medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The infection fatality rate of COVID-19 inferred from seroprevalence data

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

E-mail: jioannid@stanford.edu

Funding: METRICS has been supported by a grant from the Laura and John Arnold Foundation

Disclosures: I am a co-author (not principal investigator) of one of the 23 seroprevalence studies.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

ABSTRACT

Objective To estimate the infection fatality rate of coronavirus disease 2019 (COVID-19) from data of seroprevalence studies.

Methods Population studies with sample size of at least 500 and published as peer-reviewed papers or preprints as of June 7, 2020 were retrieved from PubMed, preprint servers, and communications with experts. Studies on blood donors were included, but studies on healthcare workers were excluded. The studies were assessed for design features and seroprevalence estimates. Infection fatality rate was estimated from each study dividing the number of COVID-19 deaths at a relevant time point by the number of estimated people infected in each relevant region. Correction was also attempted accounting for the types of antibodies assessed.

Results 23 studies were identified with usable data to enter into calculations. Seroprevalence estimates ranged from 0.1% to 47%. Infection fatality rates ranged from 0.02% to 0.86% (median 0.26%) and corrected values ranged from 0.02% to 0.78% (median 0.25%). Among people <70 years old, infection fatality rates ranged from 0.00% to 0.26% with median of 0.05% (corrected, 0.00-0.23% with median of 0.04%). Most studies were done in pandemic epicenters and the few studies done in locations with more modest death burden also suggested lower infection fatality rates.

Conclusions The infection fatality rate of COVID-19 can vary substantially across different locations and this may reflect differences in population age structure and case-mix of infected and deceased patients as well as multiple other factors. Estimates of infection fatality rates inferred from seroprevalence studies tend to be much lower than original speculations made in the early days of the pandemic.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The infection fatality rate (IFR), the probability of dying for a person who is infected, is one of the most critical and most contested features of the coronavirus disease 2019 (COVID-19) pandemic. The expected total mortality burden of COVID-19 is directly related to the IFR. Moreover, justification for various non-pharmacological public health interventions depends crucially on the IFR. Some aggressive interventions that potentially induce also more pronounced collateral harms[1] may be considered appropriate, if IFR is high. Conversely, the same measures may fall short of acceptable risk-benefit thresholds, if the IFR is low.

Early data from China, adopted also by the World Health Organization (WHO),[2] focused on a crude case fatality rate (CFR) of 3.4%; CFR is the ratio of COVID-19 deaths divided by the number of documented cases, i.e. patients with symptoms who were tested and found to be PCR-positive for the virus. The WHO envoy who visited China also conveyed the message that there are hardly any asymptomatic infections.[3] With a dearth of asymptomatic infections, the CFR approximates the IFR. Other mathematical models suggested that 40-70%,[4] or even[5] 81% of the global population would be infected. Influential mathematical models[5,6] eventually dialed back to an IFR of 1.0% or 0.9%, and these numbers long continued to be widely cited and used in both public and scientific circles. The most influential of these models, constructed by Imperial College estimated 2.2 million deaths in the USA and over half a million deaths in the UK in the absence of lockdown measures.[5] Such grave predictions justifiably led to lockdown measures adopted in many countries. With 0.9% assumed infection fatality rate and 81% assumed proportion of people infected, the prediction would correspond to a global number of deaths comparable with the 1918 influenza, in the range of 50 million fatalities.

Since late March 2020, many studies have tried to estimate the extend of spread of the virus in various locations by evaluating the seroprevalence, i.e. how many people in population samples have developed antibodies for the virus. These studies can be useful because they may

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

inform about the extend of under-ascertainment of documenting the infection based on PCR testing. Moreover, they can help obtain estimates about the IFR, since one can divide the number of observed deaths by the estimated number of people who are inferred to have been infected.

At the same time, seroprevalence studies may have several caveats in their design, conduct, and analysis that may affect their results and their interpretation. Here, data from the first presented full papers (either peer-reviewed or preprint) as of June 7, 2020 were collected, scrutinized, and used to infer estimates of IFR in different locations where these studies have been conducted.

METHODS

Seroprevalence studies

The input data for the calculations of IFR presented here are studies of seroprevalence of COVID-19 that have been done in the general population, or in samples that might approximate the general population (e.g. with proper reweighting) and that have been published in peer-reviewed journals or have been submitted as preprints as of June 7, 2020. Only studies with at least 500 assessed samples were considered, since smaller datasets would entail extremely large uncertainty for any calculations to be based on them. Studies where results were only released through press releases were not considered here, since it is very difficult to tell much about their design and analysis, and this is fundamental in making any inferences based on their results. Some key ones that have attracted large attention (e.g. Spain seroprevalence) are nevertheless considered in the Discussion. Preprints should also be seen with caution since they have not been yet fully peer-reviewed (although some of them have already been revised based on very extensive comments from the scientific community). However, in contrast to press releases, preprints typically offer at least a fairly complete paper with information about design and analysis. Studies done of blood donors were eligible, although it is possible they may underestimate seroprevalence

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

and overestimate IFR due to healthy volunteer effect. Studies done on health care workers were not, since they deal with a group at potentially high exposure risk which may lead to seroprevalence estimates much higher than the general population and thus implausibly low IFR. Searches were made in PubMed (LitCOVID), medRxiv, bioRxiv, and Research Square using the terms "seroprevalence" and "antibodies" as of June 7, 2020. Communication with colleagues who are field experts sought to ascertain if any major studies might have been missed.

Information was extracted from each study on location, recruitment and sampling strategy, dates of sample collection, sample size, types of antibody used (IgG, IgM, IgA), estimated crude seroprevalence (positive samples divided by all samples test), and adjusted seroprevalence and features that were considered in the adjustment (sampling process, test performance, presence of symptoms, other).

Calculation of inferred IFR

Information on the population of the relevant location was collected from the papers. Whenever it was missing, it was derived based on recent census data trying to approximate as much as possible the relevant catchment area (e.g. region(s) or county(ies)), whenever the study did not pertain to an entire country. Some studies targeted specific age groups (e.g. excluding elderly people and/or excluding children) and some of them made inferences on number of people infected in the population based on specific age groups. For consistency, the entire population, as well as, separately, only the population with age <70 years were used for estimating the number of infected people. It was assumed that the seroprevalence would be similar in different age groups, but significant differences in seroprevalence according to age strata that had been noted by the original authors were also recorded to examine the validity of this assumption.

The number of infected people was calculated multiplying the relevant population with the adjusted estimate of seroprevalence. Whenever an adjusted seroprevalence estimate had not been

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

obtained, the unadjusted seroprevalence was used instead. When seroprevalence estimates with different adjustments were available, the analysis with maximal adjustment was selected. When seroprevalence studies had used sequential waves of testing over time, data from the most recent wave was used, since it would give the most updated picture of the epidemic wave.

For the number of COVID-19 deaths, the number of deaths recorded at the time chosen by the authors of each study was selected, whenever the authors used such a death count up to a specific date to make inferences themselves. If the choice of date had not been done by the authors, the number of deaths accumulated until after 1 week of the mid-point of the study period was chosen. This accounts for the differential delay in developing antibodies versus dying from the infection. It should be acknowledged that this is an averaging approximation, because some patients may die very soon (within <3 weeks) after infection (and thus are overcounted), and others may die very late (and thus are undercounted due to right censoring).

The inferred IFR was obtained by dividing the number of deaths by the number of infected people for the entire population, and separately for people <70 years old. The proportion of COVID-19 deaths that occurred in people <70 years old was retrieved from situational reports for the respective countries, regions, or counties in searches done in June 3-7. A corrected IFR is also presented, trying to account for the fact that only one or two types of antibodies (among IgG, IgM, IgA) might have been used. Correcting seroprevalence upwards (and inferred IFR downwards) by 1.1-fold for not performing IgM measurements and similarly for not performing IgA measurements may be reasonable, based on some early evidence,[7] although there is uncertainty about the exact correction factor.

RESULTS

Seroprevalence studies

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

23 studies with a sample size of at least 500 have been published either in the peer-reviewed literature or as preprints as of June 7, 2020.[8-30] Dates and processes of sampling and recruitment are summarized in Table 1, sample sizes, antibody types assessed and regional population appear in Table 2, estimated prevalence, and number of people infected in the study region are summarized in Table 3, and number of COVID-19 and inferred IFR estimates are found in Table 4. Four studies (Geneva,[10] Rio Grande do Sul[17], Zurich[28], and Milan[29]) performed repeated seroprevalence surveys at different time points, and only the most recent one is shown in these tables.

Of the 23 studies, only 6 found some modest differences in seroprevalence rates across some age groups (Oise: decreased seroprevalence in age 0-14, increased in age 15-17; Geneva: decreased seroprevalence in age >50; Netherlands: increased seroprevalence in age 18-30; New York state: decreased seroprevalence in age >55; Brooklyn: decreased seroprevalence in age 0-5, increased in age 16-20; Tokyo: increased seroprevalence in age 18-34). The patterns are not strong enough to suggest major differences in extrapolating across age groups, although higher values in adolescents and young adults and lower values in elderly individuals cannot be excluded.

As shown in Table 1, these studies varied substantially in sampling and recruitment designs. The main issue is whether they can offer a representative picture of the population in the region where they are performed. A generic problem is that vulnerable people who are at high risk of infection and/or death may be more difficult to recruit in survey-type studies. COVID-19 infection seems to be particularly widespread and/or lethal in nursing homes, among homeless people, in prisons, and in disadvantaged minorities. Most of these populations are very difficult, or even impossible to reach and sample from and they are probably under-represented to various degrees (or even entirely missed) in surveys. This would result in an underestimation of seroprevalence and thus overestimation of IFR. Seven studies (Iran,[8] Geneva,[10] Gangelt,[16] Ro

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Grande do Sul[17], Luxembourg[20], Los Angeles county[22], Brazil [133 cities][25]) explicitly aimed for random sampling from the general population. In principle, this is a stronger design. However, even with such designs, people who cannot be reached (e.g. by e-mail or phone or even visiting them at a house location) will not be recruited, and these vulnerable populations are likely to be missed.

Six the 23 studies assessed blood donors in Denmark,[12] Netherlands,[15] Scotland,[18] the Bay Area in California,[24] Zurich/Lucerne,[28] and Milan.[29] By definition these studies include people in good health and without symptoms, at least recently, and therefore may markedly underestimate COVID-19 seroprevalence in the general population. A small set of 200 blood donors in Oise, France[13] showed 3% seroprevalence, while pupils, siblings, parents, teachings and staff at a high school with a cluster of cases in the same area had 25.9% seroprevalence.

For the other studies, healthy volunteer bias may lead to underestimating seroprevalence and this is likely to have been the case in at least one case (the Santa Clara study)[19] where wealthy healthy people were rapidly interested to be recruited when the recruiting Facebook ad was released. The design of the study anticipated correction with adjustment of the sampling weights by zip code, gender, and ethnicity, but it is likely that healthy volunteer bias may still have led to some underestimation of seroprevalence. Conversely, attracting individuals who might have been concerned of having been infected (e.g. because they had symptoms) may lead to overestimation of seroprevalence in surveys. Finally studies of employees, grocery store clients, or patient cohorts (e.g. hospitalized for other reasons, or coming to the emergency room) may have sampling bias with unpredictable direction.

As shown in Table 2, all studies have tested for IgG antibodies, but only 9 have also assessed IgM and 3 have assessed IgA. Only one study assessed all three types of antibodies. All studies considered the results to be “positive” if any tested antibody type was positive, with the

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

exception of one study (Luxembourg) that considered the results to be "positive" only if both IgG and IgA were detected. The ratio of people sampled versus the total population of the region was better than 1:1000 in only 6 studies (Idaho,[9] Denmark blood donors,[12] Santa Clara,[19] Luxembourg[20], Brooklyn[27], Zurich[28]), which means that the estimates can have substantial uncertainty.

Seroprevalence estimates

As shown in Table 3, prevalence ranged from as little as 0.1% to as high as 47%. Studies varied a lot on whether they tried or not to adjust their estimates for test performance, sampling (striving to get closer to a more representative sample), and clustering effects (e.g. when including same household members) as well as other factors. The adjusted seroprevalence occasionally differed substantially from the crude, unadjusted value. In principle adjusted values are likely to be closer to the true estimate, but the exercise shows that each study alone may have some unavoidable uncertainty and fluctuation, depending on the analytical choices preferred. In studies that sampled people from multiple locations, large between-location heterogeneity could be seen (e.g. 0-25% across 133 Brazilian cities)[25].

Inferred IFR

Inferred IFR estimates ranged from 0.02% to 0.86% (median 0.26%) and corrected values ranged from 0.02% to 0.78% (median 0.25%). Corrected values exceeding 0.4% were inferred for Netherlands, Milan, Luxembourg, and in one study in New York state. The first two were extrapolated from blood donors data, therefore the IFR may be overestimated, while for New York, another study found a much higher seroprevalence in Brooklyn and thus a lower inferred IFR. Conversely, very low or low IFR (corrected, 0.02-0.07%) was seen in two studies in Japan (Kobe and Tokyo), one in Iran, and one in France.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The proportion of COVID-19 deaths that occurred in people <70 years old varied substantially across locations. All deaths in Gangelt were in elderly people while in Wuhan half the deaths occurred in people <70 years old and the proportion might have been higher in Iran, but no data could be retrieved for this country. When limited to people <70 years old, IFR ranged from 0.00% to 0.26% with median of 0.05% (corrected, 0.00-0.23% with median of 0.04%). All IFR estimates in people <70 years old did not exceed 0.1%, with the exception of New York, Wuhan and Milan.

DISCUSSION

Inferred IFR values based on emerging seroprevalence studies typically show a much lower fatality than initially speculated in the earlier days of the pandemic. It should be appreciated that IFR is not a fixed physical constant and it can vary substantially across locations, depending on the population structure, the case-mix of infected and deceased individuals and other, local factors.

The 23 studies analyzed here are the first to be made available in full papers and they are not fully representative of all countries and locations around the world. Most of them come from locations with overall COVID-19 mortality rates exceeding the global average (60 per million people as of June 7). The median inferred IFR in the locations with a COVID-19 mortality rate below the global average (two studies in Japan, Idaho, Croatia and Wuhan) is 0.15% (corrected, 0.13%), while it is substantially higher for studies done in epicenters with high population mortality rate.

Several studies in hard-hit European countries inferred modestly high IFR estimates for the overall population, but the IFR was still low in people <70 years old. Some of these studies were on blood donors and may have underestimated seroprevalence and overestimated IFR. One study in Germany aimed to test the entire population of a city and thus selection bias is minimal:

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Gangelt[16] represents a situation with a superspreader event (in a local carnival) and 7 deaths were recorded, all of them in very elderly individuals (average age 81, sd 3.5). COVID-19 has a very steep age gradient of death risk.[31] It is expected therefore that in locations where the infection finds its way into killing predominantly elderly citizens, the overall, age-unadjusted IFR would be higher. However, IFR would still be very low in people <70 in these locations, e.g. in Gangelt IFR is 0.00% in non-elderly people. Similarly, in Switzerland, 69% of deaths occurred in people >80 years old[31] and this explains the higher age-unadjusted IFR in Geneva and Zurich. Similar to Germany, very few deaths in Switzerland have been recorded in non-elderly people, e.g. only 2.5% have occurred in people <60 years old and IFR in that age-group would be ~0.01%. The majority of deaths in most of the hard hit European countries have happened in nursing homes[32] and a large proportion of deaths also in the US[33] also follow this pattern. Moreover, many nursing home deaths have no laboratory confirmation and thus should be seen with extra caution in terms of the causal impact of SARS-CoV-2.

Locations with high burdens of nursing home deaths may have high IFR estimates, but the IFR would still be very low among non-elderly, non-debilitated people. The average length of stay in a nursing home is slightly more than 2 years and people who die in nursing homes die in a median of 5 months[34] so many COVID-19 nursing home deaths may have happened in people with life expectancy of only a few months. This needs to be verified in careful assessments of COVID-19 outbreaks in nursing homes with detailed risk profiling of fatalities. If COVID-19 happened in patients with very limited life expectancy, this pattern may even create a dent of less than expected mortality in the next 3-6 months after the coronavirus excess mortality wave. As of June 7 (week 22), preliminary Euromonitor data[35] indeed already show a substantial dent below baseline mortality in France, and a less prominent dent below baseline mortality in Italy and several other European countries.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The estimated IFR of 0.31 in Wuhan may reflect the wide spread of the infection to hospital personnel and the substantial contribution of nosocomial infections to a higher death toll;[36] plus unfamiliarity with how to deal with the infection in the first location where COVID-19 arose. Massive deaths of elderly individuals in nursing homes, nosocomial infections, and overwhelmed hospitals may also explain the very high fatality in specific locations in Northern Italy[37] and New York. The highest IFR among these 23 studies was seen in a study in Milan.[29] Although the estimate may be inflated in that blood donor study, probably IFR was truly very high in Milan. That same study estimated that 2.7% of the Milan population had already been infected by the time the outbreak was first recognized. Another study of seroprevalence in health care workers and administrative hospital staff in Lombardy[38] found 8% seroprevalence in Milan hospitals and 35-43% in Bergamo hospitals, supporting the scenario for widespread nosocomial infections among vulnerable patients. The high IFR values in New York are also not surprising, given the vast death toll witnessed. A very unfortunate decision of the governors in New York and New Jersey was to have COVID-19 patients sent to nursing homes. Moreover, some hospitals in New York City hotspots reached maximum capacity and perhaps could not offer optimal care. Use of unnecessarily aggressive management (e.g. mechanical ventilation) and hydroxychloroquine may also have contributed to worse outcomes. Furthermore, New York City has an extremely busy, congested public transport system that may have exposed large segments of the population to high infectious load in close contact transmission and, thus, perhaps more severe disease. A more aggressive viral clade has also been speculated, but this needs further verification.[39] Of note, two seroprevalence studies in New York City[23,27] give substantially different results. This discrepancy demonstrates the need to conduct different studies with different sampling strategies offering complementary results.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

It may not be surprising that IFR may reach very high levels among disadvantaged populations and settings that have the worst combination of factors predisposing to higher fatalities. One may predict also very high IFRs in other select locations with atypically high death toll, e.g. Bergamo or Brescia in Italy,[37] and several locations in Belgium, England, or Spain. Preliminary press released data from Belgium (6.9% seroprevalence),[40] the United Kingdom (6.8% seroprevalence),[41] and from the ongoing Spanish national study (5.2% seroprevalence)[42] agree with this expectation. With 5.2% IgG seroprevalence in Spain, crude IFR is 1.11% (corrected to 0.70% after considering sensitivity of 79% and antibody type), but with wide variability across locations, e.g. corrected IFR is 0.73% in Madrid versus 0.17% in Canary Islands. Importantly, hotspot locations are rather uncommon exceptions in the global landscape. Moreover, even in these locations, the IFR for non-elderly individuals without predisposing conditions may remain very low. E.g. in New York City only 0.6% of all deaths happened in people <65 years without major underlying conditions.[43] Thus the IFR even in New York City would probably be lower than 0.01% in these people.

Studies with extremely low inferred IFR, Kobe, Tokyo and Oise, are also worthwhile discussing. For Kobe, the authors of the study[11] raise the question whether COVID-19 deaths have been undercounted in Japan. Both undercounting and overcounting of COVID-19 deaths is likely to be a caveat in different locations and this is difficult to settle in the absence of very careful scrutiny of medical records and autopsies. The Tokyo data[30], nevertheless, also show similarly very low IFR. Moreover, evaluation of all-cause mortality in Japan has shown no excess deaths during the pandemic, consistent with the possibility that somehow the Japanese population was spared. Former immunity from exposure to other coronaviruses, genetic differences, and other unknown factors may be speculated. IFR seems to be very low also in some other Asian countries where, in contrast to Japan, extensive PCR testing was carried out. For example, as of June 7,

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

2020, in Singapore there were only 25 deaths among 37,910 cases, suggesting an upper bound of 0.07% for IFR, even if no cases had been missed. A similar picture is seen in Qatar. The Wuhan study IFR of 0.31% seems a high-value outlier for Asia. In fact, a smaller study of asymptomatic Hubei returnees[44] (n=452, not included in the calculations given sample size <500) showed 4% seroprevalence and would translate to much lower IFR.

For the Oise sample,[13] it is possible that it may not be representative of the general population. As discussed above, there is a large difference in the estimated seroprevalence between the high school-based sampling and a small dataset of blood donors from the same area, and the true seroprevalence value may be somewhere between these two extremes that may be biased in opposite directions.

Some seroprevalence studies have also been designed to assess seroprevalence repeatedly spacing out measurements in the same population over time. Preliminary data from Brazil[17] are still early to judge for meaningful increases, but the data from Geneva suggest that seroprevalence increased more than 3-fold over three weeks.[10] An increase was seen also in Zurich[28] and in Milan.[29] Notably, the increase corresponds to continued infections during a period where strict social distancing and other lockdown measures were implemented. Data from Finland,[45] with repeated measurements over several weeks (available at the Finnish Institute website, but not submitted as full paper yet) conversely show fairly steady seroprevalence in a country that maintained a much lower overall death burden. Preliminary data from the ongoing national Spain serostudy[42] also show no major increase in seroprevalence after relaxation of lockdown measures. Serial seroprevalence measurements may offer some evidence on whether different measures were associated with curbed transmission or not, and how these might translate to different IFR values. Any causal inferences need to be extremely cautious. However, it is expected that measures that manage to avoid transmission of the virus to vulnerable high-risk populations may lead to lower

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

values of IFR. Measure packages that do not protect these high-risk populations may lead to higher values of IFR. Serial seroprevalence measurements would also provide evidence on how quickly antibody titers decrease below detection. If decrease is fast, as suggested by some preliminary data,[46] numbers of infected people may be underestimated and IFR overestimated.

The only data from a low-income country among the 23 studies examined here come from Iran[8] and the IFR estimate appears to be the same or lower than the IFR of seasonal influenza. Iran has a young population with only slightly over 1% of the age pyramid at age >80. The same applies to almost every less developed country around the world. Given the very sharp age gradient and the sparing of children and young adults from death by COVID-19, one may expect COVID-19 IFR to be fairly low in the less developed countries. However, it remains to be seen whether comorbidities, poverty and frailty (e.g. malnutrition) may have adverse impact on risk and thus increase IFR also in these countries.

One should caution that the extent of validation of the antibody assays against positive and negative controls differs across studies. Specificity has typically exceeded 99.0%, which is reassuring. However, for very low prevalence rates, even 99% specificity may be problematic. The study with the lowest estimated prevalence (Brazil)[17] has nevertheless evaluated also family members of the people who tested positive and found several family members were also infected, thus suggesting that most of the positive readings are true rather than false positives. Sensitivity also varies from 60-100% in different validation exercises and for different tests, but typically it is closer to the upper than the lower bound. One caveat about sensitivity is that typically the positive controls are patients who had symptoms and thus were tested and found to be PCR-positive. However, it is possible that symptomatic patients may be more likely to develop antibodies than patients who are asymptomatic or have minimal symptoms and thus had not sought PCR testing.[47-49] Since the seroprevalence studies specifically try to unearth these asymptomatic/mildly

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

symptomatic missed infections, a lower sensitivity for these mild infections could translate to substantial underestimates of the number of infected people and substantial overestimate of the inferred IFR.

The corrected IFR estimates are trying to account for undercounting of infected people when not all 3 antibodies (IgG, IgM, and IgA) are assessed.[7] However, the magnitude of the correction is uncertain and may also vary in different circumstances. Moreover, it is possible that an unknown proportion of people may have handled the virus using immune mechanisms (mucosal, innate, cellular) that did not generate any serum antibodies.[50,51] This would lead to an unknown magnitude of underestimation of the frequency of the infection and a respective overestimation of the IFR. At least one study has found indeed that mild SARS-CoV-2 infections may lead to nasal release of IgA, without serum antibody response.[52]

An interesting observation is that even under congested circumstances, like cruise ships, aircraft carriers or homeless shelter, the proportion of people infected does not get to exceed 20-45%.[53,54] Similarly, at a wider population level, values ~47% are the maximum values documented to-date and most values are much lower, yet epidemic waves seem to wane. It has been suggested[55,56] that differences in host susceptibility and behavior can result in herd immunity at much lower prevalence of infection in the population than originally expected. COVID-19 spreads by infecting certain groups more than others because some people have much higher likelihood of exposure. People most likely to be exposed also tend to be those most likely to spread for the same reasons that put them at high exposure risk. In the absence of random mixing of people, the epidemic wave may be extinguished even with relatively low proportions of people becoming infected. Seasonality may also play a role in the dissipation of the epidemic wave. It has also been observed that about 50% of people have CD4 cellular responses to SARS-CoV-2 even without being exposed to this virus and this may be due to prior exposure to other coronaviruses.[57]

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

It is unknown whether this proportion varies in different populations around the world and whether this immunity may contribute to SARS-CoV-2 epidemic waves waning without infecting a large share of the population.

A major limitation of the current analysis is that the calculations presented in this paper include several preprints that have not yet been fully peer-reviewed. Moreover, there is a substantially larger number of studies that have made press releases about their results and probably several more will become available in the near future. Those that include or allow calculation of IFR estimates in their press releases or preliminary agency reports seem to have values that are similar to those of the 23 studies analyzed here, and most estimates are quite low (e.g. 0.16% in Slovenia, 0.23% in Stockholm, 0.14% in Israel, 0.00% in San Miguel county in Colorado). However, some estimates are high (e.g. Spain)[42]. Obviously these preliminary results require extreme caution. The plan is to try to update this analysis with new emerging data. More clean, vetted data may make the overall picture more crisp and allow having more granularity on the determinants that lead to higher or lower IFR in different locations.

A comparison of COVID-19 to influenza is often attempted, but many are confused by this comparison unless placed in context. Based on the IFR estimates obtained here, COVID-19 may have infected as of June 7 approximately 200 million people (or more), far more than the ~7 million PCR-documented cases. The global COVID-19 death toll is still evolving, but it is still similar to a typical death toll from seasonal influenza (290,000-650,000),[58] while "bad" influenza years (e.g. 1957-9 and 1968-70) have been associated with 1-4 million deaths.[59] Notably, influenza devastates low-income countries, but is more tolerant of wealthy nations, probably because of the availability and wider use of vaccination in these countries.[58] Conversely, in the absence of vaccine and with a clear preference for elderly debilitated individuals, COVID-19 may have an inverse death toll profile, with more deaths in wealthy nations than in low-income

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

countries. However, even in the wealthy nations, COVID-19 seems to affect predominantly the frail, the disadvantaged, and the marginalized – as shown by high rates of infectious burden in nursing homes, homeless shelters, prisons, meat processing plants, and the strong racial/ethnic inequalities against minorities in terms of the cumulative death risk.[60,61]

While COVID-19 is a formidable threat, the fact that its IFR is typically much lower than originally feared, is a welcome piece of evidence. The median of 0.26% found in this analysis is very similar to the estimate recently adopted by CDC for planning purposes.[62] The fact that IFR can vary substantially also based on case-mix and settings involved also creates additional ground for evidence-based, more precise management strategies. Decision-makers can use measures that will try to avert having the virus infect people and settings who are at high risk of severe outcomes. These measures may be possible to be far more precise and tailored to specific high-risk individuals and settings than blind lockdown of the entire society. Of course, uncertainty remains about the future evolution of the pandemic, e.g. the presence and height of a second wave.[63] However, it is helpful to know that SARS-CoV-2 has relatively modest IFR overall and that possibly IFR can be made even lower with appropriate, precise non-pharmacological choices.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 1. Seroprevalence studies on COVID-19 published or depositing preprints as of June 7, 2020: dates, sampling and recruitment process

Location	Dates	Sampling and recruitment
Iran (Guilan)[8]	April (until April 21)	Population-based cluster random sampling design through phone call invitation, household-based.
Idaho (Boise)[9]	Late April	People from the Boise, Idaho metropolitan area, part of the Crush the Curve initiative.
Switzerland (Geneva)[10]	April 20-27	Randomly selected previous participants of the Bus Santé study with an email (or phone contact, if e-mail unavailable); participants were invited to bring all members of their household, aged 5 years and older.
Japan (Kobe)[11]	March 31-April 7	Randomly selected patients who visited outpatient clinics and received blood testing for any reason. Patients who visited the emergency department or the designated fever consultation service were excluded.
Denmark blood donors[12]	April 6-17	All Danish blood donors aged 17-69 years giving blood. Blood donors are healthy and must comply with strict eligibility criteria; they must self-defer for two weeks if they develop fever with upper respiratory symptoms.
France (Oise)[13]	March 30-April 4	Pupils, their parents and siblings, as well as teachers and non-teaching staff of a high-school.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

China (Wuhan)[14]	April 3-15	People applying for a permission of resume (n=1,021) and hospitalized patients during April 3 to 15 (n=381).
Netherlands blood donors[15]	April 1-15	Blood donors. Donors must be completely healthy, but they may have been ill in the past, provided that they recovered at least two weeks before.
Germany (Gangelt)[16]	March 30-April 6	600 adult persons with different surnames in Gangelt were randomly selected, and all household members were asked to participate in the study.
Brazil (Rio Grande do Sul)[17]	April 25-27	Multi-stage probability sampling was used in each of 9 cities to select 500 households, within which one resident was randomly chosen for testing.
Scotland blood donors[18]	March 21-23	Blood donors. Donors should not have felt unwell in the last 14 days, also some other deferrals applied regarding travel and COVID-19 symptoms.
California (Santa Clara)[19]	April 2-3	Facebook ad with additional targeting by zip code.
Luxembourg[20]	April 16-May 5	Representative sample (no details how ensured), 1807 of 2000 contacted provided data, were <79 years and had serology results.
Germany (Frankfurt)[21]	April 6-14	Employees of Infraserv Höchst, a large industrial

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

		site operator in Frankfurt am Main. No exclusion criteria.
California (Los Angeles)[22]	April 10-14	Proprietary database representative of the county. A random sample of these residents was invited, with quotas for enrollment for subgroups based on age, sex, race, and ethnicity distribution.
New York[23]	April 19-28	Convenience sample of patrons ≥18 years and residing in New York State, recruited consecutively upon entering 99 grocery stores and via an in-store flyer.
California (Bay Area)[24]	March	1,000 blood donors in diverse Bay Area locations (excluding those with self-reported symptoms or abnormal vital signs)
Brazil[25]	May 15-22	Sampling from 133 cities (the main city in each region), selecting 25 census tracts with probability proportionate to size in each sentinel city, and 10 households at random in each tract. Aiming for 250 participants per city.
Croatia[26]	April 23-28	DIV factory workers in Split and Sibenik-Knin invited for voluntary testing
New York (Brooklyn)[27]	Early May	Patients seen in urgent care facility in Brooklyn
Switzerland (Zurich)[28]	Prepandemic until May (patients) and April (blood	Patients at the University Hospital of Zurich and blood donors in Zurich and Lucerne

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

	donors)	
Italy (Milan) blood donors[29]	February 24 – April 8	Blood donors (free of recent symptoms possibly related to COVID-19, no close contact with confirmed cases, and, since March 26th, symptom-free during the preceding 14 days and no unprotected contacts with suspected cases. Age 18-70 years
Japan (Tokyo)[30]	April 21-May 20	Two community clinics located in the major railway stations in Tokyo (Navitas Clinic Shinjuku and Tachikawa)

Two of the studies included additional datasets of <500 participants that are not presented here (n=200 blood donors in Oise and n=387 patients in the California (Bay Area) study)

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 2. Sample size, types of antibodies, and population in relevant region

Location	Sample size	Antibody	Population in region*	Population <70 years (%)
Iran (Guilan)	551	IgG/IgM	2354848	95
Idaho (Boise)	4856	IgG	481587 (Ada county)	92
Switzerland (Geneva)	576	IgG	500000	88
Japan (Kobe)	1000	IgG	1518870	79 (Japan)
Denmark blood donors	9496	IgG/IgM	5771876*	86
France (Oise)	661	IgG	5978000 (Hauts-de-France)	89
China (Wuhan)	1401	IgG/IgM	11080000	93 (China)
Netherlands blood donors	7361	IgG/IgM/IgA	17097123	86
Germany (Gangelt)	919	IgG/IgA	12597	86
Brazil (Rio Grande do Sul)	4500	IgG	11377239	91
Scotland blood donors	500	IgG	5400000	88
California (Santa Clara)	3300	IgG/IgM	1928000	90
Luxembourg	1807	IgG/IgA**	615729	90
Germany (Frankfurt)	1000	IgG	2681000***	84 (Germany)
California (Los Angeles)	863	IgG/IgM	7892000	92
New York	15101	IgG	19450000	90
California (Bay Area)	1000	IgG	7753000	90
Brazil (133 cities)	24995	IgG/IgM	74656499	94 (Brazil)
Croatia	1494	IgG/IgM	4076000	86
New York (Brooklyn)	11092	IgG	2559903	91

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Switzerland (Zurich)	2773 patients	IgG	1520968 (canton Zurich)	88
	2524 blood donors		1930525 (Zurich+Lucerne)	
Italy (Milan) blood donors	789	IgG/IgM	3260000	82
Japan (Tokyo)	1071	IgG	13902077	79 (Japan)

*The authors of some studies preferred to focus on age-restricted populations: 17-70 years old in the Denmark blood donor study (n=3800000), those 18-79 years old in the Luxembourg study (n=483000); those <70 years old in Netherlands blood donor study (n=13745768); and those >=18 years old in the New York state study (n=15280000).
**considered positive if both IgG and IgA were positive
***participants were recruited from a large number of districts, but most districts had very few participants; here the population of the 9 districts with >1:10,000 sampling ratio is included (846/1000 participants came from these 9 districts).

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 3. Prevalence of infection and estimated number of infected people

Location	Seroprevalence (%)		Estimated
	Crude	Adjusted (adjustments)	infected
Iran (Guilan)	22.0	33.0 (test, sampling)	770000
Idaho (Boise)	1.79	ND	8620
Switzerland (Geneva)	8.7	9.7 (test, age, sex)	48500
Japan (Kobe)	3.3	2.7 (age, sex)	40999
Denmark blood donors	1.8	1.7 (test)	98122
France (Oise)	25.9	ND	1548000
China (Wuhan)	10.0	ND	1108000
Netherlands blood donors	2.7	ND	461622
Germany (Gangelt)	15.0	20.0 (test, cluster, sym)	2519
Brazil (Rio Grande do Sul)	0.133	0.309 (sampling, city size)*	35153
Scotland blood donors	1.2	ND	64800
California (Santa Clara)	1.5	2.6 (test, sampling, cluster)	51000
Luxembourg	1.9	2.06 (age, sex, district)	12684
Germany (Frankfurt)	0.6	ND	16086
California (Los Angeles)	4.06	4.65 (test, sex, race/ethnicity, income)	367000
New York	12.5	14.0 (test, sex, age race/ethnicity, region)	2723000
California (Bay Area)	0.4 (blood donors)	0.1 (test/confirmation)	7753

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Brazil (133 cities)	1.39	1.62 overall, varying from 0 to 25.0 across 133 cities (test, design)	1209435**
Croatia	1.27***	ND	51765
New York (Brooklyn)	47.0	ND	1203154
Switzerland (Zurich)	unclear	1.3 in patients in May and 1.1 in blood donors in April (multivariate Gaussian conditioning)	19773 (Zurich) 21236 (Zurich+Lucerne)
Italy (Milan) blood donors	7.1 (last 3 weeks)	5.2 (test)	169520
Japan (Tokyo)	3.83	ND	532450

*an estimate is also provided adjusting for test performance, but the assumed specificity of 99.0% seems inappropriate, since as part of the validation process the authors also found that several of the test-positive individuals had household members who were also infected, thus the estimated specificity was deemed to be at least 99.95%
**the authors calculate 760000 infected in the 90 cities that had 200-250 samples tested, but many of the other 43 cities with <200 samples may be equally or ever better represented, since they tended to be smaller than the 90 (mean population 356213 versus 659326)
***1.20% in workers in Split without mobility restrictions, 3.37% in workers in Knin without mobility restrictions, 1.57% for all workers without mobility restrictions; Split and Knin tended to have somewhat higher death rates than nation-wide Croatia, but residence of workers is not given, so the entire population of the country is used in the calculations
test: test performance; ND: no data available

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 4. Inferred infection fatality rates

Location	COVID-19 deaths (date)	Inferred IFR (corrected), %	COVID-19 deaths <70 years,*** %	IFR in <70 years (corrected), %
Iran (Guilan)	617 (4/23)	0.08 (0.07)	No data	No data
Idaho (Boise)	14 (4/24)	0.16 (0.13)	14 (Idaho)	0.02 (0.02)
Switzerland (Geneva)	243 (4/30)	0.50 (0.40)	8	0.05 (0.04)
Japan (Kobe)	10 (mid-April)	0.02 (0.02)	21 (Japan)	0.01 (0.01)
Denmark blood donors	370 (4/21)	0.38 (0.34)	12	0.05 (0.05)
France (Oise)	635 (4/7)	0.04 (0.03)	11 (France, <65 years)	0.01 (0.01)
China (Wuhan)	3869 (5/2)	0.35 (0.31)	50	0.19 (0.15)
Netherlands blood donors	3134 (4/15)	0.68 (0.68)	11	0.09 (0.09)
Germany (Gangelt)	7 (4/15)	0.28 (0.25)	0	0.00 (0.00)
Brazil (Rio Grande do Sul)	50 (4/30)	0.14 (0.11)	31 (Brazil, <60 years)	0.05 (0.04)
Scotland blood donors	47 (4/1)	0.07 (0.06)	9 (<65 years)	0.01 (0.01)
California (Santa Clara)	94 (4/22)	0.18 (0.17)	35	0.07 (0.06)
Luxembourg	92 (5/2)	0.73 (0.58)	9	0.07 (0.06)
Germany (Frankfurt)	42*(4/17)	0.26 (0.21)	14 (Germany)	0.04 (0.03)
California (Los Angeles)	724 (4/19)	0.20 (0.18)	24 (<65 years)	0.06 (0.05)
New York	18610 (4/30)^	0.68 (0.54)^	34	0.26 (0.23)^
California (Bay Area)	12 (3/22)	0.15 (0.12)	25	0.04 (0.03)

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Brazil (133 cities)	**	Median 0.30 (0.27)	31 (<60 years)	0.10 (0.9)
Croatia	79 (5/3)	0.15 (0.14)	13	0.02 (0.02)
New York (Brooklyn)	4894 (5/19)^	0.41 (0.33)^	34 (NY State)	0.15 (0.14)^
Switzerland (Zurich)	127 (5/15 Zurich), 125 (4/22, Zurich+Lucerne)	0.60 (0.48)	8 (Switzerland)	0.06 (0.05)
Italy (Milan) blood donors	1459 (3/31)	0.86 (0.77)	15 (Italy)	0.15 (0.13)
Japan (Tokyo)	189 (5/11)	0.04 (0.03)	21 (Japan)	0.01 (0.01)

*approximated from number of deaths in the Hesse province on 4/17 times the proportion of deaths in the 9 districts with key enrollment in the study among all Hesse province deaths.
**data are provided by the authors for deaths per 100,000 population in each city along with inferred IFR in each city, with wide differences across cities; the IFR shown here is the median across the 36 cities with 200-250 samples and at least one positive sample (the interquartile range for the uncorrected IFR is 0.20-0.60 and the full range across all cities is 0-2.4%, but with very wide uncertainty in each city).
*** whenever the number/proportion of COVID-19 deaths at age <70 years was not provided in the paper, the proportion of these deaths was retrieved from situational reports of the relevant location (searched in June 3-7); when this was not possible to find for the specific location, a larger geographic entity was used. For example, for Italy (Milan) blood donors the number of deaths in Milan aged <70 are approximated by multiplying the total number of deaths with the proportion of COVID-19 deaths <70 years old across all Italy (15%, no specific data available for Milan). For Brazil, the closet information that could be found was from a news report (https://www.thejakartapost.com/news/2020/05/22/in-brazil-covid-19-hitting-young-people-harder.html). For Croatia, data on age for 45/103 deaths were retrieved through Wikipedia.
^Confirmed deaths; inclusion of probable deaths would increase the IFR estimates by about a quarter.
IFR: infection fatality rate. The inferred IFR is derived by dividing the number of accumulated deaths (at the time chosen by the authors of each study, or until after 1 week of the mid-point of the study dates, whenever the authors had not arbitrated on a date for the death count) by the estimated number of infected people. The corrected IFR is obtained from the inferred IFR assuming that, as compared with assessing IgG, IgM, and IgA, 20% of the infections are missed when only IgG is assessed, and 10% of the infections are missed when two of the three antibodies are assessed.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

References:

1. Melnick ER, Ioannidis JPA. Head to Head: Should governments continue lockdown to slow the spread of covid-19? BMJ 2020;369:m1924. doi: 10.1136/bmj.m1924.
2. https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---3-march-2020. Accessed May 10, 2020.
3. World Health Organization. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19) 16-24 February 2020. Published 2020-02-28. Available from: https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19). Accessed May 10, 2020.
4. McGinty JC. How many people might one person with coronavirus infect? Wall Street Journal. February 14, 2020. https://www.wsj.com/articles/how-many-people-might-one-person-with-coronavirus-infect-11581676200. Accessed February 27, 2020.
5. https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf
6. Meyerowitz-Katz G, Merone L 2020 A systematic review and meta-analysis of published research data on COVID-19 infection-fatality rates doi: https://doi.org/10.1101/2020.05.03.20089854
7. Alberto L Garcia-Basteiro, Gemma Moncunill, Marta Tortajada, Marta Vidal, Caterina Guinovart, Alfons Jimenez, Rebeca Santano, Sergi Sanz, et al. Seroprevalence of antibodies against SARS-CoV-2 among health care workers in a large Spanish reference hospital. medRxiv. doi: https://doi.org/10.1101/2020.04.27.20082289
8. Shakiba M, Nazari S, Mehrabian F, et al. Seroprevalence of COVID-19 virus infection in Guilan province, Iran. medRxiv doi: https://doi.org/10.1101/2020.04.26.20079244

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

9. 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 Testing in Idaho. J Clin Microbiol. 2020 doi:10.1128/JCM.00941-20
10. Stringhini S, Wisniak A, Piumatti G, Azman AS, Lauer SA, Baysson H, et al. Repeated seroprevalence of anti 1 -SARS-CoV-2 IgG antibodies in a population-based sample. medRxiv https://doi.org/10.1101/2020.05.02.20088898
11. Doi A, Iwata Kuroda H, Hasuike T, Nasu S, Kanda A, et al. Seroprevalence of novel coronavirus disease (COVID-19) in Kobe, Japan. medRxiv https://doi.org/10.1101/2020.04.26.20079822
12. Erikstrup C, Hother CE, Pedersen OB, Mølbak K, Skov RL, Holm DK. Estimation of SARS-CoV-2 infection fatality rate by real-time antibody screening of blood donors. medRxiv 2020
13. Fontanet A, Tondeur L, Madec Y, Grant R, Besombes C, Jolly N, et al. Cluster of COVID-19 in northern France: A retrospective closed cohort study. medRxiv 2020 https://doi.org/10.1101/2020.04.18.20071134
14. Wu X, Fu B, Chen L, Feng Y. Serological tests facilitate identification of asymptomatic SARS□CoV□2 infection in Wuhan, China. J Med Virol 2020 https://onlinelibrary.wiley.com/doi/10.1002/jmv.25904
15. Slot E, Hogema BM, Reusken CBEM, Reimerink JH, Molier M, Karregat HM, et al. Herd immunity is not a realistic exit strategy during a COVID-19 outbreak. Research Square 2020. https://dx.doi.org/10.21203/rs.3.rs-25862/v1
16. Streeck H, Schulte B, Kümmerer BM, Richter E, Höller T, Fuhrmann C, et al. Infection fatality rate of SARS-CoV-2 infection in a German community with a super-spreading event. medRxiv 2020 https://doi.org/10.1101/2020.05.04.20090076

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

17. Silveira MF, Barros AJD, Horta BL, Pellanda LC, Dellagostin O, Struchiner CJ, et al. Repeated population-based surveys of antibodies against SARS2-CoV-2 in Southern Brazil. medRxiv 2020 https://doi.org/10.1101/2020.05.01.20087205
18. 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 https://doi.org/10.1101/2020.04.13.20060467
19. Bendavid E, Mulaney B, Sood N, et al. COVID-19 Antibody Seroprevalence in Santa Clara County, California. medRxiv doi: https://doi.org/10.1101/2020.04.14.20062463
20. 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 2020.05.11.20092916; doi: https://doi.org/10.1101/2020.05.11.20092916
21. Kraehling V, Kern M, Halwe S, Mueller H, Rohde C, Savini M, et al. Epidemiological study to detect active SARS-CoV-2 infections and seropositive persons in a selected cohort of employees in the Frankfurt am Main metropolitan area. medRxiv 2020.05.20.20107730; doi: https://doi.org/10.1101/2020.05.20.20107730
22. Sood N, Simon P, Ebner P, Eichner D, Reynolds J, Bendavid E, Bhattacharya J. Seroprevalence of SARS-CoV-2-Specific Antibodies Among Adults in Los Angeles County, California, on April 10-11, 2020. JAMA. 2020 May 18:e208279.
23. 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. medRxiv 2020; doi: https://doi.org/10.1101/2020.05.25.20113050
24. Ng D, Goldgof G, Shy B, Levine A, Balcerek J, Bapat SP, et al. SARS-CoV-2 seroprevalence and neutralizing activity in donor and patient blood from the San Francisco

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Bay Area. medRxiv 2020.05.19.20107482; doi: https://doi.org/10.1101/2020.05.19.20107482

25. Hallal PC, Hartwig FP, Horta BL, Victora GD, Silveira MF, Struchiner CJ, et al. Remarkable variability in SARS-CoV-2 antibodies across Brazilian 1 regions: nationwide serological household survey in 27 states. medRxiv 2020 doi: https://doi.org/10.1101/2020.05.30.20117531
26. Jerkovic I, Ljubic T, Basic Z, Kruzic I, Kunac N, Bezic J, et al. SARS-CoV-2 antibody seroprevalence in industry workers in Split-Dalmatia and Sibenik-Knin County, Croatia. medRxiv 2020.05.11.20095158; doi: https://doi.org/10.1101/2020.05.11.20095158
27. Reifer J, Hayum N, Heszkel B, Klagsbald I, Streva VA. SARS-CoV-2 IgG Antibody Responses in New York City. medRxiv 2020.05.23.20111427; doi: https://doi.org/10.1101/2020.05.23.20111427
28. Emmenegger M, De Cecco E, Lamparter D, Jacquat RPB, Ebner D, et al. Population-wide evolution of SARS-CoV-2 immunity tracked by a ternary immunoassay. medRxiv 2020.05.31.20118554; doi: https://doi.org/10.1101/2020.05.31.20118554
29. Valenti L, Bergna A, Pelusi S, Facciotti F, Lai A, Tarkowski M, et al. SARS-CoV-2 seroprevalence trends in healthy blood donors during the COVID-19 Milan outbreak. medRxiv 2020.05.11.20098442; doi: https://doi.org/10.1101/2020.05.11.20098442
30. Takita M, Matsumura T, Yamamoto K, Yamashita E, Hosoda K, Hamaki T, Kusumi E. Regional Difference in Seroprevalence of SARS-CoV-2 in Tokyo: Results from the community point-of-care antibody testing. medRxiv 2020.06.03.20121020; doi: https://doi.org/10.1101/2020.06.03.20121020
31. Ioannidis JPA, Axfors C, Contopoulos-Ioannidis DG. Population-level COVID-19 mortality risk for non-elderly individuals overall and for non-elderly individuals without

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

underlying diseases in pandemic epicenters. medRxiv. 2020. doi:10.1101/2020.04.05.20054361

32. https://www.theguardian.com/world/2020/apr/13/half-of-coronavirus-deaths-happen-in-care-homes-data-from-eu-suggests. Accessed April 27, 2020.
33. American Geriatrics Society. American Geriatrics Society Policy Brief: COVID-19 and Nursing Homes. J Am Geriatr Soc. 2020 Apr 8. doi: 10.1111/jgs.16477.
34. Kelly A, Conell-Price J, Covinsky K, Stijacic Cenzer I, Chang A, Boscardin WJ, et al. Lengths of Stay for Older Adults Residing in Nursing Homes at the End of Life. J Am Geriatr Soc. 2010 Sep; 58(9): 1701–1706.
35. EUROMOMO. Graphs and maps: Excess mortality. Updated week 19, 2020. Accessed May 10, 2020. https://www.euromomo.eu/graphs-and-maps/
36. Wang D, Hu B, Hu C, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA. 2020 Feb 7:e201585.
37. Boccia S, Ricciardi W, Ioannidis JPA. What Other Countries Can Learn From Italy During the COVID-19 Pandemic. JAMA Intern Med. 2020 Apr 7. doi: 10.1001/jamainternmed.2020.1447.
38. Sandri MT, Azzolini E, Torri V, Carloni S, Tedeschi M, Castoldi M, Mantovani A, Rescigno M. IgG serology in health care and administrative staff populations from 7 hospital representative of different exposures to SARS-CoV-2 in Lombardy, Italy. medRxiv doi: https://doi.org/10.1101/2020.05.24.20111245
39. Brofsky A. Distinct clades of SARS-CoV-2: implications for modeling of viral spread. J Med Virol. 2020 Apr 20. doi: 10.1002/jmv.25902. https://www.wxxinews.org/post/1-every-8-nyers-may-have-had-covid-19-testing-shows

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

40. https://www.uantwerpen.be/en/research-groups/vaxinfectio/corona-research/sero-epidemiology-co/. Accessed, June 7, 2020.

41. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/5june2020

42. https://www.isciii.es/Noticias/Noticias/Paginas/Noticias/PrimerosDatosEstudioENECOVID19.aspx. Accessed, June 6, 2020.

43. https://www1.nyc.gov/assets/doh/downloads/pdf/imm/covid-19-daily-data-summary-deaths-04212020-1.pdf. Accessed April 21, 2020.

44. To KK, Cheng VC, Cai J-P, Chan K-H, Chen L-L, Wong L-H, et al. Seroprevalence of SARS-CoV-2 in Hong Kong and in residents evacuated from Hubei province, China: a multicohort study. Lancet Microbe 2020 DOI:https://doi.org/10.1016/S2666-5247(20)30053-7

45. https://www.thl.fi/roko/cov-vaestoserologia/sero_report_weekly.html

46. Rosado J, Cockram C, Merkling SH, Demeret C, Meola A, Kerneis S, Terrier B, Fafi-Kremer S, et al. Serological signatures of SARS-CoV-2 infection: Implications for antibody-based diagnostics. medRxiv 2020. doi: https://doi.org/10.1101/2020.05.07.20093963

47. Wu F, Wang A, Liu M, et al. Neutralizing antibody responses to SARS-CoV-2 in a COVID-19 recovered patient cohort and their implications. medRxiv 2020;2020.03.30.20047365.

48. Edouard S, Colson P, Melenotte C, De Pinto F, Thomas L, La Scola B, et al. Evaluating the serological status of COVID-19 patients using an indirect immunofluorescent assay, France.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

49. Solbach W, Schiffner J, Backhaus I, Burger D, Staiger R, Tiemer B, et al. Antibody profiling of COVID-19 patients in an urban low-incidence region in Northern Germany. medRxiv 2020.05.30.20111393; doi: https://doi.org/10.1101/2020.05.30.20111393

50. Kikkert M. Innate immune evasion by human respiratory RNA viruses. J Innate Immun 2020;12:4–20

51. Krammer F. The human antibody response to influenza A virus infection and vaccination. Nature Reviews Immunology 2019;19:1.

52. Cervia C, Nilsson J, Zurbuchen Y, Valaperti A, Schreiner J, Wolfensberger A, et al. Systemic and mucosal antibody secretion specific to SARS-CoV-2 during mild versus severe COVID-19. bioRxiv 2020 doi: https://doi.org/10.1101/2020.05.21.108308

53. Moriarty LF, Plucinski MM, Marston BJ, Kurbatova EV, Knust B, Murray EL, et al. Public Health Responses to COVID-19 Outbreaks on Cruise Ships - Worldwide, February-March 2020. MMWR Morb Mortal Wkly Rep. 2020 Mar 27;69(12):347-352.

54. Maxmen A. Coronavirus is spreading under the radar in US homeless shelters. Nature. 2020 May;581(7807):129-130.

55. Britton T, Trapman P, Ball FG. The disease-induced herd immunity level for Covid-19 is substantially lower than the classical herd immunity level. doi: https://doi.org/10.1101/2020.05.06.20093336

56. Gomes GM, Corder RM, King JG, Langwig KE, Souto-Maior C, Carneiro J, et al. Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold. medRxiv https://doi.org/10.1101/2020.04.27.20081893

57. Grifoni A, Weiskopf D, Ramirez SI, et al. Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. Cell 2020;181:1-13.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253; this version posted June 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

58. Paget J, Spreeuwenberg P, Charu V, Taylor RJ, Iuliano AD, Bresee J, et al. Global mortality associated with seasonal influenza epidemics: New burden estimates and predictors from the GLaMOR Project. J Glob Health. 2019 Dec; 9(2): 020421.

59. Pandemic Influenza Preparedness and Response: A WHO Guidance Document. Geneva, World Health Organization, 2009.

60. Webb Hooper M, Nápoles AM, Pérez-Stable EJ. COVID-19 and Racial/Ethnic Disparities. JAMA. 2020 May 11. doi: 10.1001/jama.2020.8598.

61. Centers for Disease Control and Prevention. COVID-19 in Racial and Ethnic Minority Groups. Updated April 24, 2020. Accessed April 30, 2020. https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/racial-ethnic-minorities.html

62. https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html, Accessed June 6, 2020.

63. Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. Science. 2020 Apr 14:eabb5793. doi: 10.1126/science.abb5793.