

Original Paper

**Cite this article:** Campbell H, Gustafson P (2021). 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. *Epidemiology and Infection* **149**, e243, 1–14. <https://doi.org/10.1017/S0950268821002405>

Received: 4 September 2021

Revised: 25 October 2021

Accepted: 26 October 2021

**Key words:**

COVID-19; evidence synthesis; Bayesian inference; infection fatality rate

**Author for correspondence:**

Harlan Campbell,

E-mail: [harlan.campbell@stat.ubc.ca](mailto:harlan.campbell@stat.ubc.ca)

# 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  and Paul Gustafson 

Department of Statistics, University of British Columbia, Vancouver, British Columbia, Canada

## Abstract

Estimating the coronavirus disease-2019 (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 (CrI) 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% CrI 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%.

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 characterise it.

Siddhartha Mukherjee, *The New Yorker*

22 February 2021

## 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 severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and the ongoing coronavirus disease-2019 (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 [1]. In particular, Ioannidis [2], Levin *et al.* [3], Brazeau *et al.* [4] and O’Driscoll *et al.* [5] 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.* [4] and O’Driscoll *et al.* [5] are done using rather complex Bayesian models which rely on numerous detailed assumptions. For instance, Brazeau *et al.* [4] 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.* [5] 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.

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

In contrast, the analyses of Ioannidis [2] and Levin *et al.* [3] are decidedly more simple. For each seroprevalence study under consideration, Ioannidis [2] 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 [2] 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 [2]'s estimates.

The analysis of Levin *et al.* [3] is based on a standard frequentist random-effects meta-analysis model. For each age-group and seroprevalence study under consideration, Levin *et al.* [3] 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 'methods').

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 [6]). We will apply the method in 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 [7, 8]. 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.* [9] and Arora *et al.* [10] 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 [11]. We will review the data and how it was obtained following a review of the methods.

## Methods

### Bayesian Model for evidence synthesis

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

$$CC_k \sim \text{Binom}\left(T_k, \frac{C_k}{P_k}\right), \quad (1)$$

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

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

We wish to emphasise 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.* [12] (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% confidence interval (CI) for the number of infections (95% CI for  $C = [1551, 2389]$ ). Doing so, they obtain a relatively narrow 95% CI for the IFR: [0.29%, 0.45%] ( $= [7/1,551, 7/2389]$ ). 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.* [3] 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  $k = 1, \dots, K$ :

$$g(IFR_k) \sim \text{Normal}(\theta_0 + \theta_1 Z_{1k} + \theta_2 Z_{2k}, \tau^2), \quad (4)$$

and

$$g(IR_k) \sim \text{Normal}(\beta, \sigma^2), \quad (5)$$

where  $\theta_0$  represents the mean  $g(IFR)$ ,  $\tau^2$  represents between-group IFR heterogeneity,  $\beta$  represents the mean  $g(\text{infection rate})$ ,  $\sigma^2$  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 IFR 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 these as the centred and scaled logarithm of the proportion of the population aged over 65 years ( $65\text{ yo}_k$ ) and of the GDP [at purchasing power parity (PPP)] per capita ( $\text{GDP}_k$ ), 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.* [13] and Kümmerer *et al.* [14]) are set accordingly:  $g^{-1}(\theta_0) \sim \text{Beta}(0.3, 3)$ ;  $g^{-1}(\beta) \sim \text{Beta}(1, 30)$ ;  $\theta_1 \sim \text{Normal}(0, 10)$ ;  $\theta_2 \sim \text{Normal}(0, 10)$ ;  $\sigma \sim \text{half-Normal}(0, 10)$ ; and  $\tau \sim \text{half-Normal}(0, 10)$ . 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 [15, 16]. In the Supplementary Material, we show results obtained with an alternative set of priors as a sensitivity analysis.

### 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 [2]) 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.* [14].) 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 [2] 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 penalised. Ioannidis [2] notes that the strategy of 'weighting the study-specific IFRs 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 recognise that this is particularly difficult when so many studies may misrepresent the precision of their estimates [53, 54].

### 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.* [24]), but can also be several weeks or months for others (e.g. 135 days for Ward *et al.* [34]). 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 [2] 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 [55] 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.* [3], 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 the 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 [56, 57].

## The data

### Seroprevalence data

As the COVID-19 pandemic has progressed, a rapidly increasing number of SARS-CoV-2 seroprevalence studies have been conducted worldwide [10]. 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 [53]. 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 [58]. Based on the reviews of Chen *et al.* [9] and of Arora *et al.* [10], 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

**Table 1.** Seroprevalence studies selected for the analysis based on the list compiled by Chen *et al.* [9] (listed in alphabetical order of authors), with the 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.

Authors	Location	Sampling (mm/dd)	IR interval (%)	In both analyses
Barchuk <i>et al.</i> [17]	Saint Petersburg, Russia	05/27–06/26	(5.60, 12.90)	yes
Biggs <i>et al.</i> [18]	DeKalb and Fulton, GA, USA (A)	04/28–05/03	(1.40, 4.50)	yes*
Bruckner <i>et al.</i> [19]	Orange County, CA, USA	07/10–08/16	(8.10, 15.50)	yes
Carrat <i>et al.</i> [20]	Ile-de-France, France	05/04–06/14	(8.90, 11.30)	
Mahajan <i>et al.</i> [21]	Connecticut, USA	06/10–07/29	(1.70, 6.30)	
Murhekar <i>et al.</i> (A) [22]	India	05/11–06/04	(0.34, 1.13)	yes
Office of National Stat [23]	England, UK (A)	04/26–09/08	(5.40, 7.10)	
Petersen <i>et al.</i> [24]	Faroe Islands, Denmark	04/27–05/01	(0.10, 1.20)	yes
Pollan <i>et al.</i> [25]	Spain	04/27–05/11	(3.30, 6.60)	yes
Samore <i>et al.</i> [26]	Four counties in UT, USA	05/04–06/30	(0.10, 1.60)	
Santos-Hovener <i>et al.</i> [27]	Kupferzell, Germany	05/20–06/09	(10.40, 14.00)	yes
Sharma <i>et al.</i> [28]	Delhi, India	08/01–08/07	(27.65, 29.14)	yes
Snoeck <i>et al.</i> [29]	Luxembourg	04/15–05/05	(1.23, 2.77)	
Sood <i>et al.</i> [30]	Los Angeles County, CA, USA	04/10–04/14	(2.52, 7.07)	
Statistics Jersey (A) [31]	Jersey, UK (A)	04/29–05/05	(1.80, 4.40)	yes*
Streeck <i>et al.</i> [12]	Gangelt, Germany	03/31–04/06	(12.31, 24.40)	yes
Stringhini <i>et al.</i> (A) [32]	Geneva, Switzerland (A)	04/06–05/09	(8.15, 13.95)	yes*
Vos <i>et al.</i> [33]	Netherlands	03/31–05/11	(2.10, 3.70)	yes
Ward <i>et al.</i> [34]	England, UK (B)	06/20–07/13	(5.78, 6.14)	yes

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.

Chen *et al.* [9] reviewed the literature for articles published between 1 December 2019 and 22 December 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 [59]. In total, Chen *et al.* [9] identified 38 articles which considered a sample based on a ‘general population’ and which obtained a study quality grade of A or B (see the full list in Supplementary Table S8 of Chen *et al.* [9]). 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.* [10]) ( $n = 8$ ), 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.* [10] 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 [53, 60]. For analysis, we consider the 45 studies listed on serotracker.com/data (as of 5 June 2021), that are categorised as having a ‘low risk of bias’ and are categorised 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.* [10]) ( $n = 1$ ), 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.

For each of the seroprevalence studies included in each of the two analysis sets, we recorded the 95% uncertainty interval for the infection rate as reported in the study 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 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

**Table 2.** Seroprevalence studies selected for the analysis based on the list compiled by Serotracker (listed in alphabetical order of authors), with the 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.

Authors	Location	Sampling (mm/dd)	IR interval (%)	In both analyses
Álvarez-Antonio <i>et al.</i> [35]	Iquitos, Peru	07/13–07/18	(67.00, 73.00)	
Bajema <i>et al.</i> [36]	DeKalb and Fulton, GA, USA (B)	04/28–05/03	(1.49, 6.67)	yes*
Barchuk <i>et al.</i> [17]	Saint Petersburg, Russia	05/27–06/26	(5.60, 12.90)	yes
Bruckner <i>et al.</i> [19]	Orange County, CA, USA	07/10–08/16	(8.10, 15.50)	yes
Chan <i>et al.</i> [37]	Rhode Island, USA	05/05–05/22	(1.00, 6.20)	
Gov. of Andorra [38]	Andorra	05/04–05/28	(10.50, 11.50)	
Kar <i>et al.</i> [39]	Puducherry District, India	08/11–08/16	(3.50, 6.40)	
Khalagi <i>et al.</i> [40]	Iran	08/03–10/31	(13.30, 15.20)	
Malani <i>et al.</i> [41]	Tamil Nadu, India	10/19–11/30	(30.40, 32.80)	
Melotti <i>et al.</i> [42]	Gardena Valley, Italy	05/26–06/08	(25.20, 28.60)	
MoHol [43]	Israel	06/28–09/17	(5.30, 5.60)	
Murhekar <i>et al.</i> (A) [22]	India	05/11–06/04	(0.34, 1.13)	yes
Nawa <i>et al.</i> [44]	Utsunomiya City, Japan	06/14–07/05	(0.08, 2.28)	
Pagani <i>et al.</i> [45]	Castiglione d’Adda, Italy	05/18–06/07	(17.20, 29.10)	
Petersen <i>et al.</i> [24]	Faroe Islands, Denmark	04/27–05/01	(0.10, 1.20)	yes
Pollan <i>et al.</i> [25]	Spain	04/27–05/11	(3.30, 6.60)	yes
Radon <i>et al.</i> [46]	Munich, Germany	11/02–01/31	(2.90, 4.30)	
Reyes-Vega <i>et al.</i> [47]	Lima, Peru	06/28–07/09	(22.50, 28.20)	
Richard <i>et al.</i> [48]	Geneva, Switzerland (B)	04/06–06/30	(6.80, 8.90)	yes*
Santos-Hovener <i>et al.</i> [27]	Kupferzell, Germany	05/20–06/09	(10.40, 14.00)	yes
Selvaraju <i>et al.</i> [49]	Chennai, Tamil Nadu, India	07/01–07/31	(14.80, 22.60)	
Sharma <i>et al.</i> [28]	Delhi, India	08/01–08/07	(27.65, 29.14)	yes
Statistics Jersey (B) [50]	Jersey, UK (B)	06/21–06/27	(2.80, 5.20)	yes*
Streeck <i>et al.</i> [12]	Gangelt, Germany	03/31–04/06	(12.31, 24.40)	yes
Vos <i>et al.</i> [33]	Netherlands	03/31–05/11	(2.10, 3.70)	yes
Ward <i>et al.</i> [34]	England, UK (B)	06/20–07/13	(5.78, 6.14)	yes
Warszawski <i>et al.</i> [51]	Metropolitan France	05/02–06/02	(3.90, 5.00)	
Yoshiyama <i>et al.</i> [52]	Tokyo, Japan	06/01–06/07	(0.01, 0.37)	

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 ( $65\text{ yo}_k$ ) and the GDP (PPP) per capita ( $GDP_k$ ); see Supplemental Material for details and data sources.

### Mortality Data

Mortality data were 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 [61]. 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 [62]– 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 [62]’s analysis. There was evidence of substantial under-reporting of COVID-19 deaths in India [63, 64] while little could be

**Table 3.** The Chen *et al.* based dataset required for the Bayesian evidence synthesis model

Location	$P_k$	$D_k$ lower	$D_k$ upper	$T_k$	$CC_k$	65yo <sub>k</sub> (%)	GDP <sub>k</sub> (\$)
Saint Petersburg, Russia	5 351 935	1596	5128	233	21	18	30 144
DeKalb and Fulton, GA, USA (A)	1 823 234	122	136	419	11	12	58 933
Orange County, CA, USA	3 010 232	512	839	285	33	15	79 287
Ile-de-France, France	12 213 447	6816	7427	2414	243	14	82 574
Connecticut, USA	2 837 877	1136	1179	251	9	22	80 729
India	1 366 417 750	4172	45 766	1632	11	6	4735
England, UK (A)	56 550 000	25 975	30 709	3100	193	18	43 310
Faroe Islands, Denmark	52 154	0	0	592	3	17	60 421
Spain	47 351 567	17 710	17 767	643	31	20	42 362
Four counties in UT, USA	2 194 298	40	97	393	2	10	60 050
Kupferzell, Germany	6247	1	3	1263	153	16	63 885
Delhi, India	19 800 000	4188	26 901	13 966	3965	4	12 817
Luxembourg	632 275	47	58	1214	23	14	122 166
Los Angeles County, CA, USA	10 039 107	529	606	316	14	14	79 287
Jersey, UK (A)	107 800	17	18	648	19	17	48 365
Gangelt, Germany	12 597	8	8	153	27	18	53 751
Geneva, Switzerland (A)	504 128	122	155	442	48	16	68 964
Netherlands	17 344 874	2621	4186	1652	47	20	59 685
England, UK (B)	56 550 000	29 526	30 038	10 635	634	18	43 310

gathered about the reliability of official mortality data for Pakistan, Palestine,<sup>1</sup> Ethiopia,<sup>2</sup> and China (but do see [65] and [66]). As such, we excluded the Qutob *et al.* [67] ('Palestinian population residing in the West Bank') and He *et al.* [68] ('Wuhan, China') studies from the Serotracker-based analysis and excluded the Alemu *et al.* [69] ('Addis Ababa, Ethiopia') and the Nisar *et al.* [70] ('Two neighborhoods of Karachi, Pakistan') studies from the Chen *et al.*-based analysis. For India, Mukherjee *et al.* [71] and Purkayastha *et al.* [72] estimate UCRs for the entire country as well as for each individual Indian state and union territory. We used these UCRs to adjust the upper bound of the  $D_k$  interval for each of the Indian seroprevalence studies (see Supplementary Material for details).

There are two countries represented within our data that were identified by Karlinsky and Kobak [62] 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 Barchuk *et al.* [17] ('Saint Petersburg, Russia') and for the Khalagi *et al.* [40] ('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).

<sup>1</sup>Official regional death numbers for Palestine are available from the Palestinian government dashboard (see <https://corona.ps/details>; accessed July 28, 2021).

<sup>2</sup>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/](http://web.archive.org/web/2020*/https://www.covid19.et/covid-19/) and the Twitter account: [https://twitter.com/Harun\\_Asefa/status/125906983287793280](https://twitter.com/Harun_Asefa/status/125906983287793280); accessed August 4, 2021).

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.* [73]. 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. Modelling 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.* [74].

## Results

The Model as described in the Methods Section, was fit to the two datasets described above. We fit the model using JAGS (just another Gibbs sampler) [75], with five independent chains, each with two 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 CrIs obtained for  $IFR_k$ , for  $k$  in 1, ...,  $K$ , respectively; see Supplementary Figures S1 and S2 for  $IFR_k$  in the Supplementary Material. In these figures, the seroprevalence studies are listed in order of their 'fitted' IFR

**Table 4.** The Serotracker-based dataset required for the Bayesian evidence synthesis model

Location	$P_k$	$D_k$		$T_k$	$CC_k$	65yo <sub>k</sub> (%)	GDP <sub>k</sub> (\$)
		lower	upper				
Iquitos, Peru	467 000	1581	1942	895	626	4	5896
DeKalb and Fulton, GA, USA (B)	1 823 234	122	136	196	7	12	58 933
Saint Petersburg, Russia	5 351 935	1596	5128	233	21	18	30 144
Orange County, CA, USA	3 010 232	512	839	285	33	15	79 287
Rhode Island, USA	1 059 361	147	189	163	5	18	58 416
Andorra	77 543	35	35	11 236	1235	17	49 900
Puducherry District, India	1 250 000	145	393	840	41	6	9152
Iran	82 913 906	19 804	98 071	5197	740	6	12 937
Tamil Nadu, India	83 697 770	11 183	29 773	5765	1821	7	7191
Gardena Valley, Italy	10 700	15	15	2612	702	20	71 853
Israel	9 216 900	187	995	10 578	577	11	40 747
India	1 366 417 750	4172	45 766	1632	11	6	4735
Utsunomiya City, Japan	517 527	0	0	228	2	26	42 931
Castiglione d'Adda, Italy	4605	39	40	190	43	26	59 291
Faroe Islands, Denmark	52 154	0	0	592	3	17	60 421
Spain	47 351 567	17 710	17 767	643	31	20	42 362
Munich, Germany	1 563 090	219	780	2692	96	17	65 345
Lima, Peru	10 804 609	21 109	28 846	893	226	9	19 313
Geneva, Switzerland (B)	504 128	122	156	2509	196	16	68 964
Kupferzell, Germany	6247	1	3	1263	153	16	63 885
Chennai, Tamil Nadu, India	10 900 000	1295	5960	380	70	7	7191
Delhi, India	19 800 000	4188	26 901	13 966	3965	4	12 817
Jersey, UK (B)	107 800	20	20	998	39	17	48 365
Gangelt, Germany	12 597	8	8	153	27	18	53 751
Netherlands	17 344 874	2621	4186	1652	47	20	59 685
England, UK (B)	56 550 000	29 526	30 038	10 635	634	18	43 310
Metropolitan France	64 897 954	17 396	19 038	5377	238	21	49 551
Tokyo, Japan	13 960 236	270	275	1314	2	20	45 796

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). 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.<sup>3</sup> 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, 65 yo<sub>k</sub> and GDP<sub>k</sub>, 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.<sup>4</sup>

observed between IgG seroprevalence and age'). On the other hand, note that among the 181 participants in the Nawa *et al.* [44] 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.* [44] found only 3 positive cases out of a total of 742 individuals tested, inference on this is limited.

<sup>4</sup>For reference, Levin *et al.* [3] 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.* [3] - Figure 6). A linear regression with intercept results in a value of 43%. The intraclass correlation coefficient (ICC) [76] between Levin *et al.* [3]'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.

<sup>3</sup>Malani *et al.* [41] ('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.* [45] ('Castiglione d'Adda, Italy') also report that the elderly are more likely to have been infected ('strong association

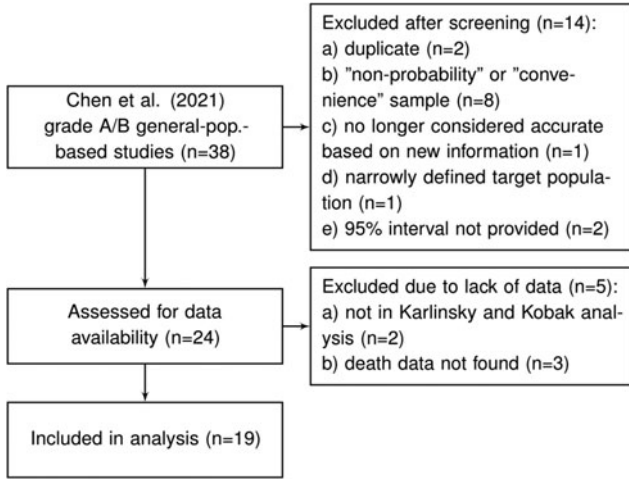


Fig. 1. Flowchart of seroprevalence studies considered for Chen *et al.* based analysis.

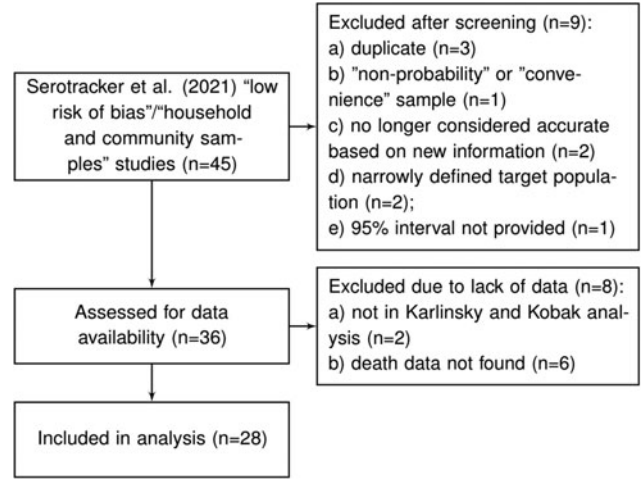


Fig. 2. Flowchart of seroprevalence studies considered for Serotracker-based analysis.

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 [77, 78]. Our estimate of  $\hat{\theta}_2 = -0.28$  (Chen *et al.*-based analysis) and  $\theta_2 = -0.11$  (Serotracker-based analysis) suggest that wealthier populations may be more likely to have lower IFRs. However, the wide CrIs 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.

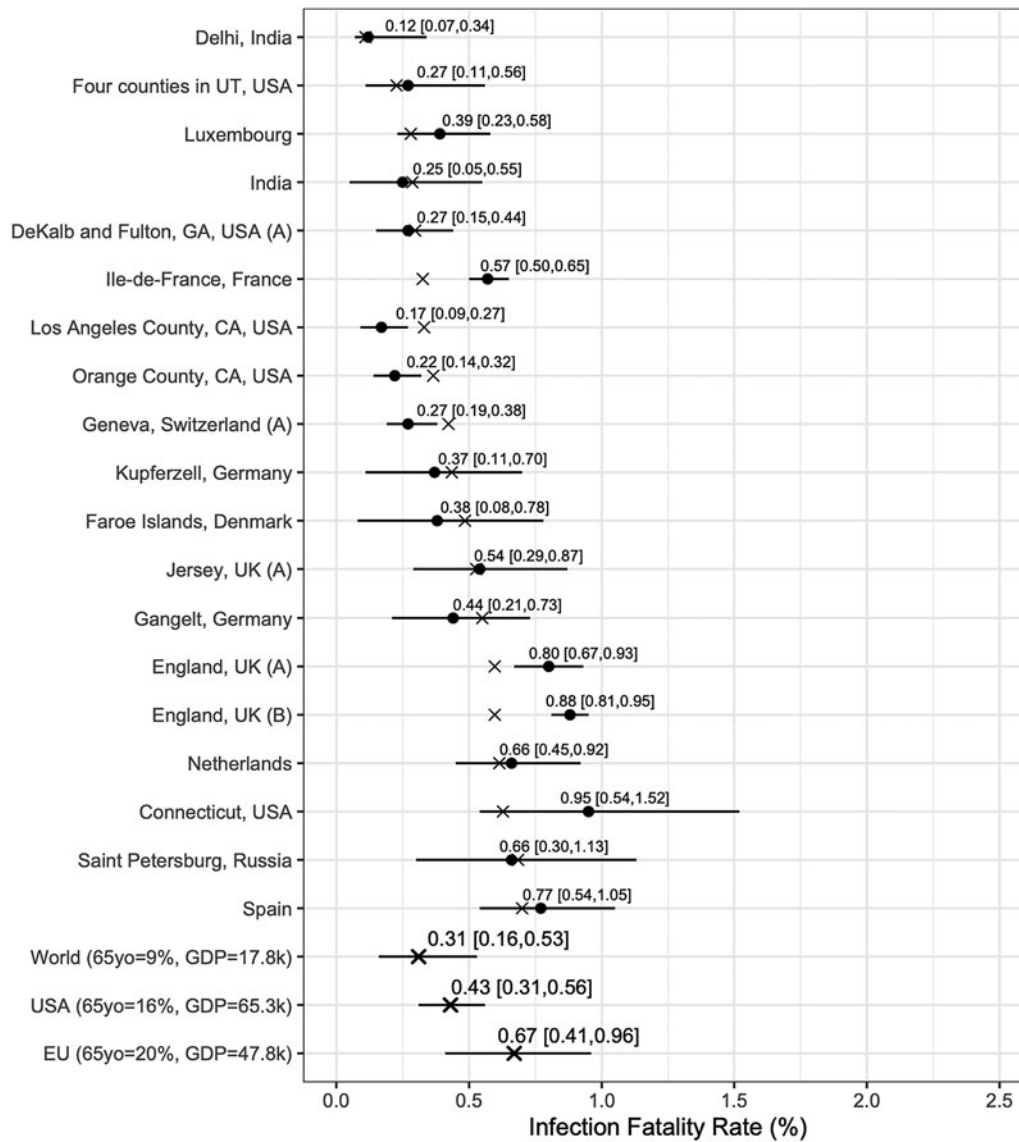
We can infer (by determining the posterior median of  $g^{-1}(\theta_0 + \theta_1 z_{1^*} + \theta_2 z_{2^*})$ , for selected values of  $z_{1^*}$  and  $z_{2^*}$ ) the typical IFR amongst populations (be they included in our study or not) having a given proportion of the populace aged over 65 and a 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 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  $GDP = \$ 65\,298$ , 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.* [9] or Arora *et al.* [10]), 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 Supplementary 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 CrIs). 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 Supplementary Figures S5 and S6 in the Supplementary Material.

Our estimates are somewhat similar to those obtained in other analyses. Brazeau *et al.* [4], 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.* [4] obtain an estimate of 1.15% (95% prediction interval of 0.78–1.79%). Ioannidis [2], using data from seroprevalence studies with sample sizes greater than 500 (and including deaths of LTC residents), obtains a 'median IFR 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.* [3], 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.* [3] 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.* [3] estimate an IFR of 1.90%.





**Fig. 3.** Results from the Chen *et al.*-based analysis: posterior median estimates (black circles) for the  $IFR_k$  variables (for  $k = 1, \dots, 19$ ) 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).

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.* [79] and the 0.75% estimate obtained by Brazeau *et al.* [4] (both of these excluding deaths of LTC residents).

Specifically, with regards to the United States, Sullivan *et al.* [80] 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.<sup>5</sup> Pei *et al.* [81] 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

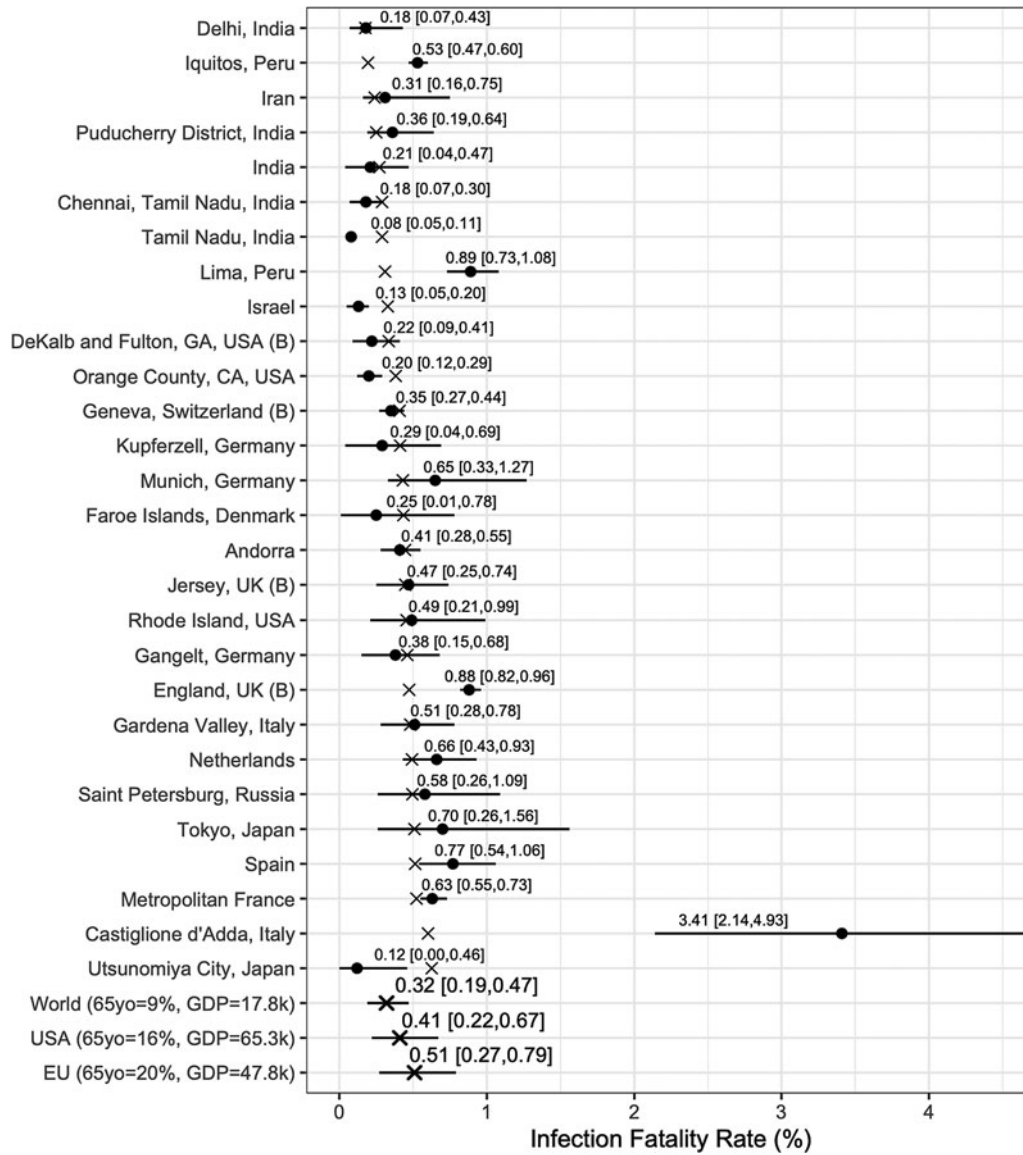
comparison, our ‘USA’ predictions of 0.43% and of 0.41% are based on data obtained mostly between April, 2020 and August, 2020 (see dates in Tables 1 and 2).

## 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 [82].

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

<sup>5</sup>While Sullivan *et al.* [80] used a nationwide representative sampling frame, the results may be subject to sizable selection bias given the low response rate of 12.6%.



**Fig. 4.** Results from the Serotracker-based analysis: posterior median estimates (black circles) for the  $IFR_k$  variables (for  $k = 1, \dots, 28$ ) 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).

**Table 5.** Parameter estimates (posterior medians and 95% HPD CrIs) obtained from the Chen *et al.*-based analysis and the Serotracker-based analysis

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

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 [8] notes that the time between infection and death may vary substantially

‘and may be shorter in developing countries where fewer people are long-sustained 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 [83].

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 [58]. 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 [84]. Notably, Axfors and Ioannidis [85] 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 overemphasised.<sup>6</sup> A major limitation of evidence synthesis is often summarised by the expression ‘garbage in, garbage out’ [86], meaning that if one includes biased studies in one’s analysis, the analysis results will themselves be biased [87]. 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.* [88] 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 [89]. Novel methods which allow evidence syntheses to appropriately incorporate biased data are urgently needed. Recently, Campbell *et al.* [90] 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 criteria, statistical modelling approaches and prior specifications [11]. 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 [91–93] and immense efforts have been made in the collection and analysis of data (e.g. Williamson *et al.* [77]). 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 [94], 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 [89, 95].

We prioritised simplicity in our modelling 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 [96]). 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 [97]. Age-group specific seroprevalence/mortality data is available for certain geographic areas [98] (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.* [3] for estimating age-group specific IFRs.

Finally, we must emphasise that the IFR is a moving target. As the pandemic changes, so 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 [81, 99, 100].

### 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.

**Supplementary material.** The supplementary material for this article can be found at <https://doi.org/10.1017/S0950268821002405>.

**Acknowledgements.** 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.

<sup>6</sup>Bobrovitz *et al.* [53] 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-Sa *et al.* [58] note that ‘it is very difficult to estimate the level of bias introduced by convenience sampling.’

**Financial support.** 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.

**Conflicts of interest.** None declared.

**Data availability statement.** Data and code used for the analysis are available in the Supplementary Material and at the OSF project 'Inferring the COVID-19 IFR 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

1. Clapham H *et al.* (1978) Seroepidemiologic study designs for determining SARS-CoV-2 transmission and immunity. *Emerging Infectious Diseases* **26**, 2020.
2. Ioannidis JPA (2021) Infection fatality rate of COVID-19 inferred from seroprevalence data. *Bulletin of the World Health Organization* **99**, 19.
3. Levin AT *et al.* (2020) Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications. *European Journal of Epidemiology* **35**, 1123–1138.
4. Brazeau N *et al.* (2020) Report 34: COVID-19 infection fatality ratio: estimates from seroprevalence. <https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-34-ifr/> (Accessed 25 October 2021).
5. O'Driscoll M *et al.* (2021) Age-specific mortality and immunity patterns of SARS-CoV-2. *Nature* **590**, 140–145.
6. Brody-Moore P (2019) Bayesian hierarchical meta-analysis of asymptomatic Ebola seroprevalence. CMC Senior Theses. 2228. [https://scholarship.claremont.edu/cmc\\_theses/2228](https://scholarship.claremont.edu/cmc_theses/2228) (Accessed 25 October 2021)
7. Bastian H (2021) Peering through the smoke at a duel over COVID-19's infection fatality rate. PLoS blogs – Absolutely Maybe: <https://absolutely-maybe.plos.org/2021/05/31/peering-through-the-smoke-at-a-duel-over-covids-infection-fatality-rate/> (Accessed 25 October 2021).
8. Ioannidis JPA (2021) Reconciling estimates of global spread and infection fatality rates of COVID-19: an overview of systematic evaluations. *European Journal of Clinical Investigation* **51**, e13554.
9. Chen X *et al.* (2021) Serological evidence of human infection with SARS-CoV-2: a systematic review and meta-analysis. *The Lancet Global Health* **9**, E598–E609.
10. Arora RK *et al.* (2021) Serotracker: a global SARS-CoV-2 seroprevalence dashboard. *The Lancet Infectious Diseases* **21**, e75–e76.
11. Gelman A and Loken E (2013) 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.
12. Streck H *et al.* (2020) Infection fatality rate of SARS-CoV-2 infection in a German community with a super-spreading event. *Nature Communications* **11**, 5829.
13. Gelman A *et al.* (2006) Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis* **1**, 515–534.
14. 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> (Accessed 25 October 2021).
15. Berger JO (2013) *Statistical Decision Theory and Bayesian Analysis*. New York: Springer Science & Business Media.
16. Lambert PC *et al.* (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**, 2401–2428.
17. Barchuk A *et al.* (2020) Seroprevalence of SARS-CoV-2 antibodies in Saint Petersburg, Russia: a population-based study. *Scientific Reports* **11**, 12930.
18. Biggs HM *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**, 965.
19. Bruckner TA *et al.* (2021) Estimated seroprevalence of SARS-CoV-2 antibodies among adults in Orange County, California. *Scientific Reports* **11**, 1–9.
20. Carrat F *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. *medRxiv*. doi: <https://doi.org/10.1101/2020.09.16.20195693>.
21. Mahajan *et al.* (2021) Seroprevalence of SARS-CoV-2-specific IgG antibodies among adults living in Connecticut: post-infection prevalence (PIP) study. *The American Journal of Medicine* **134**, 526–534.
22. Murhekar *et al.* (2020) Prevalence of SARS-CoV-2 infection in India: findings from the national serosurvey, May–June 2020. *Indian Journal of Medical Research* **152**, 48.
23. Office of National Statistics (2020) Coronavirus (COVID-19) infection survey: England. <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/coronaviruscovid19infectionsurveydata> (Accessed 25 October 2021).
24. Petersen MS *et al.* (2020) Seroprevalence of SARS-CoV-2-specific antibodies, Faroe Islands. *Emerging Infectious Diseases* **26**, 2760.
25. Pollán M *et al.* (2020) Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *The Lancet* **396**, 535–544.
26. Samore M *et al.* (2020) SARS-CoV-2 seroprevalence and detection fraction in Utah urban populations from a probability-based sample. *medRxiv*. doi: <https://doi.org/10.1101/2020.10.26.20219907>.
27. Santos-Hövenner C *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**, 2001752.
28. Sharma N *et al.* (2020) The seroprevalence and trends of SARS-CoV-2 in Delhi, India: a repeated population-based seroepidemiological study. *medRxiv*. doi: <https://doi.org/10.1101/2020.12.13.20248123>.
29. Snoeck CJ *et al.* (2020) Prevalence of SARS-CoV-2 infection in the Luxembourgish population: the CON-VINCE study. *medRxiv*. doi: <https://doi.org/10.1101/2020.05.11.20092916>.
30. Sood N *et al.* (2020) Seroprevalence of SARS-CoV-2-specific antibodies among adults in Los Angeles county, California, on April 10–11, 2020. *JAMA* **323**, 2425–2427.
31. Statistics Jersey (2020) SARS-CoV-2: prevalence of antibodies in Jersey (community survey round 2). <https://tinyurl.com/6htjn3af> (Accessed 25 October 2021).
32. Stringhini S *et al.* (2020) Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (serocov-pop): a population-based study. *The Lancet* **396**, 313–319.
33. Vos ERA *et al.* (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**, 489–495.
34. Ward H *et al.* (2020) Declining prevalence of antibody positivity to SARS-CoV-2: a community study of 365,000 adults. *medRxiv*. doi: <https://doi.org/10.1101/2020.10.26.20219725>.
35. Álvarez-Antonio *et al.* (2021) Seroprevalence of anti-sARS-CoV-2 antibodies in Iquitos, Peru in July and August, 2020: a population-based study. *The Lancet Global Health* **9**, e925–e931.
36. Bajema KL *et al.* (2020) Comparison of estimated severe acute respiratory syndrome coronavirus 2 seroprevalence through commercial laboratory residual sera testing and a community survey. *Clinical Infectious Diseases* **73**, e3120–e3123.
37. Chan PA *et al.* (2021) Seroprevalence of SARS-CoV-2 antibodies in Rhode Island from a statewide random sample. *American Journal of Public Health* **111**, 700–703.
38. Govern d'Andorra (2020) anticossos permeten diagnosticar 78 positius de la COVID-19, que podrien haver contagiats unes 360 persones. Available at <https://tinyurl.com/dybe27uk>.
39. Kar SS *et al.* (2021) Prevalence and time trend of SARS-CoV-2 infection in Puducherry, India, August–October 2020. *Emerging Infectious Diseases* **27**, 666.
40. Khalagi K *et al.* (2021) Prevalence of COVID-19 in Iran: results of the first survey of the Iranian COVID-19 serological surveillance program. *Clinical Microbiology and Infection* **27**, 1666–1671.

41. **Malani A et al.** (2021) Seroprevalence of SARS-CoV-2 in slums versus non-slums in Mumbai. *India. The Lancet Global Health* **9**, e110–e111.
42. **Melotti R et al.** (2021) Prevalence and determinants of serum antibodies to SARS-CoV-2 in the general population of the Gardena Valley. *medRxiv*. doi: <https://doi.org/10.1101/2021.03.19.21253883>.
43. **MoHoI Ministry of Health of Israel** (2020) National sero-epidemiological coverage for COVID-19. [https://www.gov.il/BlobFolder/reports/de-covid19-28062020-17092020/he/files\\_publications\\_corona\\_DE-covid19.pdf](https://www.gov.il/BlobFolder/reports/de-covid19-28062020-17092020/he/files_publications_corona_DE-covid19.pdf) (Accessed 25 October 2021).
44. **Nawa N et al.** (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*. doi: <https://doi.org/10.1101/2020.07.20.20155945>.
45. **Pagani G et al.** (2021) Prevalence of SARS-CoV-2 in an area of unrestricted viral circulation: mass seroepidemiological screening in Castiglione d'Adda, Italy. *PLoS ONE* **16**, e0246513.
46. **Radon K et al.** (2020) Protocol of a population-based prospective COVID-19 cohort study Munich, Germany (koco19). *medRxiv*. doi: <https://doi.org/10.1101/2020.04.28.20082743>.
47. **Reyes-Vega MF et al.** (2021) SARS-CoV-2 prevalence associated to low socioeconomic status and overcrowding in an LMIC megacity: a population-based seroepidemiological survey in Lima. *Peru. EClinicalMedicine* **34**, 100801.
48. **Richard A 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*. doi: <https://doi.org/10.1101/2020.12.16.20248180>.
49. **Selvaraju S et al.** (2021) Population-based serosurvey for severe acute respiratory syndrome coronavirus 2 transmission, Chennai, India. *Emerging Infectious Diseases* **27**, 586.
50. **Statistics Jersey** (2020) SARS-CoV-2: prevalence of antibodies in Jersey (community survey round 3). <https://tinyurl.com/hxd8sadb> (Accessed 25 October 2021).
51. **Warszawski J et al.** (2020) In May 2020, 4.5% of the population of metropolitan France had developed antibodies against SARS-CoV-2. <https://drees.solidarites-sante.gouv.fr/sites/default/files/2021-01/er1167-en.pdf> (Accessed 25 October 2021).
52. **Yoshiyama T et al.** (2021) Prevalence of SARS-CoV-2-specific antibodies, Japan, June 2020. *Emerging Infectious Diseases* **27**, 628.
53. **Bobrovitz N et al.** (2021) Global seroprevalence of SARS-CoV-2 antibodies: a systematic review and meta-analysis. *PLoS One* **16**, e0252617.
54. **Brownstein NC and Chen YA** (2021) Predictive values, uncertainty, and interpretation of serology tests for the novel coronavirus. *Scientific Reports* **11**, 1–12.
55. **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.
56. **Linton NM et al.** (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**, 538.
57. **Wu JT et al.** (2020) Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nature Medicine* **26**, 506–510.
58. **Shook-Sa BE, Boyce RM and Aiello AE** (2020) Estimation without representation: early severe acute respiratory syndrome coronavirus 2 seroprevalence studies and the path forward. *The Journal of Infectious Diseases* **222**, 1086–1089.
59. **Horby PW et al.** (2017) CONSIDER Statement on the reporting of seroepidemiologic studies for influenza (ROSES-I statement): an extension of the STROBE statement. *Influenza and Other Respiratory Viruses* **11**, 2–14.
60. **Munn Z et al.** (2015) Methodological guidance for systematic reviews of observational epidemiological studies reporting prevalence and cumulative incidence data. *International Journal of Evidence-Based Healthcare* **13**, 147–153.
61. **Leon DA et al.** (2020) COVID-19: a need for real-time monitoring of weekly excess deaths. *The Lancet* **395**, E81.
62. **Karlinsky A and Kobak D** (2021) Tracking excess mortality across countries during the COVID-19 pandemic with the world mortality dataset. *eLife* **10**, e69336.
63. **Banaji M** (2021) Estimating COVID-19 infection fatality rate in Mumbai during 2020. *medRxiv*. doi: <https://doi.org/10.1101/2021.04.08.21255101>.
64. **Pulla P** (2020) What counts as a COVID-19 death? *The BMJ* **370**, m2859.
65. **Liu J et al.** (2021) Excess mortality in Wuhan city and other parts of China during the three months of the COVID-19 outbreak: findings from nationwide mortality registries. *The BMJ* **372**, n415.
66. **The Economist** (2021) COVID-19 deaths in Wuhan seem far higher than the official count. Available at <https://www.economist.com/graphic-detail/2021/05/30/covid-19-deaths-in-wuhan-seem-far-higher-than-the-official-count> (Accessed 25 October 2021).
67. **Qutob N et al.** (2021) Seroprevalence of SARS-CoV-2 in the West Bank region of Palestine: a cross-sectional seroepidemiological study. *BMJ Open* **11**, e044552.
68. **He Z et al.** (2021) Seroprevalence and humoral immune durability of anti-SARS-CoV-2 antibodies in Wuhan, China: a longitudinal, population-level, cross-sectional study. *The Lancet* **397**, 1075–1084.
69. **Alemu B et al.** (2020) Sero-prevalence of anti-SARS-CoV-2 antibodies in Addis Ababa, Ethiopia. *bioRxiv*. doi: <https://doi.org/10.1101/2020.10.13.337287>
70. **Nisar MI et al.** (2020) Serial population-based sero-surveys for COVID-19 in low and high transmission. *medRxiv*. doi: <https://doi.org/10.1101/2020.07.28.20163451>.
71. **Mukherjee B et al.** (2021) Estimating the infection fatality rate from SARS-CoV-2 in India. Available at SSRN 3798552.
72. **Purkayastha S et al.** (2021) Estimating the wave 1 and wave 2 infection fatality rates from SARS-CoV-2 in India. *medRxiv*. doi: <https://doi.org/10.1101/2021.05.25.21257823>.
73. **Danis K et al.** (2020) High impact of COVID-19 in long-term care facilities, suggestion for monitoring in the EU/EEA, May 2020. *Eurosurveillance* **25**, 2000956.
74. **Pillemer K, Subramanian L and Hupert N** (2020) The importance of long-term care populations in models of COVID-19. *JAMA* **324**, 25–26.
75. **Kruschke J** (2014) *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. London: Academic Press.
76. **Vetter TR and Schober P** (2018) Agreement analysis: what he said, she said versus you said. *Anesthesia & Analgesia* **126**, 2123–2128.
77. **Williamson EJ et al.** (2020) Factors associated with COVID-19-related death using openly. *Nature* **584**, 430–436.
78. **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**, 429–439.
79. **Pastor-Barriuso R et al.** (2020) SARS-CoV-2 infection fatality risk in a nationwide seroepidemiological study. *medRxiv*. doi: <https://doi.org/10.1101/2020.08.06.20169722>.
80. **Sullivan P et al.** (2021) SARS-CoV-2 cumulative incidence, United States, August–December 2020. *Clinical Infectious Diseases*, ciab626. doi: <https://doi.org/10.1093/cid/ciab626>.
81. **Pei S et al.** (2021) Overall burden and characteristics of COVID-19 in the United States during 2020. *medRxiv*. doi: <https://doi.org/10.1101/2021.02.15.21251777>.
82. **Zelner J et al.** (2021) Accounting for uncertainty during a pandemic. *Patterns* **2**, 100310.
83. **Morfeld P et al.** (2021) COVID-19: heterogeneous excess mortality and 'burden of disease' in Germany and Italy and their states and regions, January–June 2020. *Frontiers in Public Health* **9**, 477.
84. **Ripperger TJ et al.** (2020) Orthogonal SARS-CoV-2 serological assays enable surveillance of low-prevalence communities and reveal durable humoral immunity. *Immunity* **53**, 925–933.
85. **Axfors C and Ioannidis JP** (2021) Infection fatality rate of COVID-19 in community-dwelling populations with emphasis on the elderly: an overview. *medRxiv*. doi: <https://doi.org/10.1101/2021.07.08.21260210>.
86. **Eysenck HJ** (1978) An exercise in mega-silliness. *American Psychologist* **33**, 517.

87. **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**, 881–901.
88. **Levin AT et al.** (2021) Assessing the burden of COVID-19 in developing countries: systematic review, meta-analysis, and public policy implications. *medRxiv*. doi: <https://doi.org/10.1101/2021.09.29.21264325>.
89. **Ioannidis JPA and Lau J** (1998) Can quality of clinical trials and meta-analyses be quantified? *The Lancet* **352**, 590.
90. **Campbell H et al.** (2020) Bayesian Adjustment for preferential testing in estimating the COVID-19 infection fatality rate: theory and methods. *in press at Annals of Applied Statistics*.
91. **Faust JS** (2020) Comparing COVID-19 deaths to flu deaths is like comparing apples to oranges. *Scientific American*. <https://tinyurl.com/ydxx8el8> (Accessed 25 October 2021).
92. **Ioannidis JPA** (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> (Accessed 25 October 2021).
93. **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>.
94. **Higgins JPT** (2008) Commentary: heterogeneity in meta-analysis should be expected and appropriately quantified. *International Journal of Epidemiology* **37**, 1158–1160.
95. **Berlin JA** (1995) Invited commentary: benefits of heterogeneity in meta-analysis of data from epidemiologic studies. *American Journal of Epidemiology* **142**, 383–387.
96. **Geddie J and Aravindan A** (2020) Why is Singapore’s COVID-19 death rate the world’s lowest. *Reuters*. <https://www.reuters.com/article/health-coronavirus-singapore-explainer-idUSKBN2680TF> (Accessed 25 October 2021).
97. **Westreich D et al.** (2021) Choice of outcome in COVID-19 studies and implications for policy: mortality and fatality. *in press at American Journal of Epidemiology* **kwab244**. doi: <https://doi.org/10.1093/aje/kwab244>.
98. **Riffe T, Acosta A and Zarulli V** (2021) COVerAGE-DB: a global demographic database of COVID-19 cases and deaths. *International Journal of Epidemiology* **50**, 390–390f.
99. **Pietzonka P et al.** (2021) Bayesian Inference across multiple models suggests a strong increase in lethality of COVID-19 in late 2020 in the UK. *medRxiv*. doi: <https://doi.org/10.1101/2021.03.10.21253311>.
100. **Walensky RP, Walke HT and Fauci AS** (2021) SARS-CoV-2 variants of concern in the United States - challenges and opportunities. *JAMA* **325**, 1037–1038.