Inferring the COVID-19 IFR with a simple Bayesian evidence synthesis of seroprevalence study data and imprecise mortality data

Harlan Campbell*[1] and Paul Gustafson[1]

[1]Department of Statistics, University of British Columbia

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. We estimate the COVID-19 IFR to be 0.40% (95% credible interval of (0.20%, 0.63%)) for a typical population where the proportion of those aged over 65 years old is 9% and where the GDP (at purchasing-power parity (ppp)) per capita is $17.8K (the approximate worldwide values). Our results suggest that, as one might expect, lower IFRs are associated with younger and with wealthier populations.

* Please contact: harlan.campbell@stat.ubc.ca

> 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 (Clapham et al., 2020). In particular, Ioannidis (2021*a*), Levin et al. (2020), Brazeau et al. (2020), and O'Driscoll et al. (2020) 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. (2020) and O'Driscoll et al. (2020) are done using rather complex Bayesian models which rely on numerous detailed assumptions. For instance, Brazeau et al. (2020) 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. (2020) 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" since "older individuals have fewer social contacts and are more likely to be isolated through shielding programmes." 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 (2021*a*) and Levin et al. (2020) are decidedly more simple. For each seroprevalence study under consideration, Ioannidis (2021*a*) counts the number of deaths 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 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 (2021*a*) 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 (2021*a*)'s estimates.

The analysis of Levin et al. (2020) is based on a standard frequentist random-effects meta-analysis model. For each age-group and seroprevalence study under consideration, Levin et al. (2020) calculate a 95% confidence interval (CI) for a study-specific IFR by counting the 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 fails to take into account certain important sources of uncertainty (to be discussed in Section 2).

The analysis method we propose seeks to be simple enough for researchers to easily understand and use, while at the same time properly account for important sources of uncertainty inherent in both the seroprevalence data and the mortality data. Simple Bayesian models have been used previously for evidence synthesis of seroprevalence data for other infectious diseases (e.g., Brody-Moore (2019)).

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 (Ioannidis, 2021*b*). Reviewing and evaluating the merits of the hundreds of available seroprevalence studies also involves a tremendous

amount of review work and time. Fortunately, Chen et al. (2021) have done a thorough review and assessment of potential studies to ascertain study quality (i.e., risk of bias) and eligibility for meta-analysis. We will work from the list of "grade A" and "grade B" studies compiled by Chen et al. (2021) excluding those studies with "convenience" or "non-probability" based sampling methods (as determined by Arora et al. (2021)). 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 with some final thoughts.

2 The Bayesian model for evidence synthesis

Suppose we have data from K 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 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 $k = 1, \ldots, K$:

$$CC_k \sim Binom(T_k, C_k/P_k), \tag{1}$$

$$C_k \sim Binom(P_k, IR_k), \quad \text{and} \tag{2}$$

$$D_k|C_k \sim Binom(C_k, IFR_k). \tag{3}$$

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. (2020) (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% CI for $C = [1{,}551, 2{,}389]$). Doing so, they obtain a relatively narrow 95% CI for the IFR: $[0.29\%, 0.45\%]$ ($= [7/1{,}551, 7/2{,}389]$). 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. (2020) also fail to take into account the $D|C$ binomial distribution when estimating study-specific IFRs. This leads Levin et al. (2020) to obtain overly precise IFR estimates for their meta-analysis. The result of this is a very large I^2 of 97.0 which gives the false impression that the differences in observed IFRs are almost entirely due to "unexplained variations across studies."

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

$$\mathrm{g}(IFR_k) \sim \mathcal{N}(\theta_0 + \theta_1 Z_{1k} + \theta_2 Z_{2k}, \tau^2), \quad \text{and} \tag{4}$$

$$\mathrm{g}(IR_k) \sim \mathcal{N}(\beta, \sigma^2), \tag{5}$$

where θ_0 represents the mean g(infection fatality rate), τ^2 represents between group infection fatality rate heterogeneity, β represents the mean g(infection rate), σ^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 θ_1 and θ_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 these as the centered and scaled logarithm of (1) the proportion of the population aged over 65 years ($65yo_k$), and of (2) the GDP (at

purchasing power parity (ppp)) per capita (GDP_k), 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. (2006) and Kümmerer et al. (2020)) are set accordingly: $\mathrm{g}^{-1}(\theta) \sim Beta(0.3, 30)$; $\mathrm{g}^{-1}(\beta) \sim Beta(1, 30)$; $\theta_1 \sim \mathcal{N}(0, 10)$; $\theta_2 \sim \mathcal{N}(0, 10)$; $\sigma \sim$ half-$\mathcal{N}(0, 10)$ and $\tau \sim$ 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 (Berger, 2013, Lambert et al., 2005). The priors described here represent a scenario where there is little to no *a priori* knowledge about the model parameters. Inference would no doubt be improved should appropriate informative priors be specified. In Appendix 6.3, we show results from the model fit with an alternative set of priors as a sensitivity analysis.

2.1 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 (2021*a*)) 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 Kümmerer et al. (2020) who assume that the IR follows a beta distribution with parameters chosen to match the 95% CI published in Streeck et al. (2020).) Table 1 lists the 95% uncertainty intervals obtained from each of the seroprevalence studies in our analysis and Table 2 lists the corresponding values for T_k and CC_k.

It must be noted that, as Ioannidis (2021*a*) 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 (2021*a*) 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 studies (this according to Chen et al. (2021)), 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 (Bobrovitz et al., 2020, Brownstein and Chen, 2021).

2.2 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. (2020)), but can also be several weeks or months for others (e.g., 135 days for Ward et al. (2020)). Table 1 lists the sampling window start and end dates for each of the studies in our analysis.

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 (2021*a*) 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 (2020) 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 in which the number of deaths was rapidly accelerating. Levin et al. (2020), 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. Table 2 lists 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. 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; see Wu et al. (2020) and Linton et al. (2020).[1]

3 The Data

3.1 Seroprevalence data

As the COVID-19 pandemic has progressed, a rapidly increasing number of seroprevalence surveys for antibodies to SARS-CoV-2 have been conducted worldwide (Arora et al., 2021). 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. Bobrovitz et al. (2020) 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." We seek to restrict our analysis to high quality studies which used probability-based sampling methods. Such studies are less likely

[1] Note that, for the Streeck et al. (2020) study, which used both antibody-based and PCR-based testing, the interval listed is based on the numbers noted by Streeck et al. (2020) as well as numbers noted in an investigative report on the Gangelt outbreak (see `https://medwatch.de/2020/11/26/die-ungezaehlten-todesfaelle-aus-gangelt/`; accessed on May 20, 2021).

Authors	Location	Sampling (mm/dd)	IR interval (%)
Barchuk et al.	Saint Petersburg, Russia	05/27 - 06/26	[5.60, 12.90]
Biggs et al.	Two counties in GA, USA	04/28 - 05/03	[1.40, 4.50]
Bruckner et al.	Orange county, CA, USA	07/10 - 08/16	[8.10, 15.50]
Carrat et al.	Ile-de-France, France	05/04 - 06/14	[8.90, 11.30]
Government of Jersey	Jersey, UK	04/29 - 05/05	[1.80, 4.40]
Murhekar et al. (1)	India	05/11 - 06/04	[0.34, 1.13]
Office of National Stat	England, UK (1)	04/26 - 09/08	[5.40, 7.10]
Petersen et al.	Faroe Islands, Denmark	04/27 - 05/01	[0.10, 1.20]
Pollán et al.	Spain	04/27 - 05/11	[3.30, 6.60]
Samore et al.	Four counties in UT, USA	05/04 - 06/30	[0.10, 1.60]
Santos-Hovener et al.	Kupferzell, Germany	05/20 - 06/09	[10.40, 14.00]
Sharma et al.	Delhi, India	08/01 - 08/07	[27.65, 29.14]
Snoeck et al.	Luxembourg	04/15 - 05/05	[1.23, 2.77]
Sood et al.	Los Angeles county, CA, USA	04/10 - 04/14	[2.52, 7.07]
Streeck et al.	Gangelt, Germany	03/31 - 04/06	[12.31, 24.40]
Stringhini et al.	Geneva, Switzerland	04/06 - 05/09	[8.15, 13.95]
Vos et al.	Netherlands	03/31 - 05/11	[2.10, 3.70]
Ward et al.	England, UK (2)	06/20 - 07/13	[5.78, 6.14]

Table 1: Seroprevalence studies selected for the analysis based on the list compiled by Chen et al. (2021) (listed in alphabetical order of authors), with geographic location of sampling, sampling dates, and 95% uncertainty interval for the infection rate (IR interval).

to suffer from substantial biases (Shook-Sa et al., 2020).

Chen et al. (2021) reviewed the literature for articles published between Dec 1, 2019, and Dec 22, 2020 and identified more than 400 unique seroprevalence studies. For each of these, Chen et al. (2021) determined study quality using a scoring system developed on the basis of a seroepidemiological protocol from the Consortium for the Standardization of Influenza Seroepidemiology (Horby et al., 2017). In total, Chen et al. (2021) identified 38 articles which considered a sample based on a general population and which obtained a study quality grade of A or B (see list of all 38 grade A or B general-population-based studies and citations in Chen et al. (2021), Table S8). We consider these 38 articles as a starting point for inclusion in our analysis.

Among the 38 articles, four studies represented results from different phases of the same study. For each of these we considered only the data from the earliest phase of the study. Stringhini et al. (2020) and Richard et al. (2020) are two publications that report the earlier and later phases, respectively, of the same study of Geneva, Switzerland. We considered only data from the earlier phase as reported in Stringhini et al. (2020). Murhekar et al. (2020*a*) and Murhekar et al. (2020*b*) are two publications that report the earlier and later phases, respectively, of the same study in India. We

Location	P_k	D_k lower	$D_{k,}$ upper	T_k	CC_k	$65yo_k$ (%)	GDP_k $
Saint Petersburg, Russia	5351935	2978	4776	233	21	18	29181
Two counties in GA, USA	1806672	221	247	419	11	12	58896
Orange county, CA, USA	3175692	556	979	285	33	15	80563
Ile-de-France, France	12174880	6766	7425	2414	243	14	50996
Jersey, UK	107800	27	28	648	19	17	49930
India	1028610328	4172	44760	1632	11	6	6997
England, UK (1)	56286961	27417	37264	3100	193	18	49930
Faroe Islands, Denmark	52154	1	1	592	3	17	62134
Spain	46459218	26834	26920	643	31	20	43444
Four counties in UT, USA	2200000	69	168	393	2	10	59892
Kupferzell, Germany	6247	3	3	1263	153	16	57558
Delhi, India	19800000	4188	26901	13966	3965	4	6997
Luxembourg	603951	89	109	1214	23	14	124569
Los Angeles county, CA, USA	10105518	943	1114	316	14	14	80563
Gangelt, Germany	12597	8	12	153	27	18	57558
Geneva, Switzerland	504128	221	280	442	48	16	72372
Netherlands	17181252	2955	5849	1652	47	20	61242
England, UK (2)	56286961	36229	36649	10635	634	18	49930

Table 2: All of the data required for the Bayesian evidence synthesis model.

considered only data from the first phase as reported in Murhekar et al. (2020*a*).[2]

Eight additional studies were not included due to unavailable mortality data for the specific target populations (Alemu et al., 2020, Ling et al., 2020, Mahajan et al., 2021*b*, Malani et al., 2021, Nisar et al., 2020, Poustchi et al., 2021, Shakiba et al., 2020, Tess et al., 2020); and four additional studies were not included because the articles failed to report 95% uncertainty intervals for the estimated infection rate in the target population (Borges et al., 2020, Majiya et al., 2020, Naranbhai et al., 2020, Wang et al., 2020). One study was not included due to missing dates for the sampling window (Gudbjartsson et al., 2020).

We excluded four studies that used "non-probability" or "convenience" sampling methods (e.g, studies in which participants were recruited using social media, or recruited in shopping centers). The sampling method for each study was determined by Arora et al. (2021). Specifically, we excluded: (1) McLaughlin et al. (2020) who sampled individuals from a list of volunteers (Arora et al. (2021) categorizes

[2] Note that, while similar in many ways, Ward et al. (2020) and Office of National Statistics (2020) are indeed two different large-sample studies. The Ward et al. (2020) study is based on the "REACT-2" survey which is led by Imperial College London, while the Office of National Statistics (2020) study is based on the "Coronavirus (COVID-19) Infection Survey" which is conducted by a partnership between the University of Oxford, University of Manchester, Public Health England and Wellcome Trust. Spiers (2020) discusses the differences between the two surveys.

the sampling method for this study as "self-referral"); (2) Rosenberg et al. (2020) who sampled individuals at grocery stores ("convenience"); (3) Appa et al. (2020) who recruited volunteers with support from Bolinas community leaders ("self-referral"); and (4) Bendavid et al. (2020) who recruited participants by placing targeted advertisements on Facebook ("stratified non-probability").

Finally, due to new information about the sensitivity of the Wondfo antibody tests (Silveira et al. (2021) : "Our findings cast serious doubts about the use of this brand of rapid tests for epidemiological studies."), we excluded the Hallal et al. (2020) Brazil study. Figure 1 shows a flowchart of the literature search and Table 3 in Appendix 6.1 lists all the excluded studies.

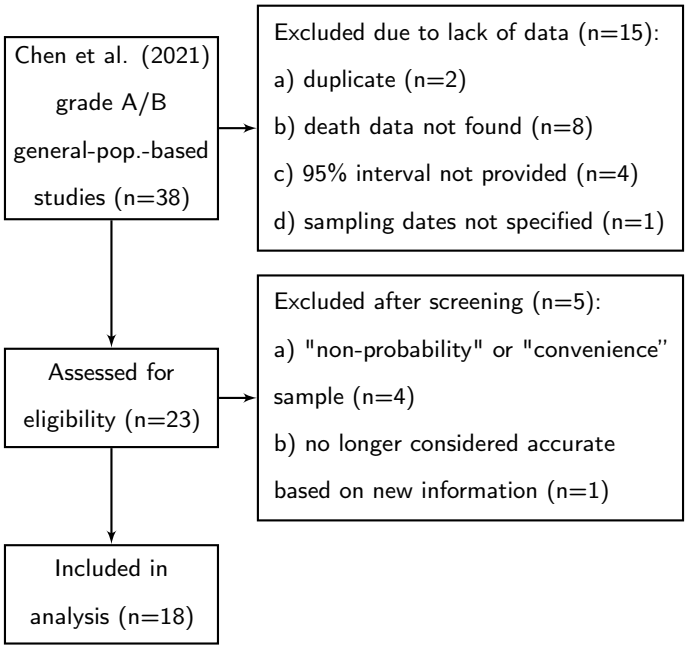

Figure 1: Flowchart of seroprevalence studies considered for analysis.

Our final set of seroprevalence studies consists of the $K = 18$ studies listed in Table 1. For each of these, we recorded the 95% uncertainty interval for the infection rate as reported in the article. If an article reported on multiple phases of a study (e.g., a longitudinal series of different surveys), or reported different results for different areas instead of an overall estimate (e.g., a series of different estimates for different regions), we selected only the first set of estimates. Furthermore, if a study reported more than one 95% uncertainty interval (e.g., different intervals corresponding to different adjustments 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 Table 1 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 Table 1 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).

3.2 Mortality data

Mortality data was obtained from various sources (e.g., academic, government, health authority); see details in Appendix 6.2. 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 are concerns that cause of death information may be very inaccurate and lead to biased COVID-19 mortality statistics. To overcome this issue, many suggest looking to "excess deaths" by comparing aggregate data for all-cause deaths from the time during the pandemic to the years prior (Leon et al., 2020). For populations with a large discrepancy between the number of deaths attributed to COVID-19 and the number of excess deaths –as suggested by the undercount ratio derived by Karlinsky and Kobak (2021)– we used excess deaths if these were available.

India is the only country represented in our data that is not included in Karlinsky and Kobak (2021)'s analysis, and according to Mukherjee et al. (2021): "no rigorous quantification of missing death numbers is currently available." However, there is evidence of potentially substantial under-reporting of COVID-19 deaths in India; see Pulla (2020). The analysis of Banaji (2021*b*) for the city of Mumbai suggests that: "there were between 11,000 and 25,000 COVID-19 deaths in the city during 2020" compared to the "11,116 recorded" by the municipal corporation. (Banaji (2021*b*): "Although Mumbai's data is far from complete, the city remains one of the few locations in India which has seen several serosurveys, and where some limited all cause mortality data is available"). Mukherjee et al. (2021) estimate the undercount ratio for each individual Indian state and Union territory. Based on Mukherjee et al. (2021)'s estimates, we multiply the upper bounds for the number

of deaths associated with the Mukherjee et al. (2021) ("India") study by a factor of 3.56 and for the Sharma et al. (2020) ("Delhi, India") study by a factor of 6.3 (see estimated underreporting factors in Figure S2 of Mukherjee et al. (2021)). (See also Bhattacharyya et al. (2021) who suggest that, for Delhi, in mid-August, 2020, the undercount ratio may be within the range of 8–13). Note that we do not change the lower bounds of the interval for these two studies. By considering the number of deaths in our analysis as interval censored data, we can account for the substantial uncertainty in these numbers and the possibility, however unlikely it may be, that the official India mortality numbers are accurate.

Of all the countries represented within our data that are included in Karlinsky and Kobak (2021)'s analysis, only Russia is associated with a large discrepancy between the official number of deaths attributed to COVID-19 and the number of excess deaths (with an estimated undercount ratio of 6.7). As such, for the "Saint Petersburg, Russia" study we use excess deaths as calculated by Kobak (2021).

4 Results

The model as described in Section 2, was fit to data as described in Section 3. We fit the model using JAGS (just another Gibbs sampler) (Kruschke, 2014), with 5 independent chains, each with 2 million draws (20% burn-in, thinning of 100); see Appendix 6.4 for details and JAGS code. Note that Z_{1k} was set equal to the centred and scaled logarithm of $65yo_k$, such that, for $k = 1, \ldots, K$: $Z_{1k} = (\log(65yo_k) - 2.63)/0.423$; and Z_{2k} was set equal to the centred and scaled logarithm of GDP_k, such that, for $k = 1, \ldots, K$: $Z_{2k} = (\log(GDP_k) - 10.75)/0.750$.

We report posterior median estimates and 95% highest probability density (HPD) credible intervals (CrI). Figure 2 plots the point estimates and credible intervals obtained for IFR_k and IR_k, for k in $1, \ldots, K$. The estimates for the study-specific IFR range from 0.13% for the Delhi, India, study, to 1.16% for Spain. Note that, in Figure 2, the seroprevalence studies are listed in order of their "fitted" IFR values (the posterior median of $g^{-1}(\theta_0 + \theta_1 Z_{1k} + \theta_2 Z_{2k})$, for k in $1, ..., K$, marked on the plot by the $\times$ symbols). For the other model parameters, we obtain:
$\hat{\theta}_0 = -5.22$, with 95% CrI of (-5.49, -4.97),

$\hat{\theta}_1 = 0.59$, with 95% CrI of (0.19, 1.00),

$\hat{\theta}_2 = -0.24$, with 95% CrI of (-0.62, 0.20),

$\hat{\tau} = 0.41$, with 95% CrI of (0.21, 0.69), and

$\hat{\sigma} = 1.17$, with 95% CrI of (0.80, 1.70).

Our estimate of $\hat{\theta}_1 = 0.59$ suggests 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 (Zimmermann and Curtis, 2021). Our estimate of $\hat{\theta}_2 = -0.24$ suggests that wealthier populations are more likely to have lower IFRs. However, we note that the wide credible interval for the θ_2 parameter overlaps zero. There are several reasons why we might have obtained this result. As with any observational data analysis, the estimate of θ_2 may suffer from bias due to unobserved confounding. Also, statistical power may be compromised by insufficient heterogeneity in the GDP per capita metric across the different populations in our analysis.

We also calculate point estimates for the average IFR amongst populations of a given age structure and wealth, 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*}. Thus we infer 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. Figure 3 plots these estimates using the 65yo and GDP values corresponding to those listed for 185 different countries by the World Bank's World Development Indicators (WDI) (Van Der Mensbrugghe, 2016).

We caution that the estimates plotted in Figure 3 do not correspond to the specific countries listed. The acutal IFR for a specific country may in fact be quite different. For example, while the estimated IFR for a population with a similar age structure and GDP as Japan is 1.47% (Japan has 28% of its population above 65 years old, and a GDP (ppp) per capita of $42,338), the actual IFR for the country of Japan is likely much lower. Consider that, as of December 31, 2020, the case fatality rate in Japan, which represents a very liberal upper-bound on the IFR, was only 1.48% (calculated as 3,414 deaths out of 230,304 reported cases, according to data from JMoHLW (2020)).[3] As another example, consider Nicaragua which has a relatively young

[3] Indeed, in order for Japan's IFR to be equal to or above 1.00%, the number of confirmed cases would need

population with only 5% above 65 and a GDP (ppp) per capita of $5,646. We suspect the IFR may be substantially higher than the estimate of 0.29% displayed in Figure 3. As of August 31, 2020, the overall population fatality rate (PFR) in Nicaragua, which represents a very conservative lower-bound on the IFR, was approximately 0.12% (calculated as 7,528 excess deaths out of a total population of 6.546 million, according to data from Karlinsky and Kobak (2021)).

We also calculate posterior point and interval estimates corresponding to age and wealth values that match the entire world, the United States (USA), and for the European Union (EU) (as listed by the World Bank's World Development Indicators (WDI)); see "World", "USA", an "EU" rows in Figure 2. For each of these estimates, in order to better understand the heterogeneity at play, we calculate a corresponding 95% prediction interval (PrI) in addition to a 95% CrI. The PrI provides the range of values within which we are likely to find the true IFR for a population, when all we know of that population is its 65yo value and its GDP value.[4] For 65yo = 9 and GDP = 17,811, the approximate worldwide values, we obtain an across-population average IFR estimate of 0.40%, with a 95% CrI of (0.20%, 0.63%) and a 95% PrI of (0.08%, 0.93%). For 65yo = 16 and GDP = 65,297, the USA values, we obtain an across-population average IFR estimate of 0.59%, with a 95% CrI of (0.43%, 0.75%) and a 95% PrI of (0.14%, 1.29%). For 65yo = 20 and GDP = 47,828, the EU values, we obtain an across-population average IFR estimate of 0.89%, with a 95% CrI of (0.56%, 1.27%) and a 95% PrI of (0.20%, 1.97%).

The robustness of our estimates was checked using a leave-one-out sensitivity analysis (Iyengar and Greenhouse, 2009); see Table 4 in the Appendix. This sensitivity analysis showed that our estimates may be somewhat sensitive to the study of Snoeck et al. (2020). For instance, without the Snoeck et al. (2020) data, our

to represent at least two thirds of the total number of infections. This seems highly unlikely given that, at least during the early stages of the pandemic, Japan was "only testing people who are already quite sick" (Wingfield-Hayes, 2020). Throughout the entire pandemic, the amount of testing in Japan was relatively low leading Omori et al. (2020) to estimate that the "ascertainment rate of non-severe cases was estimated at 0.44 (95% CI 0.37–0.50)." This would imply that a substantial proportion of cases went uncounted and suggests that Japan's IFR is substantially lower than the CFR, and likely much lower than 1%. However, note that there are also good reasons to suspect that Ioannidis (2021*a*)'s very low estimate of 0.02% is an underestimate. Ioannidis (2021*a*)'s estimate is based on likely unrepresentative seroprevalence data from Doi et al. (2020), Takita et al. (2020), and Nawa et al. (2020), which receive study quality grades of "D", "C", and "C", respectively, from Chen et al. (2021).

[4] Mathematically, the prediction interval describes the posterior distribution of $g^{-1}(\theta_0 + \theta_1 z_{1*} + \theta_2 z_{2*} + \tau\epsilon)$, where the posterior distribution is augmented to include $\epsilon \sim \mathcal{N}(0, 1)$, independently of the other parameters. For more on the relative interpretations of credible and prediction intervals, see Higgins et al. (2009), Riley et al. (2011), and IntHout et al. (2016).

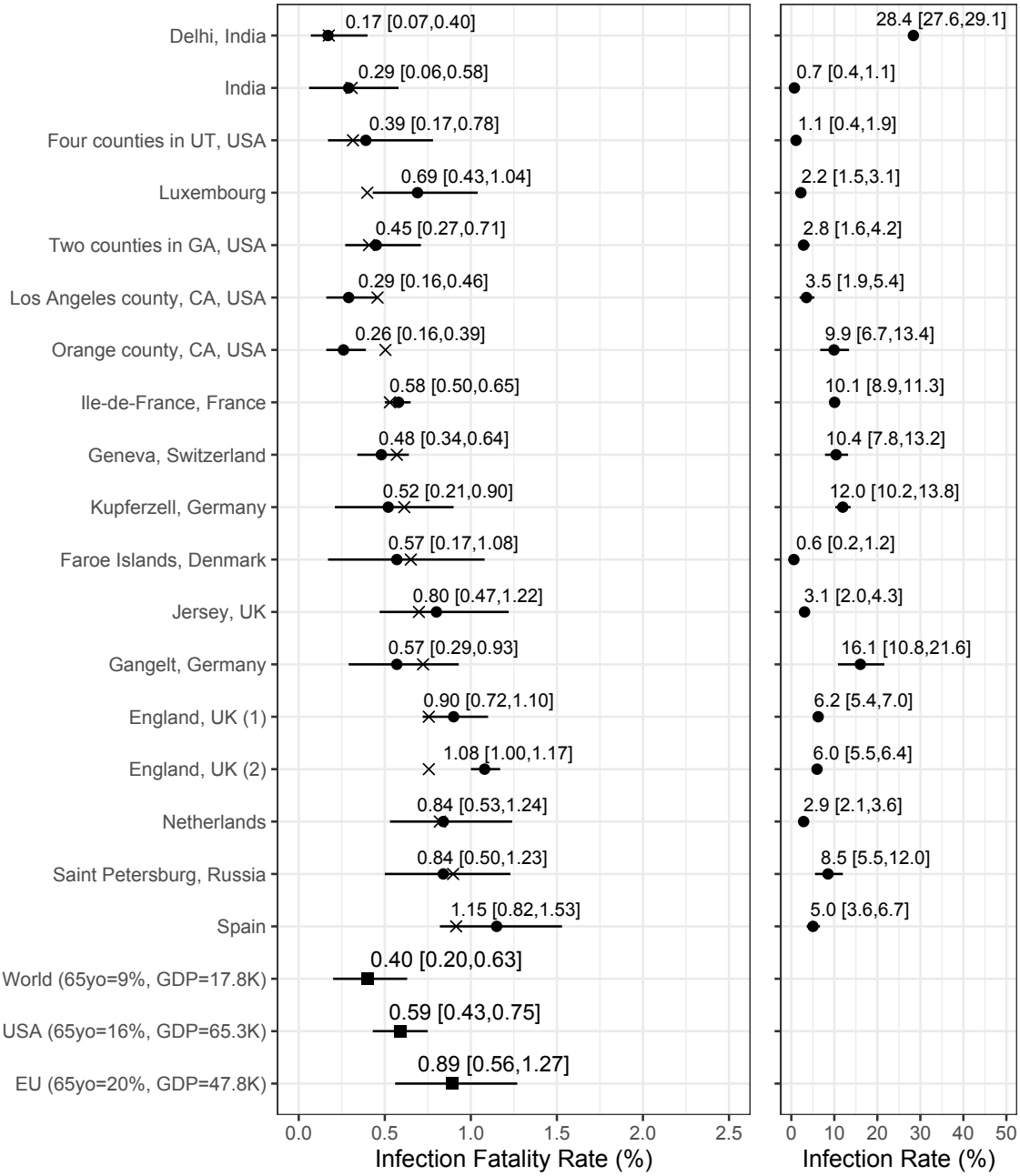

Figure 2: Posterior median estimates for the IR_k and IFR_k variables (for $k = 1, \ldots, 18$) with 95% HPD CrIs. Studies are listed from top to bottom in order of increasing fitted values (these values are indicated by $\times$). Also plotted, under the labels "World (65 yo=9%, GDP=17.8K)", "USA (65 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 \$65,297 (the USA values), and of 20% and of \$47,828 (the EU values).

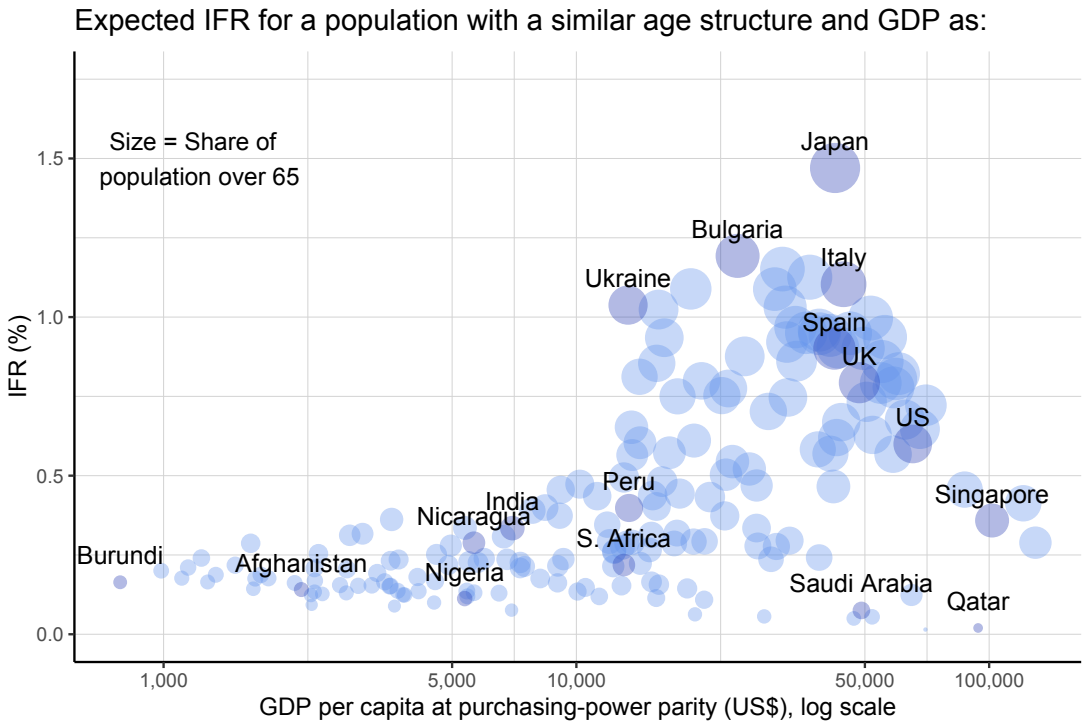

Figure 3: Each of the 185 points corresponds to an estimate for the expected IFR for a *typical* population having the same proportion of the populace aged over 65 and the same GDP per capita as each of 185 different countries. This Figure is based on a a similar Figure by Solstad et al. (2021).

estimate for τ, the between group infection fatality rate heterogeneity parameter, is 0.29 instead of 0.41. (We note that, while Arora et al. (2021) categorizes the sampling method for the Snoeck et al. (2020) study as "stratified probability," Snoeck et al. (2020) explains that that "the sample of participants was enrolled through the use of a non-probabilistic web panel" and caution that "[t]here are pros and cons of using web panels for surveys," specifically noting that "online surveys may be biased samples because the respondents are self-selected.") Elimination of each of the other studies from the evidence synthesis did not have a substantial influence on our results. We also repeated our analysis 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 analysis; see Figure 4 in the Appendix.

Our estimates are somewhat similar to those obtained in other analyses. Brazeau et al. (2020), 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. (2020) obtain an estimate of 1.15% (95% prediction interval of 0.78%-1.79%). Ioannidis (2021*a*), using data from seroprevalence studies with sample sizes greater than 500, 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. (2020), who restricted their analysis to populations in "advanced economies," do not provide an overall IFR, but instead (perhaps more appropriately) provide age-group specific estimates. For the 45–54 year old age group, Levin et al. (2020) 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%). Pei et al. (2021) using a rather complex Bayesian "metapopulation SEIR model to simulate the transmission of COVID-19" 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 estimate of 0.59% for the United States is based on data obtained mostly between April, 2020 and June, 2020 (see Table 1).

We can also compare our study-specific IFR estimates to those obtained from other analyses. Perez-Saez et al. (2021), based on the data of Stringhini et al. (2020), estimate the IFR for the canton of Geneva, Switzerland (for the period of time up until early June, 2020) to be 0.64% (95%CrI 0.38%-0.98%). Our estimate is somewhat lower at 0.48% (95%CrI 0.34%- 0.64%).

Pastor-Barriuso et al. (2020), based on the data of Pollán et al. (2020), estimate an IFR for Spain (for the period of time up to mid-July 2020) of 0.83%-1.07%, noting that their estimates "do not apply to people living in nursing homes in Spain." Our estimate for Spain (which is based on deaths both inside and outside of nursing homes) is: 1.15% (95% CrI 0.82%-1.53%). Public Health England, citing Ward et al. (2020), estimate a IFR for England of 0.90%, noting that this is "calculated excluding care home residents" (Public Health England, 2020). This compares to our estimates for England of 0.91% and 1.08% (which are calculated based on mortality data that includes care home residents). Kümmerer et al. (2020), based on the data of Streeck et al. (2020), estimate the IFR for Gangelt, Germany (for the period up until early April, 2020) to be 0.37% (95% CrI 0.12%-0.67%). This is much lower than our estimate of 0.57% (95% CrI 0.29%-0.93%). Finally, Banaji (2021*a*) provide various estimates

for the city of Chennai, India, (based on excess death data). These estimates range between 0.22% and 0.41%. Our estimate for India is 0.29% (95% CrI 0.06%-0.58%).

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.

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. We also wish to emphasize that lack of available mortality data (for the specific geographic areas defined in the seroprevalence studies) was also the main reason for excluding seroprevalence studies from our analysis (8 studies were excluded for this reason).

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 far from perfect. Seroprevalence studies are severely limited by the representativeness of the individuals they test. Certain groups of individuals are unlikely to be tested in a seroprevalence study and these groups often have very high infection rates (e.g., institutionalized populations, hospitalized populations, 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 (Shook-Sa et al., 2020).

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" (Eysenck, 1978), meaning that if one includes biased studies in one's analysis, the analysis results will themselves be biased (Sharpe, 1997). We only included data from 18 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. Indeed, while we include four different seroprevalence studies from the United States, not a single study from Africa or the Middle East was included. 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 (Ioannidis and Lau, 1998). Novel methods which allow evidence syntheses to appropriately incorporate biased data are urgently needed. Recently, Campbell et al. (2020) proposed a partially identified model to combine seroprevalence study data with data from official statistics that are known to be biased due to "preferential testing."

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 (Faust, 2020, Ioannidis, 2020, Lipsitch, 2020) and immense efforts have been made in the collection and analysis of data. And yet, even after more than a year, there is still a large amount of uncertainty and unexplained heterogeneity surrounding the COVID-19 IFR. While a certain amount of heterogeneity is to be expected (Higgins, 2008), 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 (Berlin, 1995, Ioannidis and Lau, 1998).

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 analysis fits on half a page in Table 2 and that the entire JAGS MCMC code fits on less than a single page (see Appendix 6.4). One model extension that could be pursued would involve age stratification of IFR. Age-group specific mortality data is available for many geographic areas (e.g., Riffe et al. (2021)) and such data could inform an extended version of our model, thereby offering an alternative to the approach described by Levin et al. (2020) 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, some of which were obtained more than a year ago (see dates listed in Table 1). It is likely that, with continual viral mutation of SARS-CoV-2 and advances in treatment, the current IFR in many places is now markedly different than it was earlier, and our estimates are therefore likely to be outdated (Pei et al., 2021, Pietzonka et al., 2021, Walensky et al., 2021). In particular, at the present time, India is experiencing a rapid increase in COVID-19 fatalities which suggests that the current IFR in India may be much higher now than during earlier phases of the pandemic (Padma, 2021, Thiagarajan, 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 01886-000.*

Code - Code to replicate our analysis is available in the Appendix and at `https://github.com/harlanhappydog/BayesianSeroMetaAnalysis`.

References

Alemu, B. N., Addissie, A., Mamo, G., Deyessa, N., Abebe, T., Abagero, A., Ayele, W., Abebe, W., Haile, T., Argaw, R., Amogne, W., Belachew, A., Desalegn, Z., Teka, B., Kantelhardt, E., Wossen, M., Abdella, S., Tollera, G. and Tadesse, L. (2020), 'Sero-prevalence of anti-SARS-CoV-2 antibodies in Addis Ababa, Ethiopia', *bioRxiv* .

Appa, A., Takahashi, S., Rodriguez-Barraquer, I., Chamie, G., Sawyer, A., Duarte, E., Hakim, J., Turcios, K., Vinden, J., Janson, O. et al. (2020), Universal pcr and antibody testing demonstrate little to no transmission of SARS-CoV-2 in a rural community, *in* 'Open Forum Infectious Diseases'.

Arora, R. K., Joseph, A., Van Wyk, J., Rocco, S., Atmaja, A., May, E., Yan, T., Bobrovitz, N., Chevrier, J., Cheng, M. P. et al. (2021), 'Serotracker: a global SARS-CoV-2 seroprevalence dashboard', *The Lancet Infectious Diseases* **21**(4), e75–e76.

Banaji, M. (2021*a*), 'Estimates of COVID-19 infection fatality rate (IFR) in Chennai during 2020', `http://maths.mdx.ac.uk/research/modelling-the-covid-19-pandemic/chennaiifrestimates/`.

Banaji, M. (2021*b*), 'Estimating COVID-19 infection fatality rate in mumbai during 2020', *medRxiv* .

Barchuk, A., Skougarevskiy, D., Titaev, K., Shirokov, D., Raskina, Y., Novkunkskaya, A., Talantov, P., Isaev, A., Pomerantseva, E., Zhikrivetskaya, S. et al. (2020), 'Seroprevalence of SARS-CoV-2 antibodies in Saint Petersburg, Russia: a population-based study', medRxiv.

Bendavid, E., Mulaney, B., Sood, N., Shah, S., Ling, E., Bromley-Dulfano, R., Lai, C., Weissberg, Z., Saavedra, R., Tedrow, J. et al. (2020), 'COVID-19 antibody seroprevalence in Santa Clara County, California', medRxiv.

Berger, J. O. (2013), *Statistical decision theory and Bayesian analysis*, Springer Science & Business Media.

Berlin, J. A. (1995), 'Invited commentary: benefits of heterogeneity in meta-analysis of data from epidemiologic studies', *American Journal of Epidemiology* **142**(4), 383–387.

Bhattacharyya, R., Kundu, R., Bhaduri, R., Ray, D., Beesley, L. J., Salvatore, M. and Mukherjee, B. (2021), 'Incorporating false negative tests in epidemiological models for SARS-CoV-2 transmission and reconciling with seroprevalence estimates', *Scientific reports* **11**(1), 1–14.

Biggs, H. M., Harris, J. B., Breakwell, L., Dahlgren, F. S., Abedi, G. R., Szablewski, C. M., Drobeniuc, J., Bustamante, N. D., Almendares, O., Schnall, A. H. et al. (2020), 'Estimated community seroprevalence of SARS-CoV-2 antibodies—two Georgia counties, April 28–May 3, 2020', *Morbidity and Mortality Weekly Report* **69**(29), 965.

Bobrovitz, N., Arora, R. K., Cao, C., Boucher, E., Liu, M., Rahim, H., Donnici, C., Ilincic, N., Duarte, N., Van Wyk, J. et al. (2020), 'Global seroprevalence of SARS-CoV-2 antibodies: a systematic review and meta-analysis', medRxiv.

Borges, L. P., Martins, A. F., de Melo, M. S., de Oliveira, M. G. B., de Rezende Neto, J. M., Dósea, M. B., Cabral, B. C. M., Menezes, R. F., Santos, A. A., Matos, I. L. S. et al. (2020), 'Seroprevalence of SARS-CoV-2 IgM and IgG antibodies in an asymptomatic population in Sergipe, Brazil', *Revista Panamericana de Salud Pública* **44**.

Brazeau, N., Verity, R., Jenks, S., Fu, H., Whittaker, C., Winskill, P., Dorigatti, I., Walker, P., Riley, S., Schnekenberg, R. P. et al. (2020), 'Report 34: COVID-19 infection fatality ratio: estimates from seroprevalence'.

Brody-Moore, P. (2019), 'Bayesian hierarchical meta-analysis of asymptomatic ebola seroprevalence'.

Brownstein, N. C. and Chen, Y. A. (2021), 'Predictive values, uncertainty, and interpretation of serology tests for the novel coronavirus', *Scientific Reports* **11**(1), 1–12.

Bruckner, T. A., Parker, D. M., Bartell, S. M., Vieira, V. M., Khan, S., Noymer, A., Drum, E., Albala, B., Zahn, M. and Boden-Albala, B. (2021), 'Estimated seroprevalence of SARS-CoV-2 antibodies among adults in Orange County, California', *Scientific Reports* **11**(1), 1–9.

Campbell, H., de Valpine, P., Maxwell, L., de Jong, V. M., Debray, T., Jänisch, T. and Gustafson, P. (2020), 'Bayesian adjustment for preferential testing in estimating the COVID-19 infection fatality rate: Theory and methods', arXiv - preprint arXiv:2005.08459.

Carrat, F., de Lamballerie, X., Rahib, D., Blanche, H., Lapidus, N., Artaud, F., Kab, S., Renuy, A., de Edelenyi, F. S., Meyer, L. et al. (2020), 'Seroprevalence of SARS-CoV-2 among adults in three regions of france following the lockdown and associated risk factors: a multicohort study'.

Chen, X., Chen, Z., Azman, A. S., Deng, X., Sun, R., Zhao, Z., Zheng, N., Chen, X., Lu, W., Zhuang, T. et al. (2021), 'Serological evidence of human infection with SARS-CoV-2: a systematic review and meta-analysis', *The Lancet Global Health* **9**(5), E598–E609.

Clapham, H., Hay, J., Routledge, I., Takahashi, S., Choisy, M., Cummings, D., Grenfell, B., Metcalf, C. J. E., Mina, M., Barraquer, I. R. et al. (2020), 'Seroepidemiologic study designs for determining SARS-CoV-2 transmission and immunity', *Emerging Infectious Diseases* **26**(9), 1978.

Doi, A., Iwata, K., Kuroda, H., Hasuike, T., Kanda, A., Nagao, T., Nishioka, H., Tomii, K., Morimoto, T. and Kihara, Y. (2020), 'Seroprevalence of novel coronavirus disease (COVID-19) in Kobe', medRxiv.

Eysenck, H. J. (1978), 'An exercise in mega-silliness.', *American Psychologist* **33**(5), 517.

Faust, J. S. (2020), 'Comparing COVID-19 deaths to flu deaths is like comparing apples to oranges', Scientific American, `https://tinyurl.com/ydxx8el8`.

Gelman, A. et al. (2006), 'Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper)', *Bayesian Analysis* **1**(3), 515–534.

Gudbjartsson, D. F., Norddahl, G. L., Melsted, P., Gunnarsdottir, K., Holm, H., Eythorsson, E., Arnthorsson, A. O., Helgason, D., Bjarnadottir, K., Ingvarsson, R. F. et al. (2020), 'Humoral immune response to SARS-CoV-2 in Iceland', *New England Journal of Medicine* **383**(18), 1724–1734.

Hallal, P. C., Hartwig, F. P., Horta, B. L., Silveira, M. F., Struchiner, C. J., Vidaletti, L. P., Neumann, N. A., Pellanda, L. C., Dellagostin, O. A., Burattini, M. N. et al. (2020), 'SARS-CoV-2 antibody prevalence in Brazil: results from two successive nationwide serological household surveys', *The Lancet Global Health* **8**(11), e1390–e1398.

Higgins, J. P. (2008), 'Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified', *International journal of epidemiology* **37**(5), 1158–1160.

Higgins, J. P., Thompson, S. G. and Spiegelhalter, D. J. (2009), 'A re-evaluation of random-effects meta-analysis', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **172**(1), 137–159.

Horby, P. W., Laurie, K. L., Cowling, B. J., Engelhardt, O. G., Sturm-Ramirez, K., Sanchez, J. L., Katz, J. M., Uyeki, T. M., Wood, J., Van Kerkhove, M. D. et al. (2017), 'Consise statement on the reporting of seroepidemiologic studies for influenza (roses-i statement): an extension of the strobe statement', *Influenza and other respiratory viruses* **11**(1), 2–14.

IntHout, J., Ioannidis, J. P., Rovers, M. M. and Goeman, J. J. (2016), 'Plea for routinely presenting prediction intervals in meta-analysis', *BMJ open* **6**(7).

Ioannidis, J. P. (2020), 'First Opinion: A fiasco in the making? as the coronavirus pandemic takes hold, we are making decisions without reliable data', STAT, `https://tinyurl.com/uj539o4`.

Ioannidis, J. P. (2021*a*), 'Infection fatality rate of COVID-19 inferred from seroprevalence data', *Bulletin of the World Health Organization* **99**(1), 19.

Ioannidis, J. P. (2021*b*), 'Reconciling estimates of global spread and infection fatality rates of COVID-19: an overview of systematic evaluations', *European Journal of Clinical Investigation* p. e13554.

Ioannidis, J. P. and Lau, J. (1998), 'Can quality of clinical trials and meta-analyses be quantified?', *The Lancet* **352**(9128), 590.

Iyengar, S. and Greenhouse, J. (2009), 'Sensitivity analysis and diagnostics', *Handbook of research synthesis and meta-analysis* pp. 417–433.

JMoHLW (2020), 'Japanese ministry of health, labour, and welfare - covid database', *`https://www.mhlw.go.jp/stf/houdou/houdou_list_202105.html`* (*accessed May 2, 2021*) .

Karlinsky, A. and Kobak, D. (2021), 'The world mortality dataset: Tracking excess mortality across countries during the COVID-19 pandemic', medRxiv.

Kobak, D. (2021), 'Excess mortality reveals Covid's true toll in Russia', *Significance* **18**(1), 16.

Kruschke, J. (2014), *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*, Academic Press.

Kümmerer, M., Berens, P. and Macke, J. (2020), '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`.

Lambert, P. C., Sutton, A. J., Burton, P. R., Abrams, K. R. and Jones, D. R. (2005), 'How vague is vague? A simulation study of the impact of the use of vague prior distributions in mcmc using winbugs', *Statistics in Medicine* **24**(15), 2401–2428.

Leon, D. A., Shkolnikov, V. M., Smeeth, L., Magnus, P., Pechholdová, M. and Jarvis, C. I. (2020), 'COVID-19: a need for real-time monitoring of weekly excess deaths', *The Lancet* **395**(10234), E81.

Levin, A. T., Hanage, W. P., Owusu-Boaitey, N., Cochran, K. B., Walsh, S. P. and Meyerowitz-Katz, G. (2020), 'Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications', *European Journal of Epidemiology* pp. 1–16.

Ling, R., Yu, Y., He, J., Zhang, J., Xu, S., Sun, R., Li, T., Ji, H.-L. and Wang, H.-Q. (2020), 'Seroprevalence and epidemiological characteristics of immunoglobulin M and G antibodies against SARS-CoV-2 in asymptomatic people in Wuhan, China', *China* (*6/15/2020*) .

Linton, N. M., Kobayashi, T., Yang, Y., Hayashi, K., Akhmetzhanov, A. R., Jung, S.-m., Yuan, B., Kinoshita, R. and Nishiura, H. (2020), 'Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data', *Journal of Clinical Medicine* **9**(2), 538.

Lipsitch, M. (2020), 'First Opinion: We know enough now to act decisively against COVID-19. social distancing is a good place to start', STAT, `https://tinyurl.com/yx4gf9mr`.

Mahajan et al. (2021*a*), 'SARS-CoV-2 infection hospitalization rate and infection fatality rate among the non-congregate population in Connecticut', *The American Journal of Medicine* .

Mahajan et al. (2021*b*), 'Seroprevalence of SARS-CoV-2-specific IgG antibodies among adults living in Connecticut: Post-infection prevalence (pip) study', *The American Journal of Medicine* **134**(4), 526–534.

Majiya, H., Aliyu-Paiko, M., Balogu, V. T., Musa, D. A., Salihu, I. M., Kawu, A. A., Bashir, I. Y., Sani, A. R., Baba, J., Muhammad, A. T. et al. (2020), 'Seroprevalence of COVID-19 in Niger State', medRxiv.

Malani, A., Shah, D., Kang, G., Lobo, G. N., Shastri, J., Mohanan, M., Jain, R., Agrawal, S., Juneja, S., Imad, S. et al. (2021), 'Seroprevalence of SARS-CoV-2 in slums versus non-slums in Mumbai, India', *The Lancet Global Health* **9**(2), e110–e111.

McLaughlin, C. C., Doll, M. K., Morrison, K. T., McLaughlin, W. L., O'Connor, T., Sholukh, A. M., Bossard, E. L., Phasouk, K., Ford, E. S., Diem, K. et al. (2020), 'High community SARS-CoV-2 antibody seroprevalence in a ski resort community, Blaine County, Idaho, US. preliminary results', medRxiv.

Meyerowitz-Katz, G. and Merone, L. (2020), 'A systematic review and meta-analysis of published research data on COVID-19 infection-fatality rates', *International Journal of Infectious Diseases* **101**, 138–148.

Mukherjee, B., Purkayashtha, S., Kundu, R. and Bhaduri, R. (2021), 'Estimating the infection fatality rate from SARS-CoV-2 in India', Available at SSRN 3798552.

Murhekar et al. (2020*a*), 'Prevalence of SARS-CoV-2 infection in India: Findings from the national serosurvey, May-June 2020', *Indian Journal of Medical Research* **152**(1), 48.

Murhekar et al. (2020*b*), 'SARS-CoV-2 antibody prevalence in India: Findings from the second nation-wide household serosurvey, August-September 2020'.

Naranbhai, V., Chang, C. C., Beltran, W. F. G., Miller, T. E., Astudillo, M. G., Villalba, J. A., Yang, D., Gelfand, J., Bernstein, B. E., Feldman, J. et al. (2020), 'High seroprevalence of anti-SARS-CoV-2 antibodies in Chelsea, Massachusetts', *The Journal of Infectious Diseases* **222**(12), 1955–1959.

Nawa, N., Kuramochi, J., Sonoda, S., Yamaoka, Y., Nukui, Y., Miyazaki, Y. and Fujiwara, T. (2020), 'Seroprevalence of sars-cov-2 IgG antibodies in Utsunomiya City, Greater Tokyo, after first pandemic in 2020 (u-corona): a household-and population-based study', medRxiv.

Nisar, M. I., Nadia Ansari, M. and Khalid, F. (2020), 'Serial population-based sero-surveys for COVID-19 in low and high transmission'.

O'Driscoll, M., Dos Santos, G. R., Wang, L., Cummings, D. A., Azman, A. S., Paireau, J., Fontanet, A., Cauchemez, S. and Salje, H. (2020), 'Age-specific mortality and immunity patterns of SARS-CoV-2', *Nature* pp. 1–6.

Office of National Statistics (2020), 'Coronavirus (COVID-19) infection survey: England', *https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/coronaviruscovid19infectionsurveydata* .

Omori, R., Mizumoto, K. and Nishiura, H. (2020), 'Ascertainment rate of novel coronavirus disease (COVID-19) in Japan', *International Journal of Infectious Diseases* **96**, 673–675.

Padma, T. (2021), 'India's COVID-vaccine woes—by the numbers', *Nature* **592**(7855), 500–501.

Pastor-Barriuso, R., Perez-Gomez, B., Hernan, M. A., Perez-Olmeda, M., Yotti, R., Oteo, J., Sanmartin, J. L., Leon-Gomez, I., Fernandez-Garcia, A., Fernandez-Navarro, P. et al. (2020), 'SARS-CoV-2 infection fatality risk in a nationwide seroepidemiological study', medRxiv.

Pei, S., Yamana, T. K., Kandula, S., Galanti, M. and Shaman, J. (2021), 'Overall burden and characteristics of COVID-19 in the United States during 2020', *medRxiv* .

Perez-Saez, J., Lauer, S. A., Kaiser, L., Regard, S., Delaporte, E., Guessous, I., Stringhini, S., Azman, A. S., Alioucha, D., Arm-Vernez, I. et al. (2021), 'Serology-informed estimates of SARS-CoV-2 infection fatality risk in Geneva, Switzerland', *The Lancet Infectious Diseases* **21**(4), e69–e70.

Petersen, M. S., Strøm, M., Christiansen, D. H., Fjallsbak, J. P., Eliasen, E. H., Johansen, M., Veyhe, A. S., Kristiansen, M. F., Gaini, S., Møller, L. F. et al. (2020), 'Seroprevalence of SARS-CoV-2–specific antibodies, Faroe Islands', *Emerging Infectious Diseases* **26**(11), 2760.

Pietzonka, P., Brorson, E., Bankes, W., Cates, M. E., Jack, R. L. and Adhikari, R. (2021), 'Bayesian inference across multiple models suggests a strong increase in lethality of COVID-19 in late 2020 in the UK', medRxiv.

Pollán, M., Pérez-Gómez, B., Pastor-Barriuso, R., Oteo, J., Hernán, M. A., Pérez-Olmeda, M., Sanmartín, J. L., Fernández-García, A., Cruz, I., de Larrea, N. F. et al. (2020), 'Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study', *The Lancet* **396**(10250), 535–544.

Poustchi, H., Darvishian, M., Mohammadi, Z., Shayanrad, A., Delavari, A., Bahadorimonfared, A., Eslami, S., Javanmard, S. H., Shakiba, E., Somi, M. H. et al. (2021), 'SARS-CoV-2 antibody seroprevalence in the general population and high-risk occupational groups across 18 cities in Iran: a population-based cross-sectional study', *The Lancet Infectious Diseases* **21**(4), 473–481.

Public Health England (2020), 'COVID-19: the green book, chapter 14a', https://www.gov.uk/government/publications/covid-19-the-green-book-chapter-14a.

Pulla, P. (2020), 'What counts as a cOVID-19 death?', *The BMJ* **370**.

Richard, A., Wisniak, A., Perez-Saez, J., Garrison-Desany, H., Petrovic, D., Piumatti, G., Baysson, H., Picazio, A., Pennacchio, F., De Ridder, D. et al. (2020), 'Seroprevalence of anti-SARS-CoV-2 IgG antibodies, risk factors for infection and associated symptoms in Geneva, Switzerland: a population-based study', medRxiv.

Riffe, T., Acosta, A. and Zarulli, V. (2021), 'COVerAGE-DB: A global demographic database of COVID-19 cases and deaths', *International Journal of Epidemiology* .

Riley, R. D., HIgGins, J. P. and Deeks, J. J. (2011), 'Interpretation of random effects meta-analyses', *The BMJ* **342**.

Rosenberg, E. S., Tesoriero, J. M., Rosenthal, E. M., Chung, R., Barranco, M. A., Styer, L. M., Parker, M. M., Leung, S.-Y. J., Morne, J. E., Greene, D. et al. (2020), 'Cumulative incidence and diagnosis of SARS-CoV-2 infection in New York', *Annals of Epidemiology* **48**, 23–29.

Samore, M., Looney, A., Orleans, B., Greene, T., Seegert, N., Delgado, J. C., Presson, A., Zhang, C., Ying, J., Zhang, Y. et al. (2020), 'SARS-CoV-2 seroprevalence and detection fraction in Utah urban populations from a probability-based sample', medRxiv.

Santos-Hövener, C., Neuhauser, H. K., Rosario, A. S., Busch, M., Schlaud, M., Hoffmann, R., Gößwald, A., Koschollek, C., Hoebel, J., Allen, J. et al. (2020), 'Serology-and PCR-based cumulative incidence of SARS-CoV-2 infection in adults in a successfully contained early hotspot (comolo study), Germany, May to June 2020', *Eurosurveillance* **25**(47), 2001752.

Shakiba, M., Nazari, S. S. H., Mehrabian, F., Rezvani, S. M., Ghasempour, Z. and Heidarzadeh, A. (2020), 'Seroprevalence of COVID-19 virus infection in Guilan province, Iran', medRxiv.

Sharma, N., Sharma, P., Basu, S., Saxena, S., Chawla, R., Dushyant, K., Mundeja, N., Marak, Z. S., Singh, S., Singh, G. K. et al. (2020), 'The seroprevalence and trends of SARS-CoV-2 in Delhi, India: A repeated population-based seroepidemiological study', medRxiv.

Sharpe, D. (1997), 'Of apples and oranges, file drawers and garbage: Why validity issues in meta-analysis will not go away', *Clinical Psychology Review* **17**(8), 881–901.

Shook-Sa, B. E., Boyce, R. M. and Aiello, A. E. (2020), 'Estimation without representation: early severe acute respiratory syndrome coronavirus 2 seroprevalence studies and the path forward', *The Journal of Infectious Diseases* **222**(7), 1086–1089.

Silveira, M. F., Mesenburg, M., Dellagostin, O. A., Oliveira, N. R., Maia, M. A. C., Santos, F. D. S., Vale, A. M., Barros, A., Vidaletti, L. P., Barros, F. C. et al. (2021), 'Time-dependent decay of detectable antibodies against sars-cov-2: A comparison of elisa with two batches of a lateral-flow test'.

Snoeck, C. J., Vaillant, M., Abdelrahman, T., Satagopam, V. P., Turner, J. D., Beaumont, K., Gomes, C. P., Fritz, J. V., Schröder, V. E., Kaysen, A. et al. (2020), 'Prevalence of SARS-CoV-2 infection in the Luxembourgish population: the CON-VINCE study.', medRxiv.

Solstad et al. (2021), 'Modelling Covid-19's death toll - There have been 7m-13m excess deaths worldwide during the pandemic', The Economist; `https://www.economist.com/briefing/2021/05/15/there-have-been-7m-13m-excess-deaths-worldwide-during-the-pandemic`.

Sood, N., Simon, P., Ebner, P., Eichner, D., Reynolds, J., Bendavid, E. and Bhattacharya, J. (2020), 'Seroprevalence of SARS-CoV-2–specific antibodies among adults in Los Angeles County, California, on april 10-11, 2020', *JAMA* **323**(23), 2425–2427.

Spiers, N. A. (2020), 'Rapid response: Comparing REACT with the ONS Infection Survey', *The BMJ* . **URL:** *https://www.bmj.com/content/371/bmj.m3850/rr*

Statistics Jersey (2020), 'SARS-CoV-2: prevalence of antibodies in Jersey', *https://tinyurl.com/6htjn3af* .

Streeck, H., Schulte, B., Kümmerer, B. M., Richter, E., Höller, T., Fuhrmann, C., Bartok, E., Dolscheid-Pommerich, R., Berger, M., Wessendorf, L., Eschbach-Bludau, M., Kellings, A., Schwaiger, A., Coenen, M., Hoffmann, P., Stoffel-Wagner, B., Nöthen, M. M., Eis-Hübinger, A. M., Exner, M., Schmithausen, R. M., Schmid, M. and Hartmann, G. (2020), 'Infection fatality rate of SARS-CoV-2 infection in a German community with a super-spreading event', *Nature Communications* **11**(1), 5829.

Stringhini, S., Wisniak, A., Piumatti, G., Azman, A. S., Lauer, S. A., Baysson, H., De Ridder, D., Petrovic, D., Schrempft, S., Marcus, K. et al. (2020), 'Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (serocov-pop): a population-based study', *The Lancet* **396**(10247), 313–319.

Takita, M., Matsumura, T., Yamamoto, K., Yamashita, E., Hosoda, K., Hamaki, T. and Kusumi, E. (2020), 'Geographical profiles of COVID-19 outbreak in tokyo: an analysis of the primary care clinic–based point-of-care antibody testing', *Journal of Primary Care & Community Health* **11**, 2150132720942695.

Tess, B. H., Granato, C. F., Alves, M. C., Pintao, M. C., Rizzatti, E., Nunes, M. C. and Reinach, F. C. (2020), 'SARS-CoV-2 seroprevalence in the municipality of São Paulo, Brazil, ten weeks after the first reported case', medRxiv.

Thiagarajan, K. (2021), 'Why is India having a COVID-19 surge?', *The BMJ* **373**.

Van Der Mensbrugghe, D. (2016), 'Using R to extract data from the World Bank's World Development Indicators', *Journal of Global Economic Analysis* **1**(1), 251–283.

Vos, E. R. A., den Hartog, G., Schepp, R. M., Kaaijk, P., van Vliet, J., Helm, K., Smits, G., Wijmenga-Monsuur, A., Verberk, J. D. M., van Boven, M., van Binnendijk, R. S., de Melker, H. E., Mollema, L. and van der Klis, F. R. M. (2021), 'Nationwide seroprevalence of SARS-CoV-2 and identification of risk factors in the general population of the netherlands during the first epidemic wave', *Journal of Epidemiology & Community Health* **75**(6), 489–495.

Walensky, R. P., Walke, H. T. and Fauci, A. S. (2021), 'SARS-CoV-2 variants of concern in the United States—Challenges and opportunities', *JAMA* **325**(11), 1037–1038.

Wang, X., Gao, W., Cui, S., Zhang, Y., Zheng, K., Ke, J., Lv, J., Yu, C., Sun, D., Wang, Q. et al. (2020), 'A population-based seroprevalence survey of severe acute respiratory syndrome coronavirus 2 infection in Beijing, China', medRxiv.

Ward, H., Cooke, G., Atchison, C. J., Whitaker, M., Elliott, J., Moshe, M., Brown, J. C., Flower, B., Daunt, A., Ainslie, K. E. et al. (2020), 'Declining prevalence of antibody positivity to SARS-CoV-2: a community study of 365,000 adults', medRxiv.

Wingfield-Hayes, R. (2020), 'Coronavirus: Japan's low testing rate raises questions', *https://www.bbc.com/news/world-asia-52466834* (*accessed May 2, 2021*) .

Wu, J. T., Leung, K., Bushman, M., Kishore, N., Niehus, R., de Salazar, P. M., Cowling, B. J., Lipsitch, M. and Leung, G. M. (2020), 'Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China', *Nature Medicine* **26**(4), 506–510.

Zimmermann, P. and Curtis, N. (2021), 'Why is COVID-19 less severe in children? a review of the proposed mechanisms underlying the age-related difference in severity of sars-cov-2 infections', *Archives of disease in childhood* **106**(5), 429–439.

6 Appendix

6.1 Excluded studies

Table 3 lists the studies identified by Chen et al. (2021) as grade A or B general-population-based studies that we were unable to include in our analysis. Note that, while we were unable to obtain the relevant number of deaths for Mahajan et al. (2021*b*), (deaths for the non-congregate population in Connecticut), Mahajan et al. (2021*a*) derive an estimate for the IFR of 0.95% (90% CI, 0.63%-1.90%).

6.2 Details on mortality and covariate data

- For Barchuk et al. (2020), we used excess death numbers as reported by Kobak (2021). Russian official statistics appear to underestimate the true number of fatalities by a substantial factor (Karlinsky and Kobak, 2021). Data for the proportion aged over 65 is from Barchuk et al. (2020) (see Table A3 "KOUZh-2018"). GDP value is from value listed for Russian Federation from World Bank's World Development Indicators (WDI) data (see:

Author	Location	Reason for exclusion
Hallal et al.	Brazil	new info. about the antibody test
Bendavid et al.	Santa Clara County, CA, USA	"stratified non-prob." sampling method
Rosenberg et al.	New york State, USA	"convenience" sampling method
McLaughlin et al.	Blaine County, ID, USA	"self-referral" sampling method
Appa et al.	Bolinas, CA, USA	"self-referral" sampling method
Richard et al.	Geneva, Switzerland	duplicate
Murhekar et al. (2)	India	duplicate
Mahajan et al.	Connecticut, USA	death data not found
Malani et al.	Mumbai, India	death data not found
Ling et al.	Wuhan, China	death data not found
Tess et al.	Six districts of São Paulo, Brazil	death data not found
Poustchi et al.	Iran (18 cities)	death data not found
Nisar et al.	Two neighborhoods of Karachi, Pakistan	death data not found
Shakiba et al.	Guilan Province, Iran	death data not found
Alemu et al.	Addis Ababa, Ethiopia	death data not found
Majiya et al.	Nigeria	95% interval not provided
Naranbhai et al.	Chelsea, MA, US	95% interval not provided
Wang et al.	Beijing, China	95% interval not provided
Borges et al.	Sergipe (10 cities), Brazil	95% interval not provided
Gudbjartsson et al.	Iceland	sampling dates not specified

Table 3: List of excluded studies and reason for exclusion.

`https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD` (last updated 2021-04-26)).

- For Biggs et al. (2020), the number of deaths for DeKalb, and Fulton counties was obtained from the county-level COVID-19 dataset curated by the New York Times available at: `github.com/nytimes/covid-19-data` (accessed on April 28, 2021). Data for the proportion aged over 65 is from `https://github.com/JieYingWu/COVID-19_US_County-level_Summaries/blob/master/data/README.md` (accessed on April 29, 2021). GDP value is from value listed for the state of Georgia, USA, for 2019 at `https://en.wikipedia.org/wiki/List_of_U.S._states_and_territories_by_GDP_per_capita` (accessed on April 29, 2021).
- For Bruckner et al. (2021), we obtained number of cumulative deaths for Orange County from Orange County Public Works (as referenced by Bruckner et al. (2021)) at: `data-ocpw.opendata.arcgis.com/datasets/2ec9342ffc814cf58161b1cca57365fd_0` (accessed on April 28, 2021). Data for the proportion aged over 65 is from `https://github.com/JieYingWu/COVID-19_US_County-level_Summaries/blob/master/data/README.md` (accessed on April 29, 2021). GDP value is from value listed for the state of California, USA, for 2019 at `https://en.wikipedia.org/wiki/List_of_U.S._states_`

and_territories_by_GDP_per_capita (accessed on April 29, 2021).

- For Carrat et al. (2020), we only consider Ile-de-France phase of the study (see Supp. Table 1 for sampling dates). Data for the number of deaths for Ile-de-France was obtained from the Corona Data Scraper website (coronadatascraper.com/; accessed on April 28, 2021) that pulls COVID-19 data from verified sources on national and local levels. Data for the proportion aged over 65 is from https://contrevues.paris/ile-de-france-comment-le-vieillissement-de-la-population-impacte-lhabitat/ (accessed on April 29, 2021). GDP value is from value listed for Russian Federation from World Bank's World Development Indicators (WDI) data (see: https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD (last updated 2021-04-26)).
- For Statistics Jersey (2020), data for the number of deaths was obtained from the Government of Jersey website (https://www.gov.je/datasets/listopendata?listname=COVID19DeathsClassification; accessed on April 28, 2021) summing both "probable COVID-19" deaths and "laboratory proven" COVID-19 deaths. Data for the proportion aged over 65 is from https://www.indexmundi.com/jersey/demographics_profile.html (accessed on April 29, 2021). GDP value is from value listed for United Kingdom from World Bank's World Development Indicators (WDI) data (see: https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD (last updated 2021-04-26)).
- For Murhekar et al. (2020*a*), we obtained the number of cumulative deaths for India, from the Our World in Data COVID-19 dataset available at: ourworldindata.org/coronavirus/country/india (accessed on April 28, 2021). We multiplied the number recorded for 14 days after the end of the sampling window of 12,573 by a factor of 3.56 (based on Mukherjee et al. (2021)'s estimated underreporting factor for Delhi) in order to account for potential underreporting. As such, our interval is relatively wide and reflects the uncertainty in the true number of deaths: [4,172, 44,760]. Data for the proportion aged over 65 is from World Bank's World Development Indicators (WDI) data (see https://data.worldbank.org/indicator/SP.POP.65UP.TO.ZS (last updated 2021-03-19)). GDP value is from value listed for India from World Bank's World Development Indicators (WDI) data (see: https://data.worldbank.

org/indicator/NY.GDP.PCAP.PP.CD (last updated 2021-04-26)).

- For Office of National Statistics (2020), we obtained the number of cumulative deaths for England from the UK coronavirus dashboard (https://coronavirus.data.gov.uk/details/deaths?areaType=nation&areaName=England; accessed on April 29, 2021). Seroprevalence numbers were obtained from Table 3a of https://www.ons.gov.uk/file?uri=/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/coronaviruscovid19infectionsurveydata/2020/previous/v26/covid19infectionsurveydatasets20201002.xlsx ; (accessed on April 28, 2021). Data for the proportion aged over 65 is for England from 2019 as listed by LG Inform (see https://tinyurl.com/4vhrb2uu; accessed on April 29, 2021). GDP value is from value listed for United Kingdom from World Bank's World Development Indicators (WDI) data (see: https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD (last updated 2021-04-26)).
- For Petersen et al. (2020), information on deaths for the Faroe Islands was obtained from corona.fo/hagtol, the government information website concerning COVID19 in the Faroe Islands. GDP value is from value listed for Denmark from World Bank's World Development Indicators (WDI) data (see: https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD (last updated 2021-04-26)). Data for the proportion aged over 65 is from Index mundi (see https://tinyurl.com/bb2pwrx6; accessed on April 29, 2021) which cites the CIA World Factbook as a source.
- For Pollán et al. (2020), data for the number of deaths was obtained from Wikipedia (en.wikipedia.org/wiki/COVID-19_pandemic_in_Spain; accessed on April 28, 2021) which sourced the information from the Centro Nacional de Epidemiología (cnecovid.isciii.es/covid19/). Note that, the number of deaths of for 2020-05-11 of 26,920 (14 days after the start of the sampling window), is actually higher than the number of deaths for 2020-05-25 of 26,834 (14 days after the end of the sampling window). This may be due to a reporting issue which is noted by Wikipedia: "Figures for 2020-05-24 to 2020-06-17 include

corrections in the validation of past data from several autonomous communities as a result of the transition to a new surveillance methodology implemented from 2020-05-11." We define the interval as ranging from the lowest value to the highest value, [26,834, 26,920], as listed in Table 2. GDP value is from value listed for Spain from World Bank's World Development Indicators (WDI) data (see: `https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD` (last updated 2021-04-26)). Data for the proportion aged over 65 is from World Bank's World Development Indicators (WDI) data (see `https://data.worldbank.org/indicator/SP.POP.65UP.TO.ZS` (last updated 2021-03-19)).

- For Samore et al. (2020), data for the number of deaths for the counties of Utah county, Salt Lake county, Davis county, and Summit county, was obtained from the county-level COVID-19 dataset curated by the New York Times available at: `github.com/nytimes/covid-19-data` (accessed on April 28, 2021). Data for the proportion aged over 65 is from `https://github.com/JieYingWu/COVID-19_US_County-level_Summaries/blob/master/data/README.md` (accessed on April 29, 2021). GDP value is from value listed for the state of Utah, USA, for 2019 at `https://en.wikipedia.org/wiki/List_of_U.S._states_and_territories_by_GDP_per_capita` (accessed on April 29, 2021).
- For Santos-Hövener et al. (2020), data for the number of deaths for Kupferzell, Germany was obtained directly from Santos-Hövener et al. (2020) which cites the Robert Koch Institute. Despite efforts, no publicly available dataset was found which could confirm these numbers specific these numbers. Data for the proportion aged over 65 is from: `https://ugeo.urbistat.com/AdminStat/en/de/demografia/eta/kupferzell/20172564/4` (accessed on April 29, 2021). GDP value is from value listed for Germany from World Bank's World Development Indicators (WDI) data (see: `https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD` (last updated 2021-04-26)).
- For Sharma et al. (2020), infection rate estimates are based on survey data from round 1 of the study (August 1-7). Data for the number of deaths for Delhi was obtained from Wikipedia (`en.wikipedia.org/wiki/COVID-19_pandemic_in_Delhi`; accessed on April 28, 2021) which sourced the information from the Delhi State Health Bulletin (`https://delhifightscorona.in/`). We multi-

plied the number recorded for 14 days after the end of the sampling window of 4,270 by a factor of 6.3 (based on Mukherjee et al. (2021)'s estimated underreporting factor) in order to account for potential underreporting. As such, our interval is relatively wide and reflects the uncertainty in the true number of deaths: [4,188, 26,901]. Data for the proportion aged over 65 is from Statistics Times for 2011 (`http://statisticstimes.com/demographics/india/delhi-population.php`; accessed on April 29, 2021). GDP value is from value listed for India from World Bank's World Development Indicators (WDI) data (see: `https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD` (last updated 2021-04-26)). Note that we use 18.9 million as the estimated population of Delhi, the number cited by both Sharma et al. (2020) and Bhattacharyya et al. (2021), while Chen et al. (2021) use a much different number: 30,290,936 (see Table S14 in Supplementary appendix 2 of Chen et al. (2021)).

- Snoeck et al. (2020) "recruited a representative sample of the Luxembourgish population" between April 16th and May 5th, and obtained a 95% CI of [1.23%, 2.77%]. Two different 95% CIs, obtained with and without adjustment for age, gender and canton are provided in the paper: [1.23%; 2.67%] and [1.34%; 2.77%]. As such, we record [1.23%; 2.77%] for our IR interval. Data for the number of deaths was obtained from the Our World in Data COVID-19 dataset available at: `ourworldindata.org/coronavirus/country/luxembourg` (accessed on April 28, 2021). GDP value is from value listed for Lusembourg from World Bank's World Development Indicators (WDI) data (see: `https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD` (last updated 2021-04-26)). Data for the proportion aged over 65 is from World Bank's World Development Indicators (WDI) data (see `https://data.worldbank.org/indicator/SP.POP.65UP.TO.ZS` (last updated 2021-03-19)).
- For Sood et al. (2020), data for the number of deaths for Los Angeles county, CA was obtained from the government of LA county COVID-19 dashboard (`dashboard.publichealth.lacounty.gov/covid19_surveillance_dashboard/`; accessed on April 28, 2021). Sood et al. (2020) notes that "Residents of Los Angeles County, California, within a 15-mile (24 km) radius of the testing site were eligible for participation." Data for

the proportion aged over 65 is from https://github.com/JieYingWu/COVID-19_US_County-level_Summaries/blob/master/data/README.md (accessed on April 29, 2021). GDP value is from value listed for the state of California, USA, for 2019 at https://en.wikipedia.org/wiki/List_of_U.S._states_and_territories_by_GDP_per_capita (accessed on April 29, 2021).

- For Streeck et al. (2020), the number of deaths for Gangelt, Kreis Heinsberg, Germany, is directly noted in the Streeck et al. (2020) article. No publicly available dataset was found, however we did find information on the Gangelt municipal bulletin (see www.gangelt.de/news/226-erster-corona-fall-in-nrw; accessed on April 28, 2021) and an investigative report (see https://medwatch.de/2020/11/26/die-ungezaehlten-todesfaelle-aus-gangelt/; accessed on May 20, 2021) which suggests that 8 to 12 as a reasonable range. Data for the proportion aged over 65 is from Figure S1 of Streeck et al. (2020). GDP value is from value listed for Germany from World Bank's World Development Indicators (WDI) data (see: https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD (last updated 2021-04-26)).
- For Stringhini et al. (2020), mortality data for the canton of Geneva were obtained from an excel file made publicly available by a Swiss government website at: ge.ch/document/covid-19-donnees-completes-debut-pandemie (accessed on April 28, 2021). Data for the proportion aged over 65 is from https://www.bfs.admin.ch/bfs/en/home/statistics/regional-statistics/regional-portraits-key-figures/cantons/geneva.html. GDP value is from value listed for Switzerland from World Bank's World Development Indicators (WDI) data (see: https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD (last updated 2021-04-26)).
- For Vos et al. (2021), mortality data for the Netherlands was obtained from the Our World in Data COVID-19 dataset available at: ourworldindata.org/coronavirus/country/netherlands (accessed on April 28, 2021). GDP value is from value listed for Netherlands from World Bank's World Development Indicators (WDI) data (see: https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD (last updated 2021-04-26)). Data for the proportion aged over 65 is also from World Bank's World Development Indicators (WDI) data (see

https://data.worldbank.org/indicator/SP.POP.65UP.TO.ZS (last updated 2021-03-19)).

- For Ward et al. (2020), infection rate estimates are based on survey data from from the first survey (20 June - 13 July). We obtained the number of cumulative deaths for England, from we obtained the number of cumulative deaths for England from the UK coronavirus dashboard (https://coronavirus.data.gov.uk/details/deaths?areaType=nation&areaName=England; accessed on April 29, 2021). Data for the proportion aged over 65 is for England from 2019 as listed by LG Inform (see https://tinyurl.com/4vhrb2uu; accessed on April 29, 2021). GDP value is from value listed for United Kingdom from World Bank's World Development Indicators (WDI) data (see: https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD (last updated 2021-04-26)).

6.3 Sensitivity analysis

As a sensitivity analysis, we repeated the entire analysis with an alternative set of priors. For this alternative analysis, we used: $\mathrm{g}^{-1}(\theta_0) \sim Uniform(0,1)$; $\mathrm{g}^{-1}(\beta) \sim Uniform(0,1)$; $\theta_1 \sim \mathcal{N}(0,100)$; $\theta_2 \sim \mathcal{N}(0,100)$; $\sigma \sim$ half-$\mathcal{N}(0,100)$ and $\tau \sim$ half-$\mathcal{N}(0,100)$. The results are plotted in Figure 4. We also conducted a leave-one-out sensitivity analysis whereby, for each of the $K = 18$ individual seroprevalence studies, we removed the data associated with the individual study and repeated our analysis (Higgins, 2008, Iyengar and Greenhouse, 2009). Results are listed in Table 4 and suggest that our estimates are somewhat sensitive to the data associated with Snoeck et al. (2020).

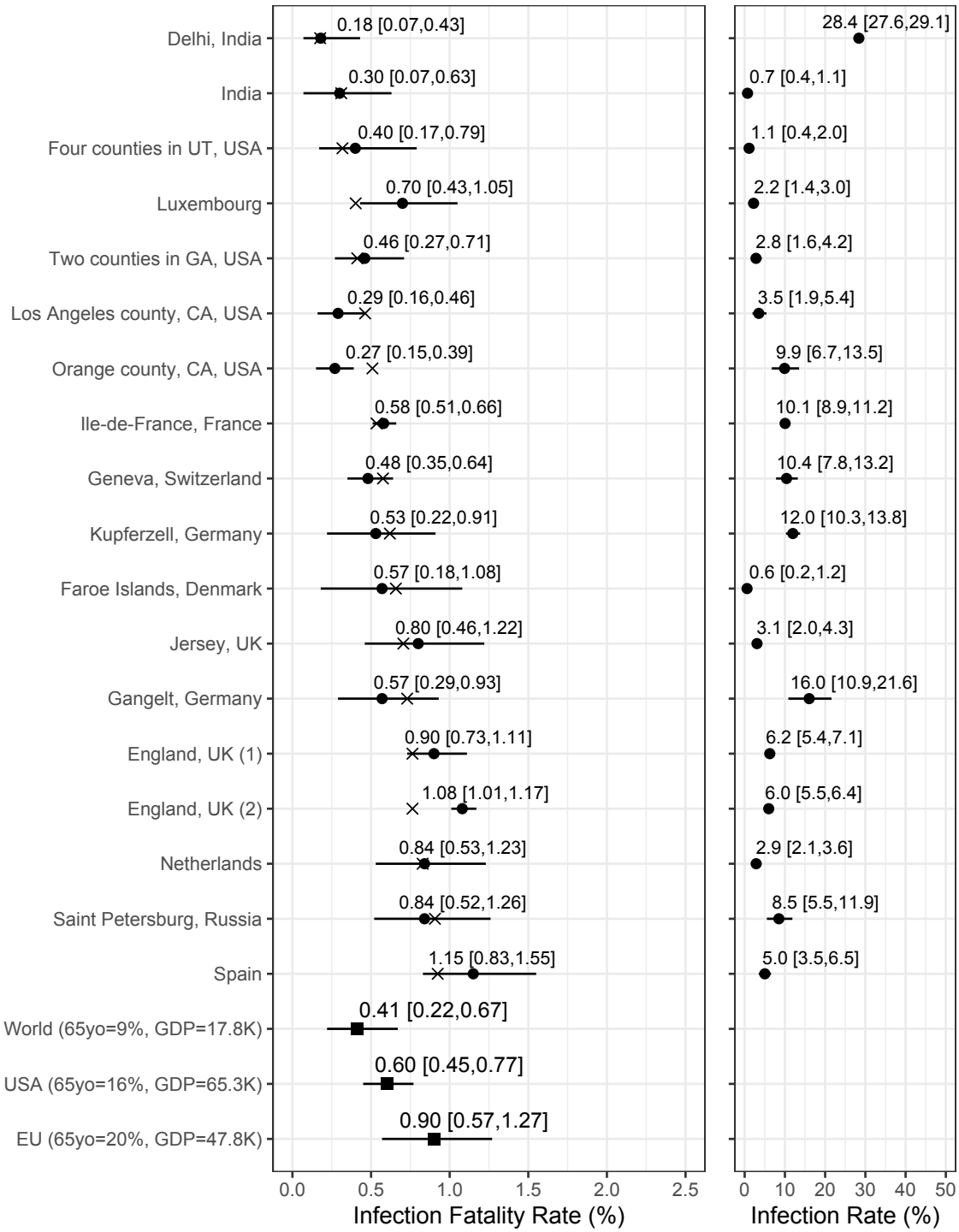

Figure 4: Analysis results with alternative priors - Posterior median estimates for the IR_k and IFR_k variables (for $k = 1, \ldots, 18$) with 95% HPD CrIs. Studies are listed from top to bottom according to increasing fitted values (these values are indicated by $\times$). Also plotted, under the labels "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 \$65,297 (the USA values), and of 20% and of \$47,828 (the EU values).

Study excluded	$IFR(z_* = 9)$	95% CrI	θ_1	95% CrI	θ_2	95% CrI	τ	95% CrI
Barchuk et al.	0.41	[0.20,0.68]	0.61	[0.13,1.07]	-0.26	[-0.73,0.23]	0.43	[0.21,0.72]
Biggs et al.	0.39	[0.19,0.64]	0.62	[0.18,1.08]	-0.25	[-0.69,0.20]	0.43	[0.22,0.72]
Bruckner et al.	0.39	[0.22,0.61]	0.57	[0.20,0.93]	-0.14	[-0.50,0.24]	0.30	[0.11,0.56]
Carrat et al.	0.39	[0.19,0.66]	0.60	[0.13,1.04]	-0.24	[-0.70,0.20]	0.44	[0.21,0.73]
Government of Jersey	0.39	[0.20,0.67]	0.59	[0.13,1.02]	-0.23	[-0.66,0.23]	0.43	[0.21,0.73]
Murhekar et al. (1)	0.44	[0.21,0.78]	0.52	[0.14,0.88]	-0.23	[-0.57,0.14]	0.41	[0.21,0.69]
Office of National Stat	0.39	[0.20,0.64]	0.57	[0.12,1.03]	-0.22	[-0.63,0.27]	0.44	[0.23,0.75]
Petersen et al.	0.39	[0.21,0.65]	0.62	[0.17,1.02]	-0.25	[-0.66,0.18]	0.40	[0.20,0.68]
Pollan et al.	0.39	[0.20,0.64]	0.52	[0.05,0.99]	-0.19	[-0.64,0.29]	0.43	[0.21,0.72]
Samore et al.	0.38	[0.20,0.62]	0.66	[0.20,1.07]	-0.29	[-0.72,0.13]	0.41	[0.21,0.68]
Santos-Hovener et al.	0.40	[0.21,0.64]	0.60	[0.17,1.02]	-0.24	[-0.66,0.20]	0.42	[0.20,0.70]
Sharma et al.	0.35	[0.11,0.72]	0.45	[0.09,0.81]	-0.16	[-0.51,0.20]	0.42	[0.20,0.70]
Snoeck et al.	0.45	[0.25,0.69]	0.82	[0.42,1.18]	-0.53	[-0.89,-0.11]	0.29	[0.10,0.54]
Sood et al.	0.40	[0.20,0.62]	0.55	[0.13,0.97]	-0.16	[-0.57,0.29]	0.38	[0.18,0.67]
Streeck et al.	0.40	[0.21,0.68]	0.63	[0.18,1.04]	-0.25	[-0.69,0.18]	0.41	[0.20,0.70]
Stringhini et al.	0.39	[0.20,0.65]	0.60	[0.14,1.03]	-0.21	[-0.64,0.26]	0.43	[0.20,0.73]
Vos et al.	0.40	[0.20,0.64]	0.58	[0.14,1.01]	-0.24	[-0.65,0.24]	0.44	[0.21,0.74]
Ward et al.	0.39	[0.20,0.65]	0.55	[0.07,0.96]	-0.21	[-0.65,0.25]	0.42	[0.20,0.73]

Table 4: Estimates obtained from the leave-one-out sensitivity analysis. For each of the $K = 18$ individual seroprevalence studies, we removed the data associated with the individual study and repeated our analysis.

6.4 MCMC details and R code

Note that, in order to improve the MCMC mixing, we replace the binomial distribution for CC_k as described in (2), with

$$CC_k \sim Binom(T_k, IR_k), \tag{6}$$

for $k = 1, \ldots, K$. For any sufficiently large P_k, this simplification will make little to no difference. Then, since the distributions of C_k and $D_k|C_k$ are both binomials (see (2) and (3)), we have that unconditionally:

$$D_k \quad \sim Binom(P_k, IFR_k \times IR_k). \tag{7}$$

The following JAGS-code was used in R for the analysis :

```
library("rjags")
metaIFR <- "model {
# Priors:
icloglog_theta0 ~ dbeta(0.3, 30);
icloglog_beta  ~ dbeta(1, 3);
theta0 <- log(-log(1-icloglog_theta0));
beta  <- log(-log(1-icloglog_beta));
inv.var_sig   <- (1/sigma)^2 ;
inv.var_tau   <- (1/tau)^2 ;
```

```
sigma     ~ dnorm(0, 1/10) T(0,);
tau     ~ dnorm(0, 1/10) T(0,);
theta1 ~ dnorm(0, 1/10);
theta2 ~ dnorm(0, 1/10);

# Likelihood:
for(k in 1:K){
cc[k] ~ dbin(ir[k], tests[k]);
    censor.index[k] ~ dinterval(deaths[k], c(deaths_lower[k], deaths_upper[k]))
deaths[k] ~ dbin(ifr[k]*ir[k], pop[k]);
cloglog(ir[k]) <- cloglog_ir[k];
cloglog(ifr[k]) <- cloglog_ifr[k];
cloglog_ir[k] ~ dnorm(beta, inv.var_sig);
cloglog_ifr[k] ~ dnorm(theta0 + theta1*Z1[k] + theta2*Z2[k], inv.var_tau);
  }}"
```