medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253.this version posted May 19, 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(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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

Conflicts of interest: None

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253.this version posted May 19, 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(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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 May 12, 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 Twelve studies were identified with usable data to enter into calculations. Seroprevalence estimates ranged from 0.113% to 25.9% and adjusted seroprevalence estimates ranged from 0.309% to 33%. Infection fatality rates ranged from 0.03% to 0.50% and corrected values ranged from 0.02% to 0.40%.

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 May 19, 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 continue to be widely cited and used in both public and scientific circles as of this writing (May 12, 2020). 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 May 19, 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 May 12, 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 May 12, 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. New York 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. Studies done on health care workers were not, since they deal

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253.this version posted May 19, 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 .

with a group at potentially high exposure risk. Searches were made in PubMed, LitCOVID, medRxiv, bioRxiv, and Research Square using the terms "seroprevalence" and "antibodies" as of May 12, 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. For two studies on blood donors where the authors stated that they would not extrapolate on people outside the recruited age group of 17-70 and on people >70 years old, respectively, all calculations were made on the 17-70 and =<70 years old groups, respectively.

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 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 to estimate seroprevalence themselves. If this had not been done by the authors, the number of deaths within 1

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253.this version posted May 19, 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 .

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.

The inferred IFR was obtained by dividing the number of deaths by the number of infected people. 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

Twelve studies with a sample size of at least 500 have been published either in the peer-reviewed literature or as preprints as of May 12, 2020.[8-19] Dates and processes of sampling and recruitment are summarized in Table 1, sample sizes, antibody types assessed and regional population appear in Table 2, and estimated prevalence, number of people infected in the study region, and inferred IFR are summarized in Table 3. Two studies (Geneva[10] and Rio Grande do Sul[17]) performed repeated seroprevalence surveys at different time points, and only the most recent one is shown in these tables.

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

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253.this version posted May 19, 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 .

degrees (or even entirely missed) in surveys. This would result in an underestimation of seroprevalence and thus overestimation of IFR. Four studies (Iran,[8] Geneva,[10] Gangelt,[16] and Ro Grande do Sul[17]) 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.

Three of the twelve studies assessed blood donors in Denmark,[12] Netherlands,[15] and Scotland.[18] 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 high school 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.

As shown in Table 2, all studies have tested for IgG antibodies, but only 5 have also assessed IgM and 2 have assessed IgA. Only one study assessed all three types of antibodies. The ratio of people sampled versus the total population of the region was better than 1:1000 in only

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253.this version posted May 19, 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 .

three studies (Idaho,[9] Denmark blood donors,[12] Santa Clara[19]), which means that the estimates can have substantial uncertainty.

Seroprevalence estimates

As shown in Table 3, crude prevalence ranged from as little as 0.133% to as high as 25.7%. 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). 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.

Inferred IFR

Interestingly, despite their differences in design, execution, and analysis, most studies provide IFR point estimates that are within a relatively narrow range. Seven of the 12 inferred IFRs are in the range 0.07 to 0.20 (corrected IFR of 0.06 to 0.16) which are similar to IFR values of seasonal influenza. Three values are modestly higher (corrected IFR of 0.25-0.40 in Gangelt, Geneva, and Wuhan) and two are modestly lower than this range (corrected IFR of 0.02-0.03 in Kobe and Oise).

DISCUSSION

Inferred IFR values based on emerging seroprevalence studies show a much lower fatality than initially speculated in the earlier days of the pandemic. Many IFR estimates are in the range of seasonal influenza IFR, but some are higher, and some others are lower than this range. 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.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253.this version posted May 19, 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 three higher values (corrected IFR of 0.25-0.40) are in Gangelt, Geneva, and Wuhan. Gangelt[16] represents a situation with a superspreader event (in a local carnival) and 7 deaths were recorded in the city, all of them in very elderly individuals (average age 81, sd 3.5). COVID-19 is known to have a very steep age gradient of death risk.[20] 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 less than 70 in these locations, e.g. in Gangelt IFR is 0.000 in non-elderly people. Similarly, in Switzerland, 69% of the deaths have occurred in people >80 years old[20] and this explains the higher age-unadjusted IFR in Geneva, which was considered a paradise for spending one's last years until the COVID-19 struck. 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 the IFR in that age-group would be in the range of 0.01%. The majority of deaths in most of the hard hit European countries have happened in nursing homes[21] and a large proportion of deaths also in the US[22] also seem to follow this pattern. Moreover, a very large number of these nursing home deaths have no laboratory confirmation and thus they need to 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 an median of 5 months[23] so it is likely that 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 the fatalities. If this pattern is prominent, it may even create a dent of less than expected mortality in the very elderly age stratum (e.g. >85 years) in the next 3-6 months following the passage of the coronavirus

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253.this version posted May 19, 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 .

excess mortality wave. Perusal of Euromonitor data[24] and similar national mortality death curves would be interesting to undertake in this regard.

Finally, 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.[25] It may also reflect 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 seen in specific locations in Northern Italy[26] and in New York and New Jersey. 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. With large proportions of medical and paramedical personnel infected, it is possible that nosocomial infections increased the death toll. Use of unnecessarily aggressive management (e.g. mechanical ventilation) 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.[27] These factors may explain why preliminary press-released information[28] on a seroprevalence survey in New York State suggests a much higher IFR. With 20% estimated crude seroprevalence in New York City, including a range between 17.3% in Manhattan to 27.6% in Bronx[28] (adjusted seroprevalence figures have not been released), IFR would be as high as 0.8% in Bronx and 1% in Queens, and even higher if probable COVID-19 deaths are included in the calculations. 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

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253.this version posted May 19, 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 .

fatalities. One may predict also very high IFRs in other select locations with atypically high death toll, e.g. Bergamo or Brescia in Italy.[26] However, these locations are very 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.[29] Thus the IFR even in New York City would probably be lower than 0.01% in these people.

The two studies with extremely low inferred IFR, Kobe 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. 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.

A few seroprevalence studies have also been designed to assess seroprevalence repeatedly spacing out measurements in the same population over time. Preliminary data from Southern Brazil[17] are still early to judge for meaningful increases, but the data from Geneva suggest that seroprevalence increased more than 3-fold over a period of three weeks.[10] This is interesting, because the increase corresponds to continued infections during a period where strict social distancing and other lockdown measures were implemented. Data from Finland[30], 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. More repeated measures results may give some stronger evidence on

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253.this version posted May 19, 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 .

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 values of IFR. Measure packages that do not protect these high-risk populations may lead to higher values of IFR.

The only data from a less developed country among the 12 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

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253.this version posted May 19, 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 .

testing.[31,32] Since the seroprevalence studies specifically try to unearth these asymptomatic/mildly 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.[33,34] This would lead to an unknown magnitude of underestimation of the frequency of the infection and a respective overestimation of the IFR.

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%.[35,36] Similarly, at a wider population level, values ~33% are the maximum values documented to-date. It has been suggested[37,38] 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.

A major limitation of the current analysis is that the calculations presented in this paper depend largely on preprints that have not yet been fully peer-reviewed. Moreover, there is a

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253.this version posted May 19, 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 .

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 seem to have values that are similar to those of the 12 studies analyzed here, and most estimates are in the range of seasonal influenza (e.g. 0.20 in Los Angeles county, 0.16 in Slovenia, 0.18 in Stockholm, 0.00 in San Miguel county in Colorado), but obviously these 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 this may be an uneven comparison. At a very broad, bird's eye view level, worldwide the IFR of COVID-19 this season may be in the same ballpark as the IFR of influenza (0.1%, 0.2% in a bad year). According to this scenario, which needs further verification, COVID-19 may have infected as of May 12 approximately 200 million people (or more), far more than the ~4.2 PCR-documented cases. However, influenza devastates developing countries, but is more tolerant of wealthy nations, probably because of the availability and wider use of vaccination in these countries.[39] 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 the developing world. 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.[40,41]

While COVID-19 is a formidable threat, the fact that its IFR is much lower than originally feared, is a welcome piece of evidence. The fact that its IFR can vary substantially also based on

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253.this version posted May 19, 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 .

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.[42] However, it is helpful to know that SARS-CoV-2 has relatively low IFR overall and that possibly its 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 May 19, 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 May 12, 2020: dates, sampling and recruitment process

Location	Dates	Sampling and recruitment
Iran (Guilan)[8]	April	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 May 19, 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.

*population <70 years old, considered relevant.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253.this version posted May 19, 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
Iran (Guilan)	551	IgG/IgM	2354848
Idaho (Boise)	4856	IgG	392365 (Ada county)
Switzerland (Geneva)	576	IgG	5000000
Japan (Kobe)	1000	IgG	1518870
Denmark blood donors	9496	IgG/IgM	3800000*
France (Oise)	661	IgG	5978000
China (Wuhan)	1401	IgG/IgM	11080000
Netherlands blood donors	7361	IgG/IgM/IgA	13745768**
Germany (Gangelt)	919	IgG/IgA	12597
Brazil (Rio Grande do Sul)	4500	IgG	11377239
Scotland blood donors	500	IgG	5400000
California (Santa Clara)	3300	IgG/IgM	1928000

*population 17-70 years old **population <70 years

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253.this version posted May 19, 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 and inferred infection fatality rates

Location	Prev	Adjusted prev (adjustments)	Estimated infected	Deaths (date)	Inferred IFI (corrected)
Iran (Guilan)	22.0	33.0 (test, samp)	770000	617 (4/23)	0.08 (0.07)
Idaho (Boise)	1.79	ND	7023	14 (4/24)	0.20 (0.16)
Switzerland (Geneva)	8.7	9.7 (test, age, sex)	48500	243 (4/30)	0.50 (0.40)
Japan (Kobe)	3.3	2.7 (age, sex)	40999	10 (mid-April)	0.02 (0.02)
Denmark blood donors	1.8	1.7 (test)	64600*	53 (4/21)*	0.08 (0.07)*
France (Oise)	25.9	ND	1548000	635 (4/7)	0.04 (0.03)
China (Wuhan)	10.0	ND	1108000	3869 (5/2)	0.35 (0.31)
Netherlands blood donors	2.7	ND	371119**	344 (4/15)**	0.09 (0.09)**
Germany (Gangelt)	15.0	20.0 (test, cluster, sym)	2519	7 (4/15)	0.28 (0.25)
Brazil (Rio Grande do Sul)	0.133	0.309 (samp, city size)	35153	50 (4/30)	0.14 (0.11)
Scotland blood donors	1.2	ND	64800	47 (4/1)	0.07 (0.06)
California (Santa Clara)	1.5	2.6 (test, samp, cluster)	51000	94 (4/22)	0.18 (0.17)

*population 17-70 years; **population <70 years; Prev: prevalence; samp: sampling; sym: symptoms; test: test performance; ND: no data available; IFR: infection fatality rate. The inferred IFR is derived by dividing the number of deaths (at the time chosen by the authors of each study, or within 1 week of the mid-point of the study dates, whenever the authors had not arbitrated on 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 May 19, 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 (in press).
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

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253.this version posted May 19, 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 .

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
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. medRxiv https://doi.org/10.1101/2020.04.27.20082362
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

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253.this version posted May 19, 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 .

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
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. 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. medRxiv. 2020. doi:10.1101/2020.04.05.20054361
21. https://www.theguardian.com/world/2020/apr/13/half-of-coronavirus-deaths-happen-in-care-homes-data-from-eu-suggests. Accessed April 27, 2020.
22. 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.
23. 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.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253.this version posted May 19, 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 .

24. EUROMOMO. Graphs and maps: Excess mortality. Updated week 19, 2020. Accessed May 10, 2020. https://www.euromomo.eu/graphs-and-maps/

25. 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.

26. 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.

27. 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

28. https://www.wxxinews.org/post/1-every-8-nyers-may-have-had-covid-19-testing-shows

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

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

31. 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 [Internet] 2020;2020.03.30.20047365.

32. 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.

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

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253.this version posted May 19, 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 .

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

35. 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.

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

37. 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

38. 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

39. 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.

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

41. 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

medRxiv preprint doi: https://doi.org/10.1101/2020.05.13.20101253.this version posted May 19, 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 .

42. 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.