

PROF. JOHN IOANNIDIS (Orcid ID : 0000-0003-3118-6859)

Article type : Review Article

## **Reconciling estimates of global spread and infection fatality rates of COVID-19: an overview of systematic evaluations**

John P.A. Ioannidis

Departments of Medicine, of Epidemiology and Population Health, of Biomedical Data Science, and of Statistics, and Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA, USA

Correspondence to: John P.A. Ioannidis, SPRC, 1265 Welch Road, Medical School Office Building Room X306, Stanford, CA 94305, USA

E-mail: [jioannid@stanford.edu](mailto:jioannid@stanford.edu)

Funding: none; the Meta-Research Innovation Center at Stanford (METRICS) has been funded with grants from the Laura and John Arnold Foundation

Conflicts of interest: none

Disclosures: I am the author of one of the 6 evaluations assessed in this article.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/ECI.13554](https://doi.org/10.1111/ECI.13554)

This article is protected by copyright. All rights reserved

Acknowledgements: I am grateful to Niklas Bobrovitz and Rahul Arora for offering clarifications on their important study.

Keywords: COVID-19, infection fatality rate, seroprevalence, global health, bias, meta-analysis

Accepted Article

## **ABSTRACT (248 words)**

**Background:** Estimates of community spread and infection fatality rate (IFR) of COVID-19 have varied across studies. Efforts to synthesize the evidence reach seemingly discrepant conclusions.

**Methods:** Systematic evaluations of seroprevalence studies that had no restrictions based on country and which estimated either total number of people infected and/or aggregate IFRs were identified. Information was extracted and compared on eligibility criteria, searches, amount of evidence included, corrections/adjustments of seroprevalence and death counts, quantitative syntheses and handling of heterogeneity, main estimates, and global representativeness.

**Results:** Six systematic evaluations were eligible. Each combined data from 10-338 studies (9-50 countries), because of different eligibility criteria. Two evaluations had some overt flaws in data, violations of stated eligibility criteria, and biased eligibility criteria (e.g. excluding studies with few deaths) that consistently inflated IFR estimates. Perusal of quantitative synthesis methods also exhibited several challenges and biases. Global representativeness was low with 78-100% of the evidence coming from Europe or the Americas; the two most problematic evaluations considered only 1 study from other continents. Allowing for these caveats, 4 evaluations largely agreed in their main final estimates for global spread of the pandemic and the other two evaluations would also agree after correcting overt flaws and biases.

**Conclusions:** All systematic evaluations of seroprevalence data converge that SARS-CoV-2 infection is widely spread globally. Acknowledging residual uncertainties, the available evidence suggests average global IFR of ~0.15% and ~1.5-2.0 billion infections by February 2021 with substantial differences in IFR and in infection spread across continents, countries, and locations.

## Highlights

- Six systematic evaluations have evaluated seroprevalence studies without restrictions based on country and have estimated either total number of people infected and/or aggregate infection fatality rates for SARS-CoV-2.
- These systematic evaluations have combined data from 10-338 studies (9-50 countries) each with partly overlapping evidence synthesis approaches.
- Some eligibility, design, and data synthesis choices are biased, while other differing choices are defensible.
- Most of the evidence (78-100%) comes from Europe or the Americas.
- All systematic evaluations of seroprevalence data converge that SARS-CoV-2 infection has been very widely spread globally.
- Global infection fatality rate is approximately 0.15% with 1.5-2.0 billion infections as of February 2021.

The extent of community spread of SARS-COV2 infection and the infection fatality rate (IFR) of COVID-19 are hotly debated. Many seroprevalence studies have provided relevant estimates. These estimates feed into projections that influence decision-making. Single studies create confusion, since they leave large uncertainty and unclear generalizability across countries, locations, settings and time points. Some overarching evaluations have systematically integrated data from multiple studies and countries.<sup>1-6</sup> These synthetic efforts probe what are typical estimates of spread and IFR, how heterogeneous they are, and what factors explain heterogeneity. An overview of these systematic evaluations comparing their methods, biases, and inferences may help reconcile their findings on these important parameters of the COVID-19 pandemic.

## **METHODS**

### **Eligible articles**

Articles were eligible if they included a systematic review of studies aiming to assess SARS-CoV-2 seroprevalence; there were no restrictions based on country; and an effort was made to estimate either a total number of people infected and/or aggregate IFRs. Articles were excluded if they considered exclusively studies of particular populations at different risk of infection than the general population (e.g. only healthcare workers), if they focused on specific countries (by eligibility criteria, not by data availability), and if they made no effort to estimate total numbers of people infected and/or aggregate IFRs.

### **Search strategy**

Searches were updated until January 14, 2021 in PubMed, medRxiv and bioRxiv with “seroprevalence [ti] OR fatality [ti] OR immunity [ti]” For feasibility, the search in PubMed was made more specific by adding “(systematic review OR meta-analysis OR analysis)”.. Communication with experts sought potentially additional eligible analyses (e.g. unindexed influential reports).

### **Extracted information**

From each eligible evaluation, the following information was extracted:

1. Types of information included (seroprevalence, other)
2. Date of last search, search sources and types of publications included (peer-reviewed, preprints, reports/other)
3. Types of seroprevalence designs/studies included

4. Number of studies, countries, locations included
5. Seroprevalence calculations: adjustment/correction for test performance, covariates, type of antibodies measured, seroreversion (loss of antibodies over time)
6. Death count calculations: done or not; adjustments for over- or under-counting, time window for counting COVID-19 deaths in relationship to seroprevalence measurements
7. Quantitative synthesis: whether data were first synthesized from seroprevalence studies in the same location/country/other level; whether meta-analyses were performed across locations/countries and methods used; handling of heterogeneity, stratification and/or regression analyses, including subgroups
8. Reported estimates of infection spread, under-ascertainment ratios (total/documentated infections) and/or IFR
9. Global representativeness of the evidence: proportion of the evidence (weight, countries, studies or locations, depending on how data synthesis had been done) from Europe and North America (sensitivity analysis: Europe and America)

### **Comparative assessment**

Based on the above, the eligible evaluations were compared against each other with focus on features that may lead to bias and trying to decipher the direction of each bias.

## **RESULTS**

### **Eligible evaluations**

Nine potentially eligible articles were retrieved<sup>1-3,5-10</sup> And 4 were rejected (Figure 1)<sup>7-10</sup> One more eligible report<sup>4</sup> was identified from communication with experts. The 6 eligible evaluations are named after their first authors or team throughout the manuscript.

### **Information used (Table 1)**

Five evaluations included only seroprevalence studies. Meyerowitz-Katz also included non-serological and modeling papers; summary IFR was smaller in the seroprevalence studies (0.60% versus 0.84% in others). The 6 evaluations differed modestly in dates of last search (range, 6/16/2020-9/9/2020) and in sources searched. Given that few studies outside of Europe and Americas were released early, evaluations with earlier searches have a more prominent dearth of low-IFR studies from countries with younger populations and fewer nursing home residents.

Eligibility criteria varied and were sometimes unclear or left room for subjectivity.

Consequently, eligible studies varied from 10 to 348 and countries covered with eligible data varied from 9 to 50. Two evaluations<sup>1,4</sup> excluded studies in overtly biased ways, leading to inflated IFR estimates.

Specifically, Meyerowitz-Katz excluded one study with low IFR<sup>5</sup> alluding that the study itself “explicitly warned against using its data to obtain an IFR”;<sup>1</sup> as co-investigator of the study, both myself and my colleagues are intrigued at this claim. They also excluded two more studies with low IFR alluding that it “was difficult to determine the numerator (i.e. number of deaths) associated with the seroprevalence estimate or the denominator (i.e. population) was not well defined”,<sup>1</sup> while one even presented IFR estimates in its published paper. Another excluded paper<sup>14</sup> tabulated several seroprevalence studies with median IFR=0.31%, half the Meyerowitz-Katz estimate.

The Imperial College COVID-19 Response Team (ICCRT) excluded studies with <100 deaths at the serosurvey midpoint.<sup>4</sup> This exclusion criterion introduces bias since number of deaths is the numerator in calculating IFR. Exclusion of studies with low numerator excludes studies likely to have low IFR. Indeed, 5 of 6 excluded studies with <100 deaths (Kenya, LA County, Rio Grande do Sul, Gangelt, Scotland)<sup>11,12,15-17</sup> have lower IFR than the 10 ICCRT-included studies; the sixth (Luxembourg)<sup>18</sup> is in the lower range of the 10 ICCRT-included studies.

The 6 evaluations varied on types of populations considered eligible. Table 2 summarizes biases involved in each study population type. General population studies are probably less biased, provided they recruit their intended sample. Conversely, studies of healthcare workers,<sup>19</sup> other high-risk exposure workers and closed/confined communities may overestimate seroprevalence; these studies were generally excluded, either upfront (5/6 evaluations) or when calculating key estimates (Bobrovitz). Other designs/populations may be biased in either direction, more frequently towards underestimating seroprevalence.<sup>20-27</sup> Three evaluations (Meyerowitz-Katz, ICCRT, O’Driscoll) were very aggressive with exclusions.

ICCRT had the most draconian exclusion criteria, excluding 165/175 identified seroprevalence studies. However, ICCRT actually dropped many general population studies (for various reasons), but included two blood donor studies<sup>28,29</sup> (out of many such) and one New York study<sup>30</sup> with convenience samples of volunteers recruited while entering grocery stores and through an in-store flyer. The latter

inclusion goes against the stated ICCRT eligibility criteria where self-selection is reason for exclusion. The New York study<sup>30</sup> had high IFR (from the worst-hit state in the first wave). The preliminary press-released report from an Italian general population survey,<sup>31</sup> was included in violation of ICCRT eligibility criteria<sup>4</sup> that a study should have performed its own antibody test validation; ICCRT “salvaged” the Italian study by transporting validation data from another study in San Francisco. The Italian study report<sup>31</sup> showed data on only 64,660 of the intended 150,000 participants (missingness 57%). Its inferred IFR estimate (2.5%) is an extreme outlier (2-20-fold larger than other reported European estimates) and simply impossible: it matches/exceeds case fatality rates despite probably major under-ascertainment of infections in Italy.<sup>32</sup>

Finally, the 6 evaluations differed markedly on how many included seroprevalence estimates came from peer-reviewed publications (journal articles listed in the references) at the time of the evaluation: from only 1 peer-reviewed estimate in Meyerowitz-Katz to 61 in Rostami. Some included seroprevalence estimates that came from preprints/reports published in peer-reviewed journals by 2/2021; final publications could have minor/modest differences versus preprints/reports. Even journal-published estimates may get revised, e.g. a re-analysis increased Indiana seroprevalence estimates by a third.<sup>33</sup>

### **Seroprevalence and death calculations (Table 3)**

Three evaluations<sup>3,4,6</sup> routinely adjusted for test performance, one<sup>5</sup> adjusted for test performance when the authors of the studies had done so, and two were unclear. Depending on test sensitivity/specificity, lack of adjustment may inflate or deflate seroprevalence. Ioannidis selected the most fully adjusted seroprevalence estimate, when both adjusted and unadjusted estimates existed; other evaluations were unclear on this issue. Ioannidis corrected the seroprevalence upward when not all three types of antibodies (IgG, IgM, IgA) were assessed. ICCRT and O’Driscoll considered seroreversion adjustments.

Rostami and Bobrovitz did not collect death counts to estimate IFR. The other 4 evaluations did not systematically adjust death counts for under- or over-counting. Finally, ICCRT and O’Driscoll used distributional approaches on the time window for counting deaths (with means between seroconversion and death differing by 1.5 days and 10 days, respectively), Ioannidis counted deaths

until 7 days after the survey mid-point (or the date survey authors made a strong case for), and Meyerowitz-Katz counted deaths up until 10 days after survey end.

#### **Quantitative synthesis, heterogeneity, and main estimates (Table 4)**

The 6 evaluations differed in quantitative synthesis approaches with implications for the main results.

Meyerowitz-Katz used random effects meta-analysis of 26 IFRs calculating a summary estimate despite extreme between-study heterogeneity ( $I^2=99.2\%$ ). Such extreme heterogeneity precludes obtaining meaningful summary estimates. Estimates from the same country/location were not combined first, and two multiply-counted countries (Italy and China) have high IFRs entered in calculations. Meta-analysis limited to seroprevalence studies yielded slightly lower summary IFR (0.60% versus 0.68%), but extreme between-study heterogeneity persisted ( $I^2=99.5\%$ ), thus summary estimates remained meaningless. Extreme between-study heterogeneity persisted also within three risk-of-bias categories ( $I^2=99.6\%$ , 98.8%, and 94.8%, respectively), within Europe and within America. There was no between-study heterogeneity for 4 Asian estimates, but none came from seroprevalence data and their IFR estimate (0.46%) is far higher than many subsequent Asian studies (outside Wuhan) using seroprevalence data<sup>5</sup> instead of modeling.

Rostami also performed random effects meta-analyses but more appropriately combined at a first step seroprevalence data from studies in the same country, and in the same region; a summary estimate across all 107 estimates in all countries was also obtained. The step-wise approach avoids the Meyerowitz-Katz analysis flaw. However, seroprevalence estimates may still vary extremely even within the same location, e.g. if done at different times. Moreover, the main estimate of the evaluation (“263.5 million exposed/infected at the time of the study”) extrapolated to the global population the pooled estimate from all 107 datasets. The more appropriate estimate is a sum of the infected per country, or at least per region. Actually, the authors did calculate numbers of people exposed/infected per world region. The sum was 641 million, 2.5-fold larger. Moreover, these numbers did not reflect “the time of the study”: the 107 seroprevalence studies were done 2 to 6 months before the Rostami evaluation was written.

Bobrovitz calculated medians (overall and across several subgroups of studies) and Ioannidis calculated sample size-weighted means per location and then medians across locations. Their

approaches avoid multiple counting of locations with many estimates available. Bobrovitz also performed random effects inverse variance meta-analysis of prevalence ratios for diverse demographics (age, sex, race, close contact, healthcare workers). The approach is defensible, since prevalence ratios were calculated within each study, but still very large between-study heterogeneity existed ( $I^2=85.1-99.4\%$  per grouping factor) making results tenuous. Bobrovitz and Ioannidis reach congruent estimates for total number infected globally (643 million by November 17 and at least 500 million by September 12, respectively) with under-ascertainment ratios of 11.9 in November and 17.2 in September. Only the latter evaluation calculated IFRs (0.23% overall; 0.05% for those <70 years old).

ICCRT and O'Driscoll focused on age-stratified estimates. ICCRT extrapolated age-stratified estimates to the age-structure of populations of typical countries, obtaining separate overall IFR estimates for low-income countries (0.22%), lower-middle-income countries (0.37%), upper-middle income countries (0.57%) and high-income countries (1.06%). O'Driscoll made extrapolations to 45 countries estimating 5.27% of their population infected by September 1.

#### **Global representativeness (Table 5)**

Seroprevalence data lacked global representativeness. 72-91% of the seroprevalence evidence came from Europe and North America (78-100% from Europe or Americas). Lack of representativeness was most prominent in Meyerowitz-Katz (only one estimate from Asia, none from Africa), ICCRT (no estimates from Asia or Africa), and O'Driscoll (only one estimate from Africa, no estimate from Asia). However, ICCRT extrapolated to all countries globally and O'Driscoll extrapolated to 45 countries including 8 in Asia.

#### **DISCUSSION**

This overview of 6 systematic evaluations of global spread and/or IFR of SARS-CoV-2 utilizing seroprevalence data highlights differences in methods, calculations, and inferences. Several choices made by some evaluations led to bias. Other choices are defensible, and reveal some unavoidable variability on how evidence on these important questions should be handled.

Choices that led to biased, inflated IFR estimates are the inclusion of modeling estimates, inappropriate exclusion of low-IFR studies despite fitting stated inclusion criteria of the evaluators, inappropriate inclusion of high-IFR studies despite not fitting stated inclusion criteria, and using low

death counts as exclusion criterion. Two evaluations (Meyerowitz-Katz and ICCRT) suffered multiple such problems each. These biases contributed to generate inflated and, sometimes, overtly implausible results. These two evaluations also cherry-picked very scant evidence (16 and 10 studies, including only 1 and 5 peer-reviewed articles, respectively), while hundreds of seroprevalence estimates are available (Appendix 1).

Differences in types of study designs and populations considered eligible may be defended with various arguments by each evaluator. Studies of healthcare workers were consistently excluded. No consensus existed on studies of blood donors, clinical samples, workers at no obvious high-risk occupations, and various convenience samples; these designs have variable reliability. Reliability increases with careful adjustment for sampling, demographics, and other key factors and when missing data are limited. General population sampling is theoretically best, but general population studies may still suffer large bias from selective missingness. Unreachable individuals, institutionalized people and non-participating invitees are typically at higher infection risk; if so, some general population studies may substantially underestimate seroprevalence (overestimate IFR). E.g., Meyerowitz-Katz included a Danish government survey press release<sup>34</sup> where only 1071 of 2600 randomly selected invitees participated (missingness 59%); the estimated IFR (0.79%) is probably substantially inflated.<sup>6,29</sup>

Differences may also ensue from seroprevalence adjustments for test performance and other factors.<sup>35,36</sup> Sometimes the change in estimated seroprevalence is substantial.<sup>37-39</sup> Special caution is needed with low seroprevalence.<sup>40</sup> When not all types of antibodies are assessed, a correction may also be useful. Adjustment for test performance may seemingly suffice. However, control samples used to estimate test sensitivity come from PCR-tested diagnosed patients, while missed diagnoses typically reflect asymptomatic or less symptomatic patients not seeking testing. Sensitivity may be much lower in these people, as many develop no or low-titer antibodies.<sup>41,42</sup> Seroreversion has a similar impact. Preliminary evidence suggests substantial seroreversion.<sup>30,43-46</sup> E.g., among healthcare personnel, 28.2% seroreverted in 2 months (64.9% in those with low titers originally).<sup>46</sup> Only ICCRT and O'Driscoll considered corrections for seroreversion, but still did not allow for high seroreversion. All these factors would result in underestimating seroprevalence (overestimating IFR).

Both over- and under-counting of COVID-19 deaths (the IFR numerator) may exist,<sup>47,48</sup> varying across countries with different testing and death coding. Correction of COVID-19 death counts through excess deaths is problematic. Excess reflects both COVID-19 deaths and deaths from measures taken.<sup>47-50</sup> Year-to-year variability is substantial, even more so within age-strata. Comparison against averages of multiple previous years are naïve, worse in countries with substantial demographic changes. E.g., in the first wave, an excess of 8071 deaths (SMR 1.03, 95%CI 1.03-1.04) in Germany became a deficit of 4926 deaths (SMR 0.98, 95%CI 0.98-0.99) after accounting for demographic changes.<sup>51</sup>

The exact timepoint when deaths are counted may affect IFR calculations when surveys happen while many deaths are still accruing. All evaluations that counted deaths allowed for greater time for death to occur than for seroconversion, but Meyerowitz-Katz used a most extreme delay, considering deaths until 10 days after survey end. Surveys take from one day to over a month, thus inferred sampling-to-death delay may occasionally exceed 6 weeks. Meyerowitz-Katz defends this choice also in another paper<sup>10</sup> choosing 4 weeks after the serosurvey mid-point. However, the argument (accounting for death reporting delays) is weak. Several situational reports plot deaths according to date-of-occurrence rather than date-of-reporting anyhow.<sup>52</sup> Moreover, infection-to-death time varies substantially and may be shorter in developing countries where fewer people are long-sustained by medical support.

Some quantitative synthesis approaches were problematic, e.g. calculating summary estimates despite  $I^2 > 99\%$ ; or no data combination within the same country/location before synthesis across countries/locations. Another generic problem with meta-analysis of such data is that it penalizes better studies that allow more appropriately for uncertainty in estimates (e.g. by accounting for test performance and adjusting for important covariates). Studies with less rigorous or no adjustments may have narrower CIs (smaller variance, thus larger weight).<sup>5</sup> Finally, for IFR meta-analysis, studies with few deaths may have higher variance (lower weight) and these studies may have the lowest IFR.

Age-stratification for IFR estimation and synthesis is a reasonable choice to reduce between-study heterogeneity driven by steep COVID-19 death risk age-gradient.<sup>53</sup> However, both analyses<sup>4,6</sup> that capitalized on granular age-stratification made tenuous extrapolations to additional countries from thin or no data. ICCRT lacked seroprevalence data on low-income and lower middle-income

countries (~half the global population); upper middle-income countries (~35% of global population) were only represented by one estimate from Brazil assuming IFR=1%, exceeding 2-5-fold other peer-reviewed estimates from Brazil.<sup>12,54</sup> Estimates used from high-income countries included an impossible Italian estimate (IFR=2.5%)<sup>31</sup> and mostly non-peer-reviewed data. O’Driscoll was more careful, but still some IFR extrapolations appear highly inflated versus data from subsequently accrued seroprevalence studies. Their ensemble model assumed highest IFR in Japan (1.09%) and lowest in Kenya (0.09%) and Pakistan (0.16%). Currently available seroprevalence studies from these countries show markedly lower IFR estimates:  $\leq 0.03\%$ ,<sup>55-57</sup>  $\leq 0.01\%$ ,<sup>15</sup> and 0.04-0.07%,<sup>58,59</sup> respectively. In Japan, infections apparently spread widely without causing detectable excess mortality.<sup>55</sup> In Kenya, under-ascertainment compared with documented cases was ~1000-fold.<sup>15</sup> While some COVID-19 deaths are certainly missed in Africa, containment measures are more deadly.<sup>60</sup>

All 6 evaluations greatly over-represented Europe and America. Only two (Rostami and Ioannidis) included meaningful amounts of data from Asia and Africa (still less than their global population share) in main estimate calculations. Currently, extensive data suggest high under-ascertainment ratios in Africa and many Asian countries<sup>5,15,55-62</sup> and thus much lower IFR in Asia (outside Wuhan) and Africa than elsewhere.

Quality of seroprevalence studies varies. Risk-of-bias assessments in prevalence studies are difficult. There are multiple risk-of-bias scales/checklists,<sup>63-66</sup> but bias scores do not translate necessarily to higher or lower IFR estimates, while assessors often disagree in scoring (Appendix 2).

Acknowledging these caveats, 4 of the 6 evaluations largely reach congruent estimates of global pandemic spread. O’Driscoll estimated 5.27% of the population of 45 countries had been infected by September 1, 2020, i.e. 180 million infected among 3.4 billion. Excluding China, the proportion of population infected among the remaining 44 countries would be ~9%, likely >10% after accounting for seroreversion. Countries not included among the 45 include some of the most populous ones with high infection rates (India, Mexico, Brazil, most African countries). Therefore, arguably at least 10% of the non-China global population (i.e. at least 630 million) would be infected as of September 1. This is very similar to the Ioannidis (at least 500 million infected as of September 12) and Rostami (641 million infected by summer, when numbers are added per region) estimates. The

Bobrovitz estimate (643 million infected as of November 17) should be increased substantially given that only 2 of 17 countries informing the calculated under-ascertainment ratio were in Asia or Africa, continents with much larger under-ascertainment ratios. National surveys in India actually estimated 60% seroprevalence in November in urban areas.<sup>67</sup> Therefore, probably infected people globally were ~1 billion (if not more) by November 17 (compared with 54 million documented cases). By extrapolation, one may cautiously estimate ~1.5-2.0 billion infections as of February 21, 2021 (compared with 112 million documented cases). This corresponds to global IFR ~0.15% - a figure open to adjustment for any over- and under-counting of COVID-19 deaths (Appendix 3).

Meyerowitz-Katz and ICCRT reach higher estimates of IFR, but, as discussed above, these are largely due to endorsing selection criteria focusing on high-IFR countries, violations of chosen selection criteria and obvious flaws that consistently cause IFR overestimation. Similar concerns apply to another publication with implausibly high age-stratified IFRs by Mayerowitz-Katz limited to countries with advanced economies, again cherry-picking some of the highest IFR locations and estimates.<sup>11</sup>

Even correcting inappropriate exclusions/inclusion of studies, errors, and seroreversion, IFR still varies substantially across continents and countries. Overall average IFR may be ~0.3-0.4% in Europe and the Americas (~0.2% among community-dwelling non-institutionalized people), and ~0.05% in Africa<sup>15</sup> and Asia (excluding Wuhan). Within Europe, IFR estimates were probably substantially higher in the first wave in countries like Spain,<sup>68</sup> UK,<sup>69</sup> and Belgium<sup>70</sup> and lower in countries like Cyprus or Faroe Islands (~0.15%, even case fatality rate is very low),<sup>71</sup> Finland (~0.15%)<sup>72</sup> and Iceland (~0.3%).<sup>73</sup> One European country (Andorra) tested for antibodies 91% of its population.<sup>74</sup> Results<sup>74</sup> suggest an IFR less than half of what sampling surveys with greater missingness have inferred in neighboring Spain. Moreover, high seroreversion was noted, even a few weeks apart,<sup>74</sup> thus IFR may be even lower. Differences exist also within a country, e.g. within the USA, IFR differs markedly in disadvantaged New Orleans districts versus affluent Silicon Valley areas. Differences are driven by population age-structure, nursing home populations, effective sheltering of vulnerable people,<sup>75</sup> medical care, use of effective (e.g. dexamethasone)<sup>76</sup> or detrimental (e.g. hydroxychloroquine)<sup>77</sup> treatments, host genetics,<sup>78</sup> viral genetics and other factors.

IFR may change over time locally<sup>79</sup> and globally. If new vaccines and treatments pragmatically prevent deaths among the most vulnerable, theoretically global IFR may decrease even below 0.1%. However, there are still uncertainties both about the real-world effectiveness of new options, the pandemic course and post-pandemic SARS-CoV-2 outbreaks or seasonal re-occurrence. IFR will depend on settings and populations involved. E.g. even “common cold” coronaviruses have IFR~10% in nursing home outbreaks.<sup>80</sup>

Admittedly, primary studies, their overviews and the current overview of overviews have limitations. All estimates have uncertainty. Interpretation unavoidably has subjective elements. This challenge is well-known in the literature of discrepant systematic reviews.<sup>81-85</sup> Cross-linking diverse types of evidence generates even more diverse eligibility/design/analytical options. Nevertheless, one should separate clear errors and directional biases from defensible eligibility/design/analytical diversity.

Allowing for such residual uncertainties, reassuringly the picture from the 6 evaluations assessed here is relatively congruent: SARS-CoV2 is widely spread, has lower average IFR than originally feared, and substantial global and local heterogeneity. Using more accurate estimates of IFR may yield more appropriate planning, predictions, and evaluation of measures.

## REFERENCES

1. Meyerowitz-Katz G, Merone L. A systematic review and meta-analysis of published research data on COVID-19 infection fatality rates. *Int J Infect Dis.* 2020;101:138-148.
2. Rostami A, Sepidarkish M, Leeflang MMG, Riahi SM, Nourollahpour Shiadeh M, Esfandyari S, Mokdad AH, Hotez PJ, Gasser RB. SARS-CoV-2 seroprevalence worldwide: a systematic review and meta-analysis. *Clin Microbiol Infect.* 2020; 24:S1198-743X(20)30651-0. doi: 10.1016/j.cmi.2020.10.020.
3. Bobrovitz N, Arora RK, Cao C, et al. Global seroprevalence of SARS-CoV-2 antibodies: a systematic review and meta-analysis. *medRxiv.* 2020;doi: <https://doi.org/10.1101/2020.11.17.20233460>.
4. Imperial College COVID-19 response team. Report 34: COVID-19 Infection Fatality Ratio: Estimates from Seroprevalence. October 29, 2020.
5. Ioannidis JPA. The infection fatality rate of COVID-19 inferred from seroprevalence data. *Bulletin of the WHO* 2021;99:19-33F.
6. O'Driscoll M, Ribeiro Dos Santos G, Wang L, Cummings DAT, Azman AS, Paireau J, Fontanet A, Cauchemez S, Salje H. Age-specific mortality and immunity patterns of SARS-CoV-2. *Nature.* 2020;doi: 10.1038/s41586-020-2918-0.
7. Franceschi VB, Santos AS, Glaeser AB, Paiz JC, Caldana GD, Machado Lessa CL, et al. Population-based prevalence surveys during the Covid-19 pandemic: A systematic review. *Reviews in Medical Virology.* 2020;n/a(n/a):e2200.
8. Chen X, Chen Z, Azman AS, Deng X, Chen X, Lu W, et al. Serological evidence of human infection with SARS-CoV-2: a systematic review and meta-analysis. *medRxiv.* 2020:2020.09.11.20192773.
9. Aurora RK, Joseph A, Van Wyk J, Rocco S, Atmaja A, May E, et al. SeroTracker: a global SARS-CoV-2 seroprevalence dashboard. *Lancet Infect Dis.* 2020:S1473-3099(20)30631-9.
10. Levin AT, Hanage WP, Owusu-Boaitey N, Cochran KB, Walsh SP, Meyerowitz-Katz G. Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications. *Eur J Epidemiol.* 2020;35(12):1123-1138.

- Accepted Article
11. Sood N, Simon P, Ebner P, Eichner D, Reynolds J, Bendavid E, et al. Seroprevalence of SARS-CoV-2-specific antibodies among adults in Los Angeles County, California, on April 10–11, 2020. *JAMA*. 2020;323(23):2425–7.
  12. Silveira MF, Barros AJD, Horta BL, Pellanda LC, Victora GD, Dellagostin OA, Struchiner CJ, Burattini MN, Valim ARM, Berlezi EM, Mesa JM, Ikeda MLR, Mesenburg MA, Mantesso M, Dall'Agnol MM, Bittencourt RA, Hartwig FP, Menezes AMB, Barros FC, Hallal PC, Victora CG. Population-based surveys of antibodies against SARS-CoV-2 in Southern Brazil. *Nat Med*. 2020;26(8):1196-1199.
  13. Bryan A, Pepper G, Wener MH, Fink SL, Morishima C, Chaudhary A, et al. Performance characteristics of the Abbott Architect SARS-CoV-2 IgG assay and seroprevalence in Boise, Idaho. *J Clin Microbiol*. 2020;58(8):e00941-20.
  14. Wilson L. SARS-CoV2 COVID-19 infection fatality rate implied by the serology antibody testing in New York City. SSRN 2020; [http://refhub.elsevier.com/S1201-9712\(20\)32180-9/sbref0350](http://refhub.elsevier.com/S1201-9712(20)32180-9/sbref0350).
  15. Uyoga S, Adetifa IMO, Karanja HK, Nyagwange J, Tuju J, Wanjiku P, Aman R, Mwangangi M, Amoth P, Kasera K, Ng'ang'a W, Rombo C, Yegon C, Kithi K, Odhiambo E, Rotich T, Orgut I, Kihara S, Otiende M, Bottomley C, Mupe ZN, Kagucia EW, Gallagher KE, Etyang A, Voller S, Gitonga JN, Mugo D, Agoti CN, Otieno E, Ndwiga L, Lambe T, Wright D, Barasa E, Tsofa B, Bejon P, Ochola-Oyier LI, Agweyu A, Scott JAG, Warimwe GM. Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Kenyan blood donors. *Science*. 2021;371(6524):79-82.
  16. Streeck H, Schulte B, Kümmerer BM, 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 MM, Eis-Hübinger AM, Exner M, Schmithausen RM, Schmid M, Hartmann G. Infection fatality rate of SARS-CoV2 in a super-spreading event in Germany. *Nat Commun*. 2020;11(1):5829.
  17. Thompson C, Grayson N, Paton RS, Lourenco J, Penman BS, Lee L. Neutralising antibodies to SARS coronavirus 2 in Scottish blood donors – a pilot study of the value of serology to determine population exposure. *medRxiv*. 2020; doi.org/10.1101/2020.04.13.20060467.

- Accepted Article
18. Snoeck CJ, Vaillant M, Abdelrahman T, Satagopam VP, Turner JD, Beaumont K, et al. Prevalence of SARS-CoV-2 infection in the Luxembourgish population: the CON-VINCE study. medRxiv. 2020; doi.org/10.1101/2020.05.11.
  19. Galanis P, Vraka I, Fragkou D, Bilali A, Katelidou D. Seroprevalence of SARS-CoV-2 antibodies and associated factors in health care workers: a systematic review and meta-analysis. medRxiv. 2020;doi.org/10.1101/2020.10.23.20218289.
  20. Imbert E, Kinley PM, Scarborough A, et al. Coronavirus Disease 2019 (COVID-19) outbreak in a San Francisco homeless shelter. Clin Infect Dis. 2020;ciaa1071.
  21. <https://covidtracking.com/data/long-term-care>, accessed January 18, 2021.
  22. Burnett JJ. Psychographic and demographic characteristics of blood donors. Journal of Consumer Research 1981;8:62-66.
  23. Ganguli M, Lytle ME, Reynolds MD, Dodge HH. Random versus volunteer selection for a community-based study. J Gerontol A Biol Sci Med Sci. 1998;53(1):M39-46.
  24. Patel EU, Bloch EM, Grabowski MK. Sociodemographic and behavioral characteristics associated with blood donation in the United States: a population-based study. Transfusion. 2019;59(9):2899-2907.
  25. Lattimore S, Wickenden C, Brailsford SR. Blood donors in England and North Wales: demography and patterns of donation. Transfusion. 2015;55:91-9.
  26. Burgdorf KS, Simonsen J, Sundby A, Rostgaard K, Pedersen OB, Sørensen E, et al. Socio-demographic characteristics of Danish blood donors. PLoS ONE 2017;12: e0169112.
  27. Volken T, Weidmann C, Bart T, Fischer Y, Klüter H, Rüesch P. Individual characteristics associated with blood donation: a cross-national comparison of the German and Swiss population between 1994 and 2010. Transfus Med Hemother. 2013;40(2):133-8.
  28. Erikstrup C, Hother CE, Pedersen OBV, Mølbak K, Skov RL, Holm DK, et al. Estimation of SARS-CoV-2 infection fatality rate by real-time antibody screening of blood donors. Clin Infect Dis. 2020 Jun 25; ciaa849.
  29. Emmenegger M, De Cecco E, Lamparter D, Jacquat RPB, Ebner D, Schneider MM, et al. Early plateau of SARS-CoV-2 seroprevalence identified by tripartite immunoassay in a large population. medRxiv 2020; doi.org/10.1101/2020.05.31

- Accepted Article
30. Rosenberg ES, Tesoriero JM, Rosenthal EM, Chung R, Barranco MA, Styer LM, et al. Cumulative incidence and diagnosis of SARS-CoV-2 infection in New York. *Ann Epidemiol.* 2020 Aug;48:23–29.e4.
  31. Primi risultati dell' indagine di sieroprevalenza sul SARS-CoV2, <https://www.istat.it/it/archivio/246156>.
  32. Russell TW, Golding N, Hellewell J, et al. Reconstructing the early global dynamics of under-ascertained COVID-19 cases and infections. *BMC Medicine* 2020;18:332.
  33. Yiannoutsos CT, Halverson PK, Menachemi N. Bayesian estimation of SARS-CoV-2 prevalence in Indiana by random testing. *Proc Natl Acad Sci U S A* 2021;118:e2013906118.
  34. Statens Serum Institut. De første foreløbige resultater af undersøgelsen for COVID-19 i befolkningen er nu klar. In: <https://www.ssi.dk/aktuelt/nyheder/2020/de-forste-forelobige-resultater-af-undersogelsen-for-covid-19-i-befolkningen-er-nu-klar>, last accessed January 18, 2021.
  35. Burgess S, Ponsford MJ, Gill D. Are we underestimating seroprevalence of SARS-CoV-2? *BMJ.* 2020;370:m3364.
  36. Deeks JJ, Dinnes J, Takwoingi Y, et al., Cochrane COVID-19 Diagnostic Test Accuracy Group. Antibody tests for identification of current and past infection with SARS-CoV-2. *Cochrane Database Syst Rev* 2020;6:CD013652.
  37. Bendavid E, Mulaney B, Sood N, Shah S, Ling E, Bromley-Dulfano R, et al. COVID-19 Antibody Seroprevalence in Santa Clara County, California. *Intern J Epidemiol* (in press).
  38. Shakiba M, Nazemipour M, Salari A, Mehrabian F, Nazari SSH, Rezvani SM, Ghasempour Z, Heidarzadeh A, Mansournia MA. Seroprevalence of SARS-CoV-2 in Guilan Province, Iran, April 2020. *Emerg Infect Dis.* 2020;27(2).
  39. Nawa N, Kuramochi J, Sonoda S, Yamaoka Y, Nukui Y, Miyazaki Y, et al. 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.* 2020; doi.org/10.1101/2020.07.20.20155945.
  40. Cai B, Ioannidis JP, Bendavid E, Tian L. Exact inference for disease prevalence based on a test with unknown specificity and sensitivity. *arXiv.* 2020; <https://arxiv.org/abs/2011.14423>.

- Accepted Article
41. Long Q-X, Tang X-J, Shi Q-L, et al. Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nat Med.* 2020;26:1200-4.
  42. Solbach W, Schiffner J, Backhaus I, et al. Antibody profiling of COVID-19 patients in an urban low-incidence region in Northern Germany. *Front Public Health.* 2020;8:570543.
  43. Stringhini S, Wisniak A, Piumatti G, Azman AS, Lauer SA, Baysson H, et al. Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study. *Lancet.* 2020 Aug 1;396(10247):313–19.
  44. Ibarondo FJ, Fulcher JA, Goodman-Meza D, et al. Rapid decay of anti-SARS-CoV-2 antibodies in persons with mild covid-19. *N Engl J Med.* 2020; . doi:10.1056/NEJMc2025179.
  45. Shioda K, Lau MS, Kraay AN, Nelson KN, Siegler AJ, Sullivan PS, Collins MH, Weitz JS, Lopman BA. Estimating the cumulative incidence of SARS-CoV-2 infection and the infection fatality ratio in light of waning antibodies. *medRxiv.* 2020; Nov 16:2020.11.13.20231266.
  46. Self WH, Tenforde MW, Stubblefield WB, Feldstein LR, Steingrub JS, Shapiro NI, Ginde AA, Prekker ME, Brown SM, Peltan ID, Gong MN, Aboodi MS, Khan A, Exline MC, Files DC, Gibbs KW, Lindsell CJ, Rice TW, Jones ID, Halasa N, Talbot HK, Grijalva CG, Casey JD, Hager DN, Qadir N, Henning DJ, Coughlin MM, Schiffer J, Semenova V, Li H, Thornburg NJ, Patel MM; CDC COVID-19 Response Team; IVY Network. Decline in SARS-CoV-2 antibodies after mild infection among frontline health care Personnel in a Multistate Hospital Network - 12 States, April-August 2020. *MMWR Morb Mortal Wkly Rep.* 2020;69:1762-1766.
  47. Ioannidis JPA. Global perspective of COVID-19 epidemiology for a full-cycle pandemic. *Eur J Clin Invest.* 2020 Oct 7:e13423.
  48. VanderWeele TJ. Challenges estimating total lives lost in COVID-19 decisions: Consideration of mortality related to unemployment, social isolation, and depression. *JAMA.* 2020;324:445-446.
  49. Bauchner H, Fontanarosa PB. Excess deaths and the Great Pandemic of 2020. *JAMA.* 2020;324:1504-1505.
  50. Woolf SH, Chapman DA, Sabo RT, Weinberger DM, Hill L. Excess deaths from COVID-19 and other causes, March-April 2020. *JAMA.* 2020;324(5):510-513.

- Accepted Article
51. Stang A, Standl F, Kowall B, et al. Excess mortality due to COVID-19 in Germany. *J Infect.* 2020;81(5):797-801.
  52. Ioannidis JPA, Axfors C, Contopoulos-Ioannidis DG. Population-level COVID-19 mortality risk for non-elderly individuals overall and for non-elderly individuals without underlying diseases in pandemic epicenters. *Environ Res.* 2020 Sep;188:109890.
  53. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, Curtis HJ, Mehrkar A, Evans D, Inglesby P, Cockburn J, McDonald HI, MacKenna B, Tomlinson L, Douglas IJ, Rentsch CT, Mathur R, Wong AYS, Grieve R, Harrison D, Forbes H, Schultze A, Croker R, Parry J, Hester F, Harper S, Perera R, Evans SJW, Smeeth L, Goldacre B. Factors associated with COVID-19-related death using OpenSAFELY. *Nature.* 2020;584(7821):430-436.
  54. Buss LF, Prete CA Jr, Abraham CMM, Mendrone A Jr, Salomon T, de Almeida-Neto C, França RFO, Belotti MC, Carvalho MPSS, Costa AG, Crispim MAE, Ferreira SC, Fraiji NA, Gurzenda S, Whittaker C, Kamaura LT, Takecian PL, da Silva Peixoto P, Oikawa MK, Nishiya AS, Rocha V, Salles NA, de Souza Santos AA, da Silva MA, Custer B, Parag KV, Barral-Netto M, Kraemer MUG, Pereira RHM, Pybus OG, Busch MP, Castro MC, Dye C, Nascimento VH, Faria NR, Sabino EC. Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic. *Science.* 2021;371(6526):288-292.
  55. Hibino S, Hayashida K, Ahn AC, Hayashida Y. Dynamic change of COVID-19 seroprevalence among asymptomatic population in Tokyo during the second wave. *medRxiv* 2020; doi: <https://doi.org/10.1101/2020.09.21.20198796>.
  56. Takita M, Matsumura T, Yamamoto K, Yamashita E, Hosoda K, Hamaki T, Kusumi E. Geographical Profiles of COVID-19 Outbreak in Tokyo: An analysis of the primary care clinic-based point-of-care antibody testing. *J Prim Care Community Health.* 2020 Jan-Dec;11:2150132720942695.
  57. Doi A, Iwata K, Kuroda H, Hasuike T, Nasu S, Kanda A, et al. Estimation of seroprevalence of novel coronavirus disease (COVID-19) using preserved serum at an outpatient setting in Kobe, Japan: a cross-sectional study. *medRxiv* 2020; doi.org/10.1101/2020.04.26.20079822.

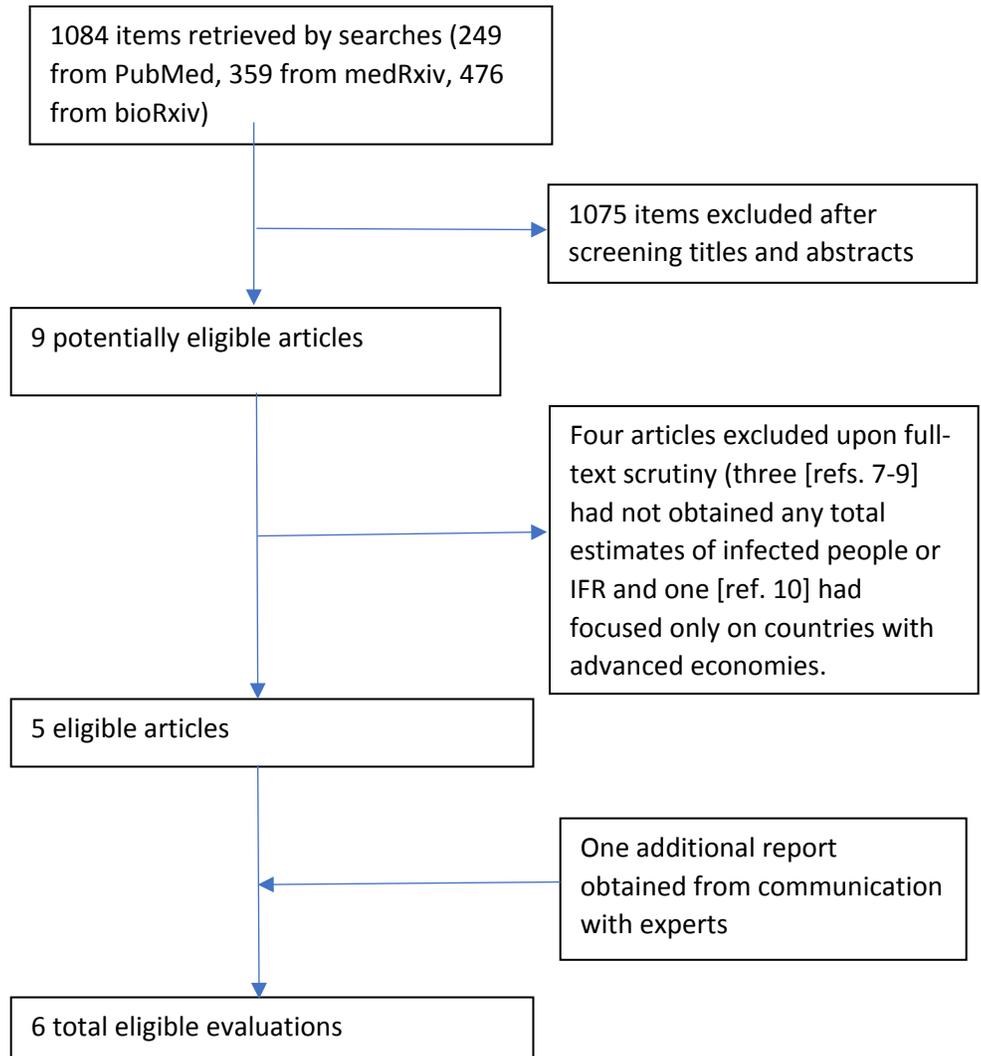
58. Nisar I, Ansari N, Amin M, Khalid F, Hotwani A, Rehman N, et al. Serial population based serosurvey of antibodies to SARS-CoV-2 in a low and high transmission area of Karachi, Pakistan. medRxiv 2020;doi.org/10.1101/2020.07.28.20163451
59. Javed W, Baqar J, Abidi SHB, Farooq W. Sero-prevalence findings from metropolises in Pakistan: implications for assessing COVID-19 prevalence and case-fatality within a dense, urban working population. medRxiv. 2020;doi.org/10.1101/2020.08.13.20173914.
60. Robertson T, Carter ED, Chou VB, et al. Early estimates of the indirect effects of the COVID-19 pandemic on maternal and child mortality in low-income and middle-income countries: a modelling study. *Lancet Glob Health*. 2020;8(7):e901-e908.
61. Poustchi H, Darvishian M, Mohammadi Z, Shayanrad A, Delavari A, Bahadorimonfared A, Eslami S, Javanmard SH, Shakiba E, Somi MH, Emami A, Saki N, Hormati A, Ansari-Moghaddam A, Saeedi M, Ghasemi-Kebria F, Mohebby I, Mansour-Ghanaei F, Karami M, Sharifi H, Pourfarzi F, Veisi N, Ghadimi R, Egtesad S, Niavarani A, Ali Asgari A, Sadeghi A, Sorouri M, Anushiravani A, Amani M, Kaveh S, Feizesani A, Tabarsi P, Keyvani H, Markarian M, Shafighian F, Sima A, Sadjadi A, Radmard AR, Mokdad AH, Sharafkhah M, Malekzadeh R. 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. *Lancet Infect Dis*. 2020;S1473-3099(20)30858-6. doi: 10.1016/S1473-3099(20)30858-6.
62. Abu Raddad LJ, Chemaitelly H, Ayoub HH, Al Kanaani Z, Al Khal A, Al Kuwari E, et al. Characterizing the Qatar advanced-phase SARS-CoV-2 epidemic. medRxiv. 2020; doi.org/10.1101/2020.07.16 .20155317.
63. Shamliyan T, Kane RL, Dickinson S. A systematic review of tools used to assess the quality of observational studies that examine incidence or prevalence and risk factors for diseases. *J Clin Epidemiol*. 2010;63:1061-1070.
64. Munn Z, Moola S, Lisy K, Riitano D, Tufanaru C. Methodological guidance for systematic reviews of observational epidemiological studies reporting prevalence and incidence data. *Int J Evid Based Health*. 2015;13:147–153.

65. Munn Z, Moola S, Riitano D, Lisy K. The development of a critical appraisal tool for use in systematic reviews addressing questions of prevalence. *Int J Health Policy Manag.* 2014;3:123-8.
66. Hoy D, Brooks P, Woolf A, Blyth F, March L, Bain C, Baker P, Smith E, Buchbinder R. Assessing risk of bias in prevalence studies: modification of an existing tool and evidence of interrater agreement. *J Clin Epidemiol.* 2012;65:934-9.
67. Ghosh A. India is missing about 90 infections for every COVID case, latest government analysis shows. <https://theprint.in/health/india-is-missing-about-90-infections-for-every-covid-case-latest-govt-analysis-shows/567898/>, last accessed January 18, 2021.
68. Pollán M, Pérez-Gómez B, Pastor-Barriuso R, Oteo J, Hernán MA, Pérez-Olmeda M, et al.; ENE-COVID Study Group. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *Lancet.* 2020;396(10250):535–44.
69. Ward H, Atchinson C, Whitaker M, Ainslie KED, Elliott J, Okell L, et al. Antibody prevalence for SARS-CoV-2 following the peak of the pandemic in England: REACT2 study in 100,000 adults. *medRxiv.* 2020;doi.org/10.1101/2020.08.12.
70. Herzog S, De Bie J, Abrams S, Wouters I, Ekinci E, Patteet L, et al. Seroprevalence of IgG antibodies against SARS coronavirus 2 in Belgium – a prospective cross sectional study of residual samples. *medRxiv.* 2020;doi.org/10.1101/2020.06.08.20125179.
71. Worldometer, [www.worldometer.com](http://www.worldometer.com), Cyprus and Faroe Islands, accessed February 21, 2021.
72. Weekly report of the population serology survey of the corona epidemic. Helsinki: Finland Department of Health and Welfare; 2020. [Finnish]. In: [https://www.thl.fi/roko/cov-vaestoserologia/sero\\_report\\_weekly.html](https://www.thl.fi/roko/cov-vaestoserologia/sero_report_weekly.html), accessed January 18, 2021.
73. Gudbjartsson DF, Norddahl GL, Melsted P, Gunnarsdottir K, Holm H, Eythorsson E, et al. Humoral immune response to SARS-CoV-2 in Iceland. *N Engl J Med.* 2020 Sep 1. *NEJMoa2026116* [online ahead of print]. doi: <http://dx.doi.org/10.1056/NEJMoa2026116> PMID: 32871063.
74. Royo-Cebrecos C, Vilanova D, Lopez J, et al. Mass SARS-CoV-2 serological screening for the Principality of Andorra. *Research Square* 2020.

75. Ioannidis JPA. Precision shielding for COVID-19: metrics of assessment and feasibility of deployment. medRxiv doi: <https://doi.org/10.1101/2020.11.01.20224147>. BMJ Global Health (in press).
76. Horby P, Lim WS, Emberson JR, et al. Dexamethasone in hospitalized patients with Covid-19 - preliminary report. *New Engl J Med*. 2020.Epub ahead of print.
77. Axfors C, Schmitt AM, Janiaud P, van 't Hooft J, Abd-Elsalam S, Abdo EF, et al. Mortality outcomes with hydroxychloroquine and chloroquine in COVID-19: an international collaborative meta-analysis of randomized trials. *Nature Communications* (in press).
78. Chen HH, Shaw DM, Petty LE, Graff M, Bohlender RJ, Polikowsky HG, Zhong X, Kim D, Buchanan VL, Preuss MH, Shuey MM, Loos RJF, Huff CD, Cox NJ, Bastarache JA, Bastarache L, North KE, Below JE. Host genetic effects in pneumonia. *Am J Hum Genet*. 2021;108(1):194-201.
79. Ioannidis JPA, Axfors CA, Contopoulos-Ioannidis DG. Second versus first wave of COVID-19 deaths: shifts in age distribution and in nursing home fatalities. *Environ Res*. 2021;195:110856.
80. Veiga ABGD, Martins LG, Riediger I, Mazetto A, Debur MDC, Gregianini TS. More than just a common cold: Endemic coronaviruses OC43, HKU1, NL63, and 229E associated with severe acute respiratory infection and fatality cases among healthy adults. *J Med Virol*. 2021;93(2):1002-1007.
81. Ioannidis JPA. Meta-research: the art of getting it wrong. *Res Synth Methods* 2010;1:169-84.
82. Campbell J, Bellamy N, Gee T. Differences between systematic reviews/meta-analyses of hyaluronic acid/hyaluronan/hylan in osteoarthritis of the knee. *Osteoarthritis Cartilage*. 2007;15:1424-36.
83. Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic reviews. *Can Med Assoc J*. 1997;156:1411-1416.
84. Moayyedi P, Deeks J, Talley NJ, Delaney B, Forman D. An update of the Cochrane systematic review of *Helicobacter pylori* eradication therapy in nonulcer dyspepsia: resolving the discrepancy between systematic reviews. *Am J Gastroenterol*. 2003;98: 2621-2626.

85. Linde K, Willich SN. How objective are systematic reviews? Differences between reviews on complementary medicine. *J Royal Soc Med.* 2003;96:17-22.

Figure 1. Flow diagram



**Table 1. Key features for eligible systematic data syntheses**

Features	Meyerowitz-Katz	Rostami	Bobrovitz	Imperial College COVID-19 response team	Ioannidis	O’Driscoll
Types of information included	SP, non-serological and modeling studies	SP studies	SP studies	SP studies	SP studies	SP studies
Last search	June 16	August 14	August 28	Unclear	September 9	Unclear (September 1?)
Search sources	PubMed, preprints (medRxiv, SSRN), Google, Twitter searches, government agency reports eligible	PubMed, Scopus, Embase, medRxiv, bioRxiv, research reports eligible	Medline, EMBASE, Web of Science, and Europe PMC, Google, communication with experts	Serotracker searches (see Bobrovitz)	PubMed (LitCOVID), medRxiv, bioRxiv, Research Square, national reports, communication with experts for additional studies	Unclear
Types of SP studies included	Excluded targeted populations with selection bias, also 4 other	Excluded at-risk populations (e.g. HCW), known diseases (e.g. dialysis, cancer)	All studies included if they reported on sample, date, region, and SP	Studies with defined sampling framework, defined geographic area, with availability of test performance,	General population or approximations (including blood donors, excluding high-risk, e.g. HCW, communities),	Unclear, but eventually it includes some general population studies, some blood donors,

	studies*		estimate	preferentially validation done as part of the study (not just by manufacturers), >100 deaths at SP study mid-point**; excluded healthcare workers, symptoms of COVID-19, self- referral or self- selection, narrow age range, confined settings, clinical samples	sample size >500, area with population >5000	and some hospital samples
Number of studies, countries, locations	24-27 studies***, of which 16 serological from 14 countries	107 datasets from 47 studies from 23 countries	338 studies (184 from general population) from 50 countries (36 from general population)****	10 studies (6 national, 4 subnational), 9 countries*****	82 estimates, 69 studies, 51 locations, 36 countries (main analysis at the location level)	25 studies from 20 countries (only 22 national representing 16 countries used in the ensemble model)
Studies published in peer-review journals	1/16	61/107	4/40 included in final analysis of	5/10	35/82	6/20 countries

at the time of the evaluation			under-ascertainment ratio			
-------------------------------	--	--	---------------------------	--	--	--

SP: seroprevalence, HCW: healthcare workers, IFR: infection fatality rate

\*One study (LA County)<sup>11</sup> with very low IFR was excluded with the justification that it “explicitly warned against using its data to obtain an IFR”; as a co-investigator of the study both myself and my colleagues are intrigued at the rationale for exclusion; in the publication of the study in JAMA<sup>11</sup> we did list limitations and caveats, as it is appropriate for any seroprevalence study to do; excluding studies that are honest to discuss limitations would keep only the worst studies that discuss no limitations. Two other studies with low IFR were excluded as well. One was done in Rio Grande do Sul<sup>12</sup> where its authors even report IFR estimates in their paper (0.29%, 0.23%, 0.38% in the three rounds of the serosurvey); the other was done in Boise,<sup>13</sup> where its authors properly discuss limitations but an approximation of IFR is possible; even if not perfectly accurate it is certainly lower than the IFR estimates included in the Meyerowitz-Katz meta-analysis. For the fourth excluded study,<sup>14</sup> the justification offered for its exclusion is that it “calculated an IFR, but did not allow for an estimate of confidence bounds”.<sup>1</sup> However, this study presents results of a New York study that Meyerowitz-Katz did include in their meta-analysis. Of note, that fourth study<sup>14</sup> also presents a cursory review of seroprevalence studies arriving at a median IFR=0.31%, half of the summary estimate of Meyerowitz-Katz.

\*\*clear bias introduced since number of deaths is the numerator itself in the calculation of IFR, and exclusion of studies with low numerator is thus excluding studies likely to have low IFR

\*\*\*different numbers provided by the authors for total studies in abstract (n=24), text of the paper (n=25), tabulated studies (n=27) and forest plot studies (n=26)

\*\*\*\*39 estimates from 17 countries used in main calculation of median under-ascertainment ratio (N. Bobrovitz, personal communication)

\*\*\*\*\*one of the 10 included studies violates the eligibility criterion of the investigators having validated themselves the antibody test used; the ICCRT included this study invoking validation data for the same antibody kit done by a different team in a study in a completely different setting and continent (San Francisco); based on this rationale perhaps many other studies could have been included, if the same violation of the eligibility criteria were tolerated. The included study was an Italian survey<sup>31</sup> which had only been released in the press with a preliminary report

at the time of the ICCRT evaluation, and which included crude results on only 64,660 of the intended 150,000 participants (missingness 57%). Its inferred IFR estimate (2.5%) is an extreme outlier, as it is 2-20-fold larger than other typical estimates reported from numerous European countries. Moreover, that IFR estimate even matches/exceeds case fatality rates and thus it is simply impossible. It is widely accepted that IFR must be several times smaller than case fatality rate, even in locations with substantial testing. Italy had very limited testing in the first wave and modest testing in the second wave. One estimate suggests that the number of infections in Italy at the peak of the first wave was 12 times more than the number of documented cases, i.e. the IFR would be more than an order of magnitude lower than the case fatality rate.<sup>32</sup>

**Table 2. Direction of potential bias in studies with different types of populations**

<b>Type of sampling</b>	<b>Direction of bias</b>
General population (entire population or design for representative sample)	Depends on characteristics of individuals who cannot be reached and/or decline participation. If they are more likely to be more disadvantaged (e.g. have no address/phone/e-mail) and thus also at higher risk of infection, SP may be underestimated. Potential for bias is more prominent when non-response/non-participation is larger. Institutionalized populations and homeless people are typically not included, and these populations often have very high infection rates; <sup>20,21</sup> thus SP is underestimated.
Convenience sample (including self-referral and response to adverts)	Bias could be in either direction. Volunteer bias is common and would tend to recruit more health conscious, low-risk individuals, <sup>22</sup> leading to SP underestimation. Conversely, interest to get tested because of worrying in the presence of symptoms may lead to SP overestimation.
Blood donors	Bias could be in either direction, but SP underestimation is more likely, since blood donors tend to be more health conscious and thus more likely to avoid also risky exposures. An early classic assessment <sup>23</sup> described blood donors as “low risk takers, very concerned with health, better educated, religious, and quite conservative” - characteristics that would lead to lower infection risk. In countries with large shares of minorities (e.g. USA and UK), minorities are markedly under-represented among blood donors. <sup>24,25</sup> E.g, in the USA, donation rates are 37-40% lower in blacks and Hispanics versus whites <sup>24</sup> and in the UK, donation rates range from

	<p>1.59 per 1000 among Asian Bangladeshi origin, compared to 22.1 per 1000 among white British origin.<sup>25</sup> These minorities were hit the most by COVID-19. In European countries, donations are lower in low-income and low-education individuals;<sup>26,27</sup> these are also risk factors for COVID-19 infection. Bobrovitz<sup>3</sup> found median seroprevalence of 3.2% in blood donor studies versus 4.1% in general community/household samples (risk ratio 0.80 in meta-regression). SP may be overestimated if blood donation is coupled to a free COVID-19 test in a poor population (as in the case of a study in Manaus, Brazil).</p>
<p>Clinical residual samples and patients (e.g. dialysis, cancer, other)</p>	<p>Bias could be in either direction, but SP underestimation is more likely since patients with known health problems may be more likely to protect themselves in a setting of a pandemic that poses them at high risk. Conversely, repeated exposure to medical facilities may increase risk. Demographic features and socioeconomic status may also affect the size and direction of bias. Bobrovitz<sup>3</sup> found median seroprevalence of 2.9% in studies of residual samples versus 4.1% in general community/household samples (risk ratio 0.63 in meta-regression). Hospital visitors' studies had even lower seroprevalence (median 1.4%).</p>
<p>Healthcare workers, emergency response, other workers with obvious high risk of exposure</p>	<p>Bias very likely to lead to SP overestimation compared with the general population, because of work-related contagion hazard; however, this may not always be the case (e.g. most infections may not happen at work) and any increased risk due to work exposure sometimes may be counterbalanced by favorable socioeconomic profile for some healthcare workers (e.g. wealthy physicians). Bias may have been more prominent in early days of the pandemic, especially in places lacking protective gear. Across 8 studies with data on healthcare workers</p>

	and other participants, seroprevalence was 1.74-fold in the former. <sup>3</sup>
Other workers	Bias could be in either direction and depends on work experience during the pandemic period and socioeconomic background, e.g. SP may be underestimated compared with the general population for workers who are wealthy and work from home during the pandemic and overestimated for essential workers.
Communities (shelters, religious, other shared-living)	Likely very strong bias due to high exposure risk leading to SP overestimation compared with the general population. Some of these communities were saturated with very high levels of infection very early. <sup>20,21</sup>

SP: seroprevalence

**Table 3. Adjustments and corrections for seroprevalence and death counts**

<b>Features</b>	<b>Meyerowitz-Katz</b>	<b>Rostami</b>	<b>Bobrovitz</b>	<b>Imperial College COVID-19 response team</b>	<b>Ioannidis</b>	<b>O'Driscoll</b>
Adjustment of SP for test performance	Unclear selection rule	Unclear selection rule	Yes (Bayesian)	Yes	Yes, when done by authors of SP study	Yes (24/25 studies)
Adjustment of SP for confounders	Unclear selection rule	Unclear selection rule	Unclear selection rule	Unclear selection rule	Selecting most fully adjusted SP estimated	Unclear selection rule
Other SP correction	No	No	No	Seroreversion	Type of antibodies*	Seroreversion, in secondary analysis
Death count adjustments	No adjustments	Deaths not assessed	Deaths not assessed	No adjustments	No adjustments	No adjustments

Time window for death counts	10 d after completion of SP study	Deaths not assessed	Deaths not assessed	Distributional (truncated gaussian and beta), mean 18.3 d from onset to seroconversion, 19.8 d from onset to death	7 d after mid-point of SP survey or as chosen by its authors	Distributional (gamma), mean 10 d from onset to seroconversion, 20 d from onset to death
------------------------------	-----------------------------------	---------------------	---------------------	--	--	--

SP: seroprevalence, IFR: infection fatality rate, d: days \*one-tenth adjustment per each not tested antibody (IgG, IgM, IgA)

**Table 4. Quantitative synthesis approaches, stratification and/or regression, and main estimates**

	<b>Meyerowitz-Katz</b>	<b>Rostami</b>	<b>Bobrovitz</b>	<b>Imperial College COVID-19 response team</b>	<b>Ioannidis</b>	<b>O'Driscoll</b>
Quantitative synthesis	26 IFR estimates combined at one-step with D-L RE model, $I^2=99.4\%$	First step 107 SP estimates combined separately for each country with D-L RE model, then per region. Also D-L RE for all 107 estimates, $I^2=99.7\%$	Median SP calculated overall and per subgroup of interest.	Log-linear model for pooling age-stratified IFR, then age-stratified estimates extrapolated to the age-structure of populations of typical countries	First step, sample-size weighted summary of SP per location; then median estimated across locations	The ensemble model eventually models age-stratified IFR in a total of 45 countries with available age-stratified death counts, but data are used as input from only 16 countries that have IFR data with some age stratification

Stratification and/or regression	Subgroup analyses per continent, month of publication, modeling versus serological, and risk of bias	Subgroup analyses per age, gender, type of population, serological method, race/ethnicity, income, human development index, latitude/longitude, humidity, temperature, days from onset of pandemic; also RE meta-regressions	Subgroup analyses per GBD region, scope (national, regional, local, sublocal), risk of bias, days since 100 <sup>th</sup> case (also explored in meta-regressions); RE inverse variance meta-analysis of prevalence ratios for demographics (age, sex, race, close contact, HCW status) with I <sup>2</sup> =85.1-99.4% per grouping factor	Focus on age-strata, also IFR estimates with and without seroreversion, and (for some countries) excluding nursing home deaths	Separate analyses for age <70 years; also subgroup analyses according to level of overall mortality in the location	Focus on age-strata; also per sex/gender and per country
Main estimates	Summary IFR 0.68 (95% CI-0.53-0.82%), 0.60 when limited to serological studies	263.5 million exposed/infected at the time of the study based on the pooled SP from all 107 datasets; when estimated per region the total is	643 million infected as of November 17, based on estimated median under-ascertainment factor of 11.9 (using 9 days before study end date for PCR counts)**	Overall IFR: LIC 0.22 (0.14, 0.39), LMIC) 0.37 (0.25, 0.61), UMIC 0.57 (0.38, 0.92), HIC 1.06 (0.73, 1.64)	Over 500 million infected as of September 12 (vs. 29 million documented cases) globally; median IFR 0.23% in the available studies (0.09% in locations with < 118	5.27% of the population of the 45 modeled countries had been infected by September 1.

		641 million*			deaths/million), 0.20% in locations with 118– 500 deaths/million, 0.57% in locations with > 500 deaths/million	
--	--	--------------	--	--	--	--

SP: seroprevalence, IFR: infection fatality rate, RE: random effects

\*in millions: Europe+North America 47, East+Southeast Asia 47, Latin America 9, South America 6, Sub Saharan Africa 62, Central and South Africa 446, North Africa and West Asia 24

\*\* median under-ascertainment was 14.5 overall based on 125 study estimates and 11.9 in national estimates, 15.7 in regional estimates, and 24.0 in local estimates

**Table 5. Global representativeness**

	<b>Meyerowitz-Katz</b>	<b>Rostami</b>	<b>Bobrovitz</b>	<b>Imperial College COVID-19 response team</b>	<b>Ioannidis</b>	<b>O'Driscoll**</b>
Estimates (countries)*						
Europe	11 (11)	52 (13)	33 (13)	8 (7)	22 (21)	13 (13)
North America	3 (1)	22 (1)	1 (1)	1 (1)	15 (2)	1 (1)
Latin America	1 (1)	17 (2)	3 (1)	1 (1)	3 (3)	1 (1)
Asia	1 (1)	14 (5)	2 (1)	0 (0)	10 (9)	0 (0)
Africa	0 (0)	2 (2)	1 (1)	0 (0)	1 (1)	1 (1)
Oceania	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Information from Europe and North America	91% of weight	72% of datasets	85% of datasets (82% of countries)	90% of datasets	73% of location estimates	87% of countries
Information from Europe and America	98% of weight	85% of datasets	93% of datasets (87% of countries)	100% of datasets	78% of location estimates	94% of countries

\*geographic location of estimates (countries) included in main calculations \*\* The extrapolated 45 countries on which age-stratified IFR estimates are obtained also include countries outside the regions that have at least one country represented (Pakistan, Philippines, Bangladesh, Indonesia, China, Thailand, South Korea, Japan) even though not directly measured in any of them

## **Appendix 1. Evaluations with multiple flaws and with choices consistently inflating IFR: technical competence or bias issues**

The evaluations by ICCRT<sup>4</sup> and Meyerowitz-Katz<sup>1</sup> have multiple flaws as well as eligibility, design, and analytical choices that consistently lead to higher IFR estimates. This raises questions of technical competence and/or bias.

In multiple main media interviews and quotes Meyerowitz-Katz is presented professionally as an “epidemiologist”, but apparently he has not received yet a PhD degree as of this writing and he is still a student at the University of Wollongong in Australia. Neither he nor his co-author of the evaluation (apparently another PhD student) had published any peer-reviewed systematic review or meta-analysis on any topic prior to the pandemic. By the end of 2019, Meyerowitz-Katz had published 2 PubMed-indexed papers (both on diabetes) that had received 2 citations and 1 self-citation in Scopus. Meyerowitz-Katz is very active also in Twitter through an account called Health Nerd (56,800 tweets as of January 19, 2021). The Twitter account has interesting, smart content with strong advocacy, often supporting worthy causes. The same account has also been generating tweetorial content reviewing various COVID-19 papers, including many critical/highly negative comments on my papers, e.g. on the IFR evaluation.<sup>5</sup> For fairness, readers may consult these Twitter criticisms of Meyerowitz-Katz and of another prolific Twitter critic with highly similar views as Meyerowitz-Katz (Atomsk’s Sanakan [64,200 tweets as of January 19, 2021], self-described as “Christian; Science, Denialism Debunked, Philosophy, Manga, Death Metal, Pokémon, Immunology FTW; Fan of Bradford Hill + Richard Joyce”, also supporting several worthy causes, e.g. debunking denialism). The tweetorials have been posted in Pubpeer (<https://pubpeer.com/publications/C2A5DD4ED8B5A0B13F63A47FEC143A>). Comparison against the present manuscript may hopefully help knowledgeable readers generate an informed opinion as to the merits of arguments raised. I don’t have a personal Twitter account, but was alerted to the negative tweetorials by Meyerowitz-Katz several months ago. At that time, the name of the Twitter account owner was not obviously visible (the photo showed an unrecognizable figure with big glasses and a cat), but Meyerowitz-Katz seemed to use the Twitter account prolifically to promote his own work and criticize work contradicting his work. The identity of the Health Nerd Twitter account has become transparent now, since the owner has added a photo of him (wearing a T-shirt that writes

“Trust me, I am an epidemiologist”). The identity of the reverberating Atomsk’s Sanakan Twitter account is still unclear (to me at least) and its relationship to Meyerowitz-Katz, if any, is unknown.

Overall, one potential explanation is that the flaws of the Meyerowitz-Katz evaluation may simply reflect lack of experience and technical expertise of otherwise well-intentioned and smart authors with a heightened sense of advocacy during a serious pandemic that represents undoubtedly a major crisis. It is well-known that most published systematic reviews and meta-analyses in the literature have substantial flaws anyhow. For students performing their first evidence synthesis ever, choosing a topic that requires advanced expertise due to unusual cross-design features, difficult methodological challenges and convoluted and often erratic data, a highly-flawed final product should not be surprising. Perusal of the voluminous Twitter comments of Health Nerd similarly demonstrates immediately the wonderful enthusiasm, but also lack of adequate expertise required to conduct such analyses in any rigorous way. Nevertheless, it is worrisome that trustworthy media like Scientific American and The Guardian have espoused Meyerowitz-Katz’s views and serious organizations may guide their planning based on a flawed paper. Meyerowitz-Katz is a columnist also at the American Council on Science and Health (<https://www.acsh.org/profile/gideon-meyerowitz-katz>), a pro-industry advocacy group. He reports no conflicts of interest.

The ICCRT evaluation presents a very different case. Even though the first author had also published no peer-reviewed systematic reviews and meta-analyses prior to the pandemic, he has a stronger overall publication record, and there are also 21 other scientists co-authoring that evaluation. The author list includes senior names with unquestionable competence, especially in modeling. The ICCRT evaluation is also extremely sophisticated methodologically in many of its analytical processes. Its fatal flaws pertain to issues of more fundamental clinical epidemiology issues (e.g. excluding studies with few deaths, extreme selection bias in the choice of eligible studies, inappropriate extrapolations/generalizations). Sophisticated modeling can do absolutely nothing to salvage an evaluation once it fails at such fundamental principles.

Perhaps the ICCRT simply paid little or no attention to these issues given its traditional strength and focus on modeling. It is worrisome, however, that the ICCRT work has been extremely influential in shaping the dominant narrative and major decisions about measures taken to deal with the pandemic in the UK, USA, and many other countries. The early models of the ICCRT that

Accepted Article

assumed an IFR of 0.9% and that predicted a death toll of 2.2 million in the USA and 500 thousand in the UK were key scientific drivers supporting the choice of aggressive lockdowns. The ICCRT publication in Nature (Flaxman et al., <https://www.nature.com/articles/s41586-020-2405-7>) that claimed that “across 11 countries 3.1 (2.8–3.5) million deaths have been averted owing to interventions since the beginning of the epidemic” and that “only the effect of lockdown is identifiable, and that it has a substantial effect (81% (75–87%) reduction in  $R_t$ )” was a key driver for the re-introduction of lockdown strategies in many countries in the fall of 2020. That publication has been repeatedly challenged (see for example <https://www.nature.com/articles/s41586-020-3025-y>, <https://www.medrxiv.org/content/10.1101/2020.07.22.20160341v3>) because the results are highly dependent on the model used and assumptions made. In that Nature paper, ICCRT published the results of a model that showed that lockdown worked, while ICCRT itself had developed by that time also another model with much better fit to the data and which showed that lockdown did not work (for detailed discussion see: <https://www.medrxiv.org/content/10.1101/2020.07.22.20160341v3>). Moreover, they reported on data on 11 European countries, while they failed to report on additional European countries for which data were available and which showed little or no benefit from lockdown with either analytical model. This behavior represents a mixture of major selection bias and confirmation bias, of the same character as the biases in mis-shaping and in violating eligibility criteria in the IFR report of the ICCRT.

Overall, ICCRT seems to consistently violate basic evidentiary practices in order to defend at all cost its original narrative of very high IFR and need for draconian lockdown measures. ICCRT does not seem to have any obvious conflict with for-profit entities and it is funded by prestigious public research agencies and not-for-profit philanthropists. Given the pervasive problems that permeate its work, the influence it exerts, and the unquestionable excellence of the scientific team, ICCRT and its oversight academic/research organizations and/or funders should audit the processes with which research is conducted and disseminated within ICCRT in order to identify how such biases can be contained or mitigated.

## Appendix 2. Risk of bias assessment in COVID-19 seroprevalence studies

Meyerowitz-Katz used the scoring tool developed by Hoy et al.<sup>66</sup> that has 10 items, and Rostami as well as Bobrovitz used the Joanna Briggs scoring tool, but Rostami used an earlier version that has 10 items,<sup>65</sup> while Bobrovitz used a later version that has 9 items.<sup>64</sup> “High risk of bias” studies need to be seen with more caution, but the “low risk of bias” and “high risk of bias” tags depend on the scale used and also they depend on the rater. Importantly, understanding the study design and conduct without a full publication is difficult or impossible. Often it is difficult or impossible to do this even with a full publication available, since reporting is insufficiently complete plus there may be differences between what is reported and what was actually done. Scales of observational studies are thus among the weakest frontiers of evidence-based medicine, despite continued interesting efforts.

Meyerowitz-Katz scored all 10 checklist items on all the included studies, even though almost all of the studies included in their evaluation were just press releases or preliminary reports or preprints. Scoring on such limited information in particular the Hoy et al. tool seems excessively confident and problematic. The Joanna Briggs checklists may be easier to score, but they are also ambiguous in the absence of detailed information about a study.

Furthermore, even when there was a relatively full description of a seroprevalence study, assessments by different raters reached different conclusions. For example, Bobrovitz classified overall only 12 studies out of 338 as low risk of bias, acknowledging in many cases that the information was unclear regarding some items of the scoring checklist. Conversely, Rostami classified 44 out of 107 studies as low risk of bias and Meyerowitz-Katz classified 6 out of 16 studies as low risk of bias. Moreover, Meyerowitz-Katz made a claim that the IFR was higher in the 6 studies with low risk of bias. However, 3 of these studies were not classified as low risk of bias by Bobrovitz.

To illustrate the confusion that can arise from divergent scoring between raters, let us consider here one study, the Santa Clara study.<sup>37</sup> Bobrovitz scored the study as “yes” on 5 items, as “unclear” in 1 item, and as “no” in 3 items of the Joanna Briggs checklist. Conversely, according to Rostami the study scored “yes” on 9 out of 10 items of the earlier Joanna Briggs checklist and was classified as overall “low risk of bias”. According to Meyerowitz-Katz, the study scored “yes” on only 5 of 10 items of the Hoy et al. checklist and was thus classified as “high risk of bias”.

Accepted Article

According to Bobrovitz the study scored “no” on the following items: item 1 (appropriate sample frame), item 5 (good coverage of sample), and item 6 (valid methods used for the identification of the condition). Conversely, according to Rostami, the study scored “yes” on all these three items, but it scored “no” on an item that has been dropped from the updated Joanna Briggs tool: “Are all important confounding factors/subgroups/differences identified and accounted for?” Obviously, this question is extremely difficult to answer with confidence for any study (seroprevalence or other) and the default answer should be “no”. The essential question is to what extent residual factors affect the validity of the results, not whether all of them have been identified and accounted for, which would be a mission impossible. Also of note, Bobrovitz operationalized item 6 as “sensitivity >90% and specificity >95%”. This operationalization allows some standardization, but is rather arbitrary, and violates the original checklist item.

Overall, quality and risk of bias scoring of prevalence studies is a very difficult task and despite valiant efforts by the evaluators, it is precarious to draw any strong conclusions.

### **Appendix 3. Optimizing the calculation of a global infection fatality rate and spread of the infection**

Current estimates of infection spread and of infection fatality rate try to extrapolate from data available on a limited set of countries, with heavy representation of data from European and American countries. Extrapolations to countries that do not have direct measures of seroprevalence need to correct for differences in age structure of the population, and proportion of nursing home residents among the overall population and among COVID-19 fatalities. Other factors that need to be considered are the impact of seroreversion (later conducted seroprevalence surveys may underestimate total prior infections), re-infections, and changes over time in the testing intensity and in the use of effective or detrimental treatments and management options.

The validity of the data and inferences depend on the capacity/availability of a country/continent in testing, data acquisition and reporting. For example, most African countries have far more limited capacity. However, this does not mean that unbiased studies need to be available for all 200+ countries around the world. E.g. data from 5-10 countries in Africa would offer substantially solid evidence that should be possible to extrapolate to the other African countries. Data need to be collected also from a selection of approximately 10 countries from each of the main economic strata other than high-income countries (low-income countries, lower-middle-income countries, upper-middle income countries). For countries with limited resources, some centralized planning and execution of such studies, e.g. under the auspices of the World Health Organization, may be needed. All age subgroups need to be properly represented in such evaluations and efforts should be made to avoid under-participation of some population strata that are at higher risk of infection (see Table 2 regarding deficiencies in general population strata). Definition of COVID-19 deaths is a particularly contentious issue and efforts at standardization would be useful. Granular detailed information on risk profile of COVID-19 fatalities not only in terms of comorbidities, but also regarding their severity and the overall prior functional capacity and life expectancy of the deceased would help map more appropriately the death burden and person-years lost. Crude estimates of infection fatality rate may be over-emphasizing burden of disease, if COVID-19 deaths occur in people with minimal life expectancy. E.g. the infection fatality rate for an 80-year old person may be vastly (10-100-fold)

different in a community dwelling 80-year old versus an 80-year old nursing home resident in palliative hospice care.

Accepted Article