} = 0.90$ (Serotracker-based analysis). This suggests that the two covariates, $65yo_k$ and $GDP_k$, account for approximately 45% ($= (0.62^2 - 0.46^2) / (0.62^2)$; Chen et al.-based analysis) and 13% ($= (0.90^2 - 0.84^2) / (0.90^2)$; Serotracker-based analysis) of the heterogeneity in the IFR.

3Malani et al. [49] (“Tamil Nadu”) note that: “Seroprevalence among the elderly (70+: 25.8%) is significantly lower than among the working age populations (age 40-49: 31.6%; $p < 0.001$) or the young (18-29: 30.7%; $p < 0.001$).” Pagani et al.[61] (“Castiglione d’Adda, Italy”) also report that the elderly are more likely to have been infected (“strong association observed between IgG seroprevalence and age”). On the other hand, note that among the 181 participants in the Nawa et al. [57] study (“Utsunomiya City, Japan”) who were aged 65 years or older, none were positive. This suggests that infection in the elderly may be lower than in the general population. However, since the Nawa et al. [57] found only 3 positive cases out of a total of 742 individuals tested, inference on this is limited.

4For reference, Levin et al. [43] conclude that 87% of the heterogeneity in the IFR (of advanced economies) can be explained by variations in age composition and age-specific prevalence of COVID-19. However, note that the linear regression analysis used to obtain this 87% result is done without an intercept term (see Levin et al. [43] - Figure 6). A linear regression with intercept results in a value of 43%. The intraclass correlation coefficient (ICC) 90 between Levin et al. [43]’s predicted IFRs and the observed IFRs is 0.65 (after removing one outlier, “Portugal”), suggesting that the extent of agreement is reasonably high but nowhere near perfect.
Our estimates of $\hat{\theta}_1 = 0.61$ (Chen et al.-based analysis) and $\hat{\theta}_1 = 0.42$ (Serotracker-based analysis) suggest that older populations are more likely to have higher IFRs. This is as expected since age is known to be a very important risk factor [96, 100]. Our estimate of $\hat{\theta}_2 = -0.28$ (Chen et al.-based analysis) and $\hat{\theta}_2 = -0.11$ (Serotracker-based analysis) suggest that wealthier populations may be more likely to have lower IFRs. However, the wide credible intervals obtained for the $\theta_2$ parameter (in both analyses) suggest a much less definitive conclusion. There are several reasons which might explain this result. As with any observational data analysis, the estimate of $\theta_2$ may suffer from bias due to unobserved confounding and statistical power may be compromised by the presence of outliers and insufficient heterogeneity in the GDP per capita metric across the different populations included in our analyses.

We can infer (by determining the posterior median of $g^{-1}(\theta_0 + \theta_1 z_{1*} + \theta_2 z_{2*})$, for selected values of $z_{1*}$ and $z_{2*}$) the typical IFR amongst populations (be they included in our study or not) having a given proportion of the populace aged over 65 and a given GDP per capita. Thus we calculate posterior point and interval estimates corresponding to age and wealth values that match the population of the entire world (World), the United States (USA), and the European Union (EU) (as listed by the

<table>
<thead>
<tr>
<th>Location</th>
<th>$P_k$</th>
<th>$D_{k_{lower}}$</th>
<th>$D_{k_{upper}}$</th>
<th>$T_k$</th>
<th>$CC_k$</th>
<th>$65yo_k$ (%)</th>
<th>$GDP_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iquitos, Peru</td>
<td>467000</td>
<td>1581</td>
<td>1942</td>
<td>895</td>
<td>626</td>
<td>4</td>
<td>5896</td>
</tr>
<tr>
<td>DeKalb and Fulton, GA, USA (B)</td>
<td>1823234</td>
<td>122</td>
<td>136</td>
<td>196</td>
<td>7</td>
<td>12</td>
<td>58933</td>
</tr>
<tr>
<td>Saint Petersburg, Russia</td>
<td>5351935</td>
<td>1596</td>
<td>5128</td>
<td>233</td>
<td>21</td>
<td>18</td>
<td>30144</td>
</tr>
<tr>
<td>Orange County, CA, USA</td>
<td>3010232</td>
<td>512</td>
<td>839</td>
<td>285</td>
<td>33</td>
<td>15</td>
<td>79287</td>
</tr>
<tr>
<td>Rhode Island, USA</td>
<td>1059361</td>
<td>147</td>
<td>189</td>
<td>163</td>
<td>5</td>
<td>18</td>
<td>58416</td>
</tr>
<tr>
<td>Andorra</td>
<td>77543</td>
<td>35</td>
<td>35</td>
<td>11236</td>
<td>1235</td>
<td>17</td>
<td>49900</td>
</tr>
<tr>
<td>Puducherry District, India</td>
<td>1250000</td>
<td>145</td>
<td>393</td>
<td>840</td>
<td>41</td>
<td>6</td>
<td>9152</td>
</tr>
<tr>
<td>Iran</td>
<td>82913906</td>
<td>19804</td>
<td>98071</td>
<td>5197</td>
<td>740</td>
<td>6</td>
<td>12937</td>
</tr>
<tr>
<td>Tamil Nadu, India</td>
<td>83697770</td>
<td>11183</td>
<td>29773</td>
<td>5765</td>
<td>1821</td>
<td>7</td>
<td>7191</td>
</tr>
<tr>
<td>Gardena Valley, Italy</td>
<td>10700</td>
<td>15</td>
<td>15</td>
<td>2612</td>
<td>702</td>
<td>20</td>
<td>57833</td>
</tr>
<tr>
<td>Israel</td>
<td>9216900</td>
<td>187</td>
<td>995</td>
<td>10578</td>
<td>577</td>
<td>11</td>
<td>4747</td>
</tr>
<tr>
<td>India</td>
<td>1366417750</td>
<td>4172</td>
<td>45766</td>
<td>1632</td>
<td>11</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>Utsunomiya City, Japan</td>
<td>517527</td>
<td>0</td>
<td>0</td>
<td>228</td>
<td>2</td>
<td>26</td>
<td>12931</td>
</tr>
<tr>
<td>Castiglione d’Adda, Italy</td>
<td>4605</td>
<td>39</td>
<td>40</td>
<td>190</td>
<td>43</td>
<td>26</td>
<td>39291</td>
</tr>
<tr>
<td>Faroe Islands, Denmark</td>
<td>52154</td>
<td>0</td>
<td>0</td>
<td>592</td>
<td>3</td>
<td>17</td>
<td>60421</td>
</tr>
<tr>
<td>Spain</td>
<td>47351567</td>
<td>17710</td>
<td>17767</td>
<td>643</td>
<td>35</td>
<td>20</td>
<td>42362</td>
</tr>
<tr>
<td>Munich, Germany</td>
<td>1563090</td>
<td>219</td>
<td>780</td>
<td>2692</td>
<td>14</td>
<td>17</td>
<td>65345</td>
</tr>
<tr>
<td>Lima, Peru</td>
<td>10804609</td>
<td>21109</td>
<td>28846</td>
<td>893</td>
<td>226</td>
<td>9</td>
<td>19313</td>
</tr>
<tr>
<td>Geneva, Switzerland (B)</td>
<td>504128</td>
<td>122</td>
<td>156</td>
<td>2697</td>
<td>26</td>
<td>16</td>
<td>68964</td>
</tr>
<tr>
<td>Kupferzell, Germany</td>
<td>6247</td>
<td>1</td>
<td>3</td>
<td>1262</td>
<td>53</td>
<td>16</td>
<td>63885</td>
</tr>
<tr>
<td>Chennai, Tamil Nadu, India</td>
<td>10900000</td>
<td>1295</td>
<td>59608</td>
<td>886</td>
<td>70</td>
<td>7</td>
<td>7191</td>
</tr>
<tr>
<td>Delhi, India</td>
<td>19800000</td>
<td>4188</td>
<td>26903</td>
<td>13960</td>
<td>3965</td>
<td>4</td>
<td>12817</td>
</tr>
<tr>
<td>Jersey, UK (B)</td>
<td>107800</td>
<td>20</td>
<td>30</td>
<td>998</td>
<td>39</td>
<td>17</td>
<td>48365</td>
</tr>
<tr>
<td>Gargelt, Germany</td>
<td>12597</td>
<td>8</td>
<td>15</td>
<td>153</td>
<td>27</td>
<td>18</td>
<td>53751</td>
</tr>
<tr>
<td>Netherlands</td>
<td>17344874</td>
<td>186</td>
<td>186</td>
<td>1652</td>
<td>47</td>
<td>20</td>
<td>59685</td>
</tr>
<tr>
<td>England, UK (B)</td>
<td>56550000</td>
<td>2920</td>
<td>3368</td>
<td>10635</td>
<td>634</td>
<td>18</td>
<td>43310</td>
</tr>
<tr>
<td>Metropolitan France</td>
<td>64897972</td>
<td>17373</td>
<td>9038</td>
<td>5377</td>
<td>238</td>
<td>21</td>
<td>49551</td>
</tr>
<tr>
<td>Tokyo, Japan</td>
<td>13960238</td>
<td>275</td>
<td>275</td>
<td>1314</td>
<td>2</td>
<td>20</td>
<td>45796</td>
</tr>
</tbody>
</table>

Table 4: The Serotracker based dataset required for the Bayesian evidence synthesis model.
Table 5: Parameter estimates (posterior medians and 95% HPD credible intervals) obtained from the Chen et al.-based analysis and the Serotracker-based analysis.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Chen et al. analysis</th>
<th>Serotracker analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_0$</td>
<td>-5.50, with 95% CrI of (-5.78, -5.22)</td>
<td>-5.56, with 95% CrI of (-5.93, -5.19)</td>
</tr>
<tr>
<td>$\theta_1$</td>
<td>0.61, with 95% CrI of (0.21, 1.00)</td>
<td>0.42, with 95% CrI of (-0.40, 1.16)</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>-0.28 with 95% CrI of (-0.67, 0.14)</td>
<td>-0.11, with 95% CrI of (-0.86, 0.69)</td>
</tr>
<tr>
<td>$\tau$</td>
<td>0.46 with 95% CrI of (0.25, 0.75)</td>
<td>0.84, with 95% CrI of (0.55, 1.23)</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>1.17 with 95% CrI of (0.80, 1.68)</td>
<td>1.53, with 95% CrI of (1.13, 2.05)</td>
</tr>
</tbody>
</table>

World Bank’s World Development Indicators (WDI)); see “World”, “USA”, and “EU” rows in Figures 3 and 4. For 65yo = 9% and GDP = $17,811, the approximate worldwide values, we obtain, from the Chen et al.-based analysis, an across-population average IFR estimate of 0.31%, with a 95% CrI of (0.16%, 0.53%). With the Serotracker-based analysis, we obtain a similar estimate of 0.32%, with a 95% CrI of (0.19%, 0.47%). For 65yo = 16% and GDP = $55,298, the USA values, we obtain across-population average IFR estimates of 0.43%, with a 95% CrI of (0.31%, 0.56%) (Chen et al.-based analysis) and of 0.41%, with a 95% CrI of (0.22%, 0.67%) (Serotracker-based analysis). Finally, for 65yo = 20% and GDP = $47,828, the EU values, we obtain across-population average IFR estimates 0.67%, with a 95% CrI of (0.41%, 0.96%) (Chen et al.-based analysis) and of 0.51%, with a 95% CrI of (0.27%, 0.79%) (Serotracker-based analysis). Note that for the “World” predictions, the Serotracker-based analysis has the more precise estimates, while the Chen et al.-based estimates are more precise for the “USA” predictions. This is likely due to the fact that the Serotracker-based analysis considers several younger and less wealthy populations, whereas the Chen et al.-based analysis considers fewer outliers.

While the infection-rate estimates obtained from the seroprevalence studies should be relatively reliable (due to having satisfied the risk of bias assessments of either Chen et al. [20] or Arora et al. [8]), the mortality data we collected may be less reliable depending on the target population and source. The data which were not obtained from official and reliable sources may be particularly suspect. With this in mind, as a sensitivity analysis, we repeated both analyses with these data excluded; see results in Figures S3 and S4 in the Supplementary Material. Without the excluded studies, we are unable to provide a reasonable “World” estimate (see the extremely wide credible intervals). However, the “USA” and “EU” estimates are relatively similar. We also repeated the two analyses using a different set of priors to verify that our results were not overly sensitive to our particular choice of priors. The results of this alternative analysis are very similar to the results of our original analyses; see Figures S5 and S6 in the Supplementary Material.

Our estimates are somewhat similar to those obtained in other analyses. Brazeau et al. [13], using data from 10 representative seroprevalence studies (identified after screening 175 studies), infer “the overall IFR in a typical low-income country, with a population structure skewed towards younger...
individuals, to be 0.23% (0.14%-0.42% 95% prediction interval range).” For a “typical high income country, with a greater concentration of elderly individuals,” Brazeau et al. [13] obtain an estimate of 1.15% (95% prediction interval of 0.78%-1.79%). Ioannidis [33], using data from seroprevalence studies with sample sizes greater than 500 (and including deaths of LTC residents), obtains a “median infection fatality rate across all 51 locations” of 0.27% and (and of 0.23% following an ad-hoc correction to take into account “that only one or two types of antibodies” may have been tested in some seroprevalence studies). Levin et al. [43], who restricted their analysis to populations in “advanced economies” (and included deaths of LTC residents) provide age-group specific estimates and country-specific estimates. For instance, for the 45–54 year old age group, Levin et al. [43] estimate the IFR to be 0.23% (95% CI of 0.20%–0.26%), and for the 55–64 year old age group, 0.75% (95% CI of 0.66%–0.87%). For Spain, Levin et al. [43] estimate an IFR of 1.90%. For comparison, we estimate the IFR for the community-dwelling population (i.e., excluding deaths of LTC residents) of Spain to be 0.77% (in both analyses). This is similar to the 0.83% estimate obtained by Pastor-Barriuso et al. [62] and the 0.75% estimate obtained by Brazeau et al. [13] (both of these excluding deaths of LTC residents).

Specifically with regards to the United States, Sullivan et al. [88] estimate the IFR for adults to be 0.85% (95% CrI of 0.76%-0.97%) based on a US nationwide seroprevalence survey conducted between August and December, 2020 [9]. Pei et al. [63] using a rather complex Bayesian “metapopulation” model conclude that, for the United States during 2020, the IFR likely “decreased from around 1% in March to about 0.25% in December.” For comparison, our “USA” predictions of 0.43% and of 0.41% are based on data obtained mostly between April, 2020 and August, 2020 (see dates in Tables [1] and [2]).

5 Conclusion

Estimation of the IFR can be incredibly challenging due to the fact that it is a ratio of numbers where both the numerator and the denominator are subject to a wide range of biases. Our proposed method seeks to address some of these biases in a straightforward manner. In our analysis, proper handling of the various sources of uncertainty was a primary focus [99].

With regards to the numerator, we considered the number of deaths as interval censored data so as to account for the uncertainty in selecting the most relevant number of deaths. While we consider this an improvement over other methods that use a single fixed number, we acknowledge that the specific choice of a 14 day offset is somewhat arbitrary and that the data for deaths also suffer from other sources of bias. Ioannidis [34] notes that the time between infection and death may vary substantially “and may be shorter in developing countries where fewer people are long-sustained.
Figure 3: Results from the Chen et al.-based analysis: posterior median estimates (black circles) for the \( IFR_k \) variables (for \( k = 1, \ldots, 19 \)) with 95% HPD CrIs. Studies are listed from top to bottom in order of increasing fitted values (these values are indicated by \( \times \)). Also plotted, under the labels “World (65 yo=9%, GDP=17.8k)”, “USA (65 yo=16%, GDP=65.3k)”, “EU (65 yo=20%, GDP=47.8k)”, are the posterior median estimate and 95% HPD CrIs for the typical IFR corresponding to values for the proportion of the population aged 65 years and older of 9% and for GDP per capita of $17,811 (the worldwide values), of 16% and of $65,298 (the USA values), and of 20% and of $47,828 (the EU values).
Figure 4: Results from the Serotracker-based analysis: posterior median estimates (black circles) for the IFR, variables (for $k = 1, \ldots, 28$) with 95% HPD CrIs. Studies are listed from the top to the bottom in order of increasing fitted values (these values are indicated by $\times$). Also plotted, under the labels “World (65 yo=9%, GDP=17.8k)”, “USA (65 yo=16%, GDP=65.3k)”, “EU (65 yo=20%, GDP=47.8k)”, are the posterior median estimate and 95% HPD CrIs for the typical IFR corresponding to values for the proportion of the population aged 65 years and older of 9% and for GDP per capita of $17,811$ (the worldwide values), of 16% and of $65,298$ (the USA values), and of 20% and of $47,828$ (the EU values).
by medical support.” In addition to official numbers, we used mortality data based on “excess deaths” statistics for Russia and Iran, since official mortality statistics appeared to be potentially highly inaccurate. We also used adjusted mortality numbers for India based on the best available information. These adjustments are certainly not perfect and we note that “excess deaths” statistics may also suffer from substantial inaccuracies [53].

With regards to the denominator, we looked to data from “high-quality” seroprevalence studies in an effort to avoid biased estimates. However, these data are also not perfect. Seroprevalence studies are severely limited by the representativeness of the individuals they test. Certain groups of individuals who may have very high infection rates are unlikely to be tested in a seroprevalence study (e.g., homeless people). On the other hand, those individuals who have reason to believe they may have been infected, may be more likely to volunteer to participate in a seroprevalence study [81]. It is also likely that seroreversion (loss of detectable antibodies over time) may lead to a seroprevalence study underestimating the true number of infections if the time between the main outbreak and the subsequent antibody testing is substantial [75]. Notably, Axfors and Ioannidis [4] employ a “X-fold”-based correction factor to adjust seroprevalence estimates for this type of bias. The need to improve the quality and reporting of seroprevalence studies cannot be overemphasized.

A major limitation of evidence synthesis is often summarized by the expression “garbage in, garbage out” [23], meaning that if one includes biased studies in one’s analysis, the analysis results will themselves be biased [80]. In our two analyses, we only included data from 19 and 28 out of potentially hundreds of seroprevalence studies due primarily to the fact that so few studies were considered reliable and at low risk of bias. Excluding low-quality/biased studies from our analysis was necessary, at least to a certain degree, in order to obtain valid estimates. However, as a consequence of our strict exclusion criteria, much of the world’s population is severely under-represented in our data. In related work, Levin et al. [44] review the available literature and “informally assess studies for risk of bias” in an attempt to estimate the COVID-19 IFR specifically for developing countries. If the quality of studies were to be correlated with unmeasured factors that impact the IFR, excluding studies based on their perceived quality could lead to unmeasured confounding at a meta-analytic level [35]. Novel methods which allow evidence syntheses to appropriately incorporate biased data are urgently needed. Recently, Campbell et al. [17] proposed a partially identified model to combine seroprevalence study data with biased data from official statistics.

Outside of biased data, perhaps the foremost challenge in evidence synthesis using observational data is that necessarily one is forced to make an array of design choices around inclusion/exclusion

---

6Bobrovitz et al. [12] conclude that a majority of COVID-19 seroprevalence studies are “at high risk of bias […], often for not statistically correcting for demographics or for test sensitivity and specificity, using non-probability sampling methods, and using non-representative sample frames.” Shook-Ša et al. [61] note that “it is very difficult to estimate the level of bias introduced by convenience sampling.”
criteria, statistical modeling approaches, and prior specifications [26]. With the two separate
analyses and the various additional sensitivity analyses, we were quite encouraged by the stability
of our results to perturbations of these inputs.

Reducing the uncertainty around the severity of COVID-19 was of great importance to policy makers
and the public during the early stages of the pandemic [24, 32, 46] and immense efforts have been
made in the collection and analysis of data (e.g., Williamson et al. [96]). And yet, even after more
than a year, there is still a large amount of uncertainty and unexplained heterogeneity surrounding
the COVID-19 IFR, particularly with respect to populations in less affluent countries. While a
certain amount of heterogeneity is to be expected [30], identifying factors associated with higher
IFRs is the ultimate goal and investigating potential variables that can account for the observed
heterogeneity may lead to important insights [10, 35].

We prioritized simplicity in our modeling so as to promote transparency in our findings, and to
facilitate adaptations to similar, but not identical, data structures. While “simple” is a relative term,
note that the entire dataset used for our analyses fits on a single page (in Tables 3 and 4) and that the
entire JAGS MCMC code fits on less than a single page (see Supplementary Material). One model
extension that could be pursued would involve age stratification of IFR.

Including age-stratification in the model could represent a substantial improvement given that
infection in some populations is far from homogeneous (e.g., about 95% of Singapore’s COVID-
19 infections were among young migrant workers (as of September, 2020), which explains the
incredibly low case fatality rate [25]). If a factor, such as age, impacts both the risk of infection
and the risk of death given infection, then estimating the IFR as we have done in our analysis
could be subject to confounding [95]. Age-group specific seroprevalence/mortality data is available
for certain geographic areas [74] (although not always consistently reported) and such data could
inform an extended version of our model, thereby offering an alternative to the approach described
by Levin et al. [43] for estimating age-group specific IFRs.

Finally, we must emphasize that the IFR is a moving target. As the pandemic changes, so to does
the IFR. Our estimates are based on data from 2020, most of which were obtained more than a year
ago. It is likely that, with continual viral mutation of SARS-CoV-2, advances in treatment, and the
availability of vaccines, the current IFR in many places is now markedly different than it was earlier
in 2020 [25, 65, 92].
Key Messages

- The COVID-19 IFR is estimated to be about 0.32% for a typical community-dwelling population where 9% of individuals are over 65 years old and where the GDP per capita is $17.8k (the approximate worldwide averages). For a typical community-dwelling population with the age and wealth of the United States we estimate the IFR to be approximately 0.42%.

- Any estimation of the COVID-19 IFR should take into account the various uncertainties and potential biases in both the mortality data and the seroprevalence data.

- Bayesian methods with interval censoring are well suited for complex evidence synthesis tasks such as estimating the COVID-19 IFR.

Funding - This work was supported by the European Union's Horizon 2020 research and innovation programme under ReCoDID grant agreement No 25746 and by the Canadian Institutes of Health Research, Institute of Genetics (CIHR-IG) under Grant Agreement No 01886-000.

Conflicts of interest/Competing interests - None declared.

Data Availability Statement - Data and code used for the analysis is available in the Supplementary Material and at the OSF project “Inferring the COVID-19 infection fatality rate in the community-dwelling population: a simple Bayesian evidence synthesis of seroprevalence study data and imprecise mortality data” (DOI 10.17605/OSF.IO/34SQ5); see osf.io/34sq5.

References


[26] Andrew Gelman and Eric Loken. The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. *Department of Statistics, Columbia University*, 348, 2013.


[40] Matthias Kämmerer, Philipp Berens, and Jakob Macke. A simple Bayesian analysis of the infection fatality rate in Gangelt, and an uncertainty aware extrapolation to infection-counts in Germany. [https://matthias-k.github.io/BayesianHeinsberg.html](https://matthias-k.github.io/BayesianHeinsberg.html), 2020.


[46] Marc Lipsitch. First Opinion: We know enough now to act decisively against COVID-19. Social distancing is a good place to start. *STAT*, [https://tinyurl.com/ya4g9m](https://tinyurl.com/ya4g9m), 2020.


