medRxiv preprint doi: https://doi.org/10.1101/2021.04.08.21255101; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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-ND 4.0 International license .

Estimating COVID-19 infection fatality rate in Mumbai during 2020

Murad Banaji[1], April 8, 2021

Abstract

The aim of this piece is to estimate the infection fatality rate (IFR) of COVID-19 in Mumbai during 2020, namely the fraction of SARS-CoV-2 infections which resulted in death. The estimates are based on the city's seroprevalence and fatality data. This task runs into difficulties connected with obtaining reliable estimates of both the numerator (fatalities) and the denominator (infections). In the light of the many uncertainties, it makes most sense to present a range of values with transparent discussion of assumptions and possible biases. We find that the estimates are consistent with meta-analyses of COVID-19 age-stratified IFR.

1. Introduction and main results

Arriving at estimates of the infection fatality rate (IFR) associated with a COVID-19 epidemic presents challenges associated with estimating both fatalities and infections. Given widespread evidence of COVID-19 death underreporting, both in India [1] and worldwide [2], ignoring uncertainties associated with fatalities can introduce a major downward bias into IFR estimates.

Although Mumbai's data is far from complete, the city remains one of the few locations in India which has seen several serosurveys, and where some limited all cause mortality data is available. Taking into account the various uncertainties, we obtain for the city's IFR during 2020 a median value of 0.23%, a range of $(0.10\%, 0.44\%)$, and a 95% confidence interval of $(0.15\%, 0.33\%)$. The assumptions are presented transparently, so it is easy to explore how the estimates change if, for example, we assume a higher fraction of excess deaths were COVID-19 deaths.

The estimates tally with available international data. The median estimate is close to predictions from the meta-analysis in O'Driscoll *et al* [3], while the meta-analysis of Levin *et al* [4] predicts an IFR value close to the upper end of the 95% confidence interval. Note, howewver, that the predictions from meta-analyses are made on the assumption of even spread across genders and age-groups, whereas there is some evidence of lower levels of infection amongst men, and strong evidence of lower levels of infection amongst older groups. Thus Mumbai's theoretical IFR in the event of even spread across age and gender seems likely to be higher than the estimates for 2020.

2. Basic notions

Seroprevalence surveys to estimate the prevalence of antibodies to SARS-CoV-2 ("serosurveys" for short) provide the main data used to estimate levels of prior infection. Mumbai has seen several serosurveys but the two carried out in July 2020 and August 2020 by the city corporation in collaboration with the Tata Institute of Fundamental Research [5, 6] were carefully planned and are accompanied by useful technical information.

[1]Middlesex University, London, Department of Design Engineering and Mathematics.

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/10.1101/2021.04.08.21255101; this version posted April 10, 2021. 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-ND 4.0 International license .

These are referred to as the city's first and second serosurveys. The first one forms the cornerstone of the analysis here.

Sensitivity and specificity of antibody tests to prior infection is important to consider. Given some number k of infections all occurring on a given day, we can ask how many of these would be picked up by serology using some given assay on n days later. We expect sensitivity to peak two to three weeks after infection, and then decline; however the speed of the decline depends strongly on the assay [7]. In particular, the test used in Mumbai's first and second serosurveys has fairly rapidly declining sensitivity.

Prevalence is used to refer to the fraction of the population who have been infected by SARS-CoV-2 and "recovered". Since estimates of prevalence come from serosurveys, the time to recovery is defined as the time to seroconversion. The **infection rate** refers to the number of SARS-CoV-2 recoveries plus deaths as a fraction of the total population. Note that if there can be reinfections, an infection rate of 50% does not necessarily imply that 50% of the population has been infected, and a value of over 100% is not in itself absurd. We assume that:

- Reinfections were relatively few by the time of the first serosurvey (July, 2020).
- Prevalence values estimated from serosurveys hold for the entire population, even when children below some age were omitted from a serosurvey.

Infection fatality rate (IFR) is the ratio of deaths to total infections, assuming that all those infected have either recovered or died. In practice, if we ignore deaths in the denominator this makes a very marginal difference to the estimates. For example, if there have been 3 deaths for every 1000 recoveries, removing the deaths from the denominator changes IFR from 0.299% to 0.3%.

Naïve IFR refers to any IFR estimate which ignores possible fatality underreporting and uses *recorded* fatalities in the numerator. Naïve IFR calculations are useful; but it is important to flag up the risk of downward bias if naïve IFR is used as a proxy for IFR.

Fatality delay refers to the delay used when estimating naïve IFR. If we have an estimate of prevalence on a given date and we assume a fatality delay of n days, then we would use recorded fatalities n days later when computing naïve IFR. The fatality delay needs to take into account possible delays in fatality recording. For example, suppose we estimate seroprevalence from an IgG antibody test, and suppose that IgG seroconversion typically occurs 14 days after symptom onset [8]. Using a fatality delay of 7 days is equivalent to assuming that deaths are typically *recorded* 21 days after symptom onset. Mumbai is known to have had major fatality recording delays and data reconciliations [9].

3. Recorded COVID-19 mortality and excess mortality in Mumbai during 2020

According to data released by the municipal corporation, Mumbai saw 11,116 recorded COVID-19 deaths in 2020. This number rose slowly through January, reaching 11,351 by the end of the month. On the other hand, Mumbai's all cause mortality data for the year 2020, released in a "Vital Statistics" report by the corporation [10], showed 21,850 more deaths than the previous 5 year average, amounting to a 24% rise in mortality (data in Appendix A). In other words, for each recorded COVID-19 death in 2020, the city saw close to one additional "excess" death in 2020.

There is evidence that that some forms of mortality actually fell in 2020. For example:

1. According to the Vital Statistics Report [10], infant mortality fell by 33%, perhaps largely as a consequence of a 23% drop in births in 2020. This means that there were around 1312 fewer infant deaths in 2020 compared to the previous 5 year average.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.08.21255101; this version posted April 10, 2021. 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-ND 4.0 International license .

2. According to the Hindustan Times [11], deaths on Mumbai's suburban railway lines dropped by 1575 (58%) from 2,691 in 2019 to 1,116 in 2020 (no data is given prior to 2019).

3. According to the Times of India [12], deaths on Mumbai's roads dropped by 149 (36%) from 413 in 2019 to 264 in 2020 (no data is given prior to 2019).

These three drops in mortality summed up give roughly 3000 fewer deaths in 2020 compared to earlier years in the categories of infant deaths, and rail and road accidents.

The fall in some kinds of mortality during 2020 is also consistent with a pattern noted in the early part of the year. Although month-by-month data is not publicly available, according to a report in IndiaSpend [13], there were significant falls of 24% and 16% in the city's all cause mortality during March and April 2020 relative to the previous three year average. This period corresponds to an early phase of the city's epidemic, when restrictions (including national lockdown on March 25) were coming into effect, and the COVID surge was accelerating but had not yet had time to translate into large scale fatalities. Some fall in mortality could be due to registration delays; but it could also indicate that mitigation was reducing mortality in categories such as rail and road accidents discussed above.

On the other hand, according to this report in The Indian Express [14], May and June 2020 saw huge year-on-year all cause mortality rises of 106% and 70% respectively. This was during a period when the city's cases peaked and began a slow decline. The city's health infrastructure was severely overstretched as described in this report in Article-14 [15], likely leading to additional COVID and non-COVID deaths.

To summarise: until data is available on other categories of deaths which may have been affected by NPIs – e.g., deaths from transmissible diseases other than COVID-19, pollution related deaths, and deaths related to economic distress – a broad-minded view is that there were between 11,000 and 25,000 COVID-19 deaths in the city during 2020. The median estimate of 18,000 corresponds to around 64% of those excess deaths not officially classified as COVID-19 deaths actually being COVID-19 deaths.

4. Estimating COVID-19 infections in Mumbai by the time of the first serosurvey

Mumbai's first serosurvey [5] surveyed residents 12 years or older in 3 of the city's 24 wards (F/N, M/W and R/N) for IgG antibodies to SARS-CoV-2. With an approximate mid-point of July 8, 2020, it returned seroprevalences of 54.1% (52.7% – 55.5%) in the slums surveyed, and 16.0% (14.8% – 17.2%) in the nonslum areas surveyed. These values were following reweighting to account for demographic characteristics (the raw values, inferred from Tables e2 and e6 in [5] were 57.0% and 15.7%), but were not corrected for test sensitivity or specificity of the assay used. The 95% CIs given by the authors reflect possible sampling errors.

According to 2011 data (given in Appendix A) 52.5% of the city lives in the slums. Thus slum and nonslum seroprevalence values of 54% and 16% respectively, if they held citywide, would imply 36% seroprevalence in the city at the time of the first serosurvey.

Although the serosurvey was carefully designed and had adequate sample sizes, there are various possible biases which need to be considered, especially if we hope to extrapolate to the city as a whole.

1. *Possible underestimation of prevalence by igoring sensitivity of tests.* Correcting for sensitivity and specificity, the authors of [5] estimate prevalence values upto 58.3% in the slums and upto 17.1% in nonslum areas (Table 1 in the supplementary material of [5]). These higher values would lead to a citywide prevalence estimate of 39%.

2. *Possible overestimation of prevalence by choice of slums.* The largest slums in each ward were selected – this could have introduced a bias. There are reasons why, for example, a large slum might see earlier or more rapid spread of infection.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.08.21255101; this version posted April 10, 2021. 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-ND 4.0 International license .

3. *Possible bias with unclear direction introduced by the choice of wards.* The authors of [5] note that wards were selected for surveying based on the consideration that "we had NGO partners operating in the ward's slums and the Municipal Corporation had health officers that could accompany phlebotomists as they surveyed non-slums". It is not clear how this might affect prevalence in the wards.

4. *Possible underestimation of prevalence by choice of wards.* The surveyed wards had generated approximately 13% fewer cases than we would expect based on even spread across the city, using the estimation process described in this technical document [16]. This suggests that these wards may not have been hit as hard as the city average by the time of the survey.

5. *Possible bias with unclear direction associated with non-consent in nonslum areas.* The authors of [5] note that in the nonslum areas, "There were difficulties in obtaining consent from some resident associations". This could have introduced a bias if, for example, housing societies with suspected cases were reluctant to allow the surveying.

Given the possible biases, we can check how the prevalence estimates from the first survey tally with other available data. Three other surveys give results approximately consistent with the first serosurvey.

1. *A survey in Dharavi in May-June, 2020.* According to this report in The Print [17], 36% seroprevalence was measured in Dharavi (a large slum in Mumbai) some time between May 11 and June 4. Total cases from the city's slums, estimated following the procedure in [16], approximately doubled between June 4 and July 8. If cases tracked prevalence during this period, then 54% is not a surprisingly high estimate for slum prevalence by July 8 – if anything, it is on the low side. However, few details of the Dharavi survey are available: the data was never officially released. Dharavi may have been affected earlier than other slums. And it is not clear if Dharavi as a whole was represented, or just one containment zone within Dharavi.

2. *A survey in five slums in October 2020.* Around 75% seroprevalence was found in a serosurvey in five slums in October 2020 according to this report in India Today [18]. No technical detail is available about the sampling or the assay used. Using median estimates of various quantities, 54% slum prevalence in early July results in slum prevalence of 75% by the end of the year (see Section 5). On this basis 54% does not seem unreasonably high for slum prevalence in early July.

3. *The second serosurvey in August, 2020.* This survey reported 45% seroprevalence in slums and 17.5% seroprevalence in nonslum areas of the same three wards as the first survey. The drop in the slum seroprevalence (54% to 45%), and marginal rise in nonslum seroprevalence (16% to 17.5%), may appear at face value to indicate possible overestimation in the first survey. However, analysis shows that the numbers from the two surveys are consistent provided we take into account waning sensitivity to prior infection of the assay used in both surveys. Details are given in Appendix B.

In summary, the data suggests that prevalence in slums and nonslums could have been under- or overestimated during the first serosurvey; but there is no strong evidence of bias in one particular direction. The lowest (resp., highest) values amongst *all* the 95% CI's on prevalence values given by the authors are 52.7% (resp., 59.9%) for the slums; and 14.8% (resp., 18.4%) in nonslum areas. In later simulations we use a wider range of 45% to 62% for slum prevalence by the time of the first serosurvey; and of 12% to 20% for nonslum prevalence by this point.

5. Estimating COVID-19 infections by the end of 2020

We divide the year 2020 into two periods: *Period 1* from the beginning of the epidemic to the first serosurvey mid-point (July 8, 2020); and *Period 2* from July 8, 2020 to Dec 31, 2020. When referring to fatalities (measured or estimated) the periods are shifted forwards by the fatality delay. For example, "Period 2 fatalities" refer to all fatalities associated with added prevalence in Period 2.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.08.21255101; this version posted April 10, 2021. 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-ND 4.0 International license .

We can estimate total infections by the end of 2020, by taking estimates of infections by the end of Period 1 and adding on estimates of new infections during Period 2. Since we know that detection of infections can vary substantially over time, the approach used here to estimate Period 2 infections relies on fatality rather than case data. The main assumption is that the naïve IFRs in the slums and nonslum areas did not change much after Period 2 began.

The naïve IFRs during Period 1 are calculated by the authors of [5] (using fatality data not given in the paper) as 0.076% in the slums and 0.263% in nonslum areas. These values, if applied to the city as a whole would give a naïve IFR estimate of 0.12%. They imply 2775 slum fatalities and 2573 nonslum fatalities during Period 1, assuming 54% slum prevalence and 16% nonslum prevalence and using demographic data given in Appendix A. These estimates imply that 52% of recorded fatalities during Period 1 were from the slums.

The total of 5348 Period 1 fatalities inferred from [5] is close to the 5332 COVID-19 fatalities reported in the city by July 13, 2020: thus the estimates are consistent with an assumed fatality delay of 5 days. As fatalities were rising fast at the time – they rose 37% during July – there may be some right-censoring in these naïve IFR estimates.

In order to estimate how many new infections occurred in slums and nonslum areas during Period 2, we need first to estimate recorded fatalities from the two strata during Period 2. Before doing this systematically, let's see an example calculation.

Suppose we fix a fatality delay – say 7 days. This gives 5698 recorded COVID-19 fatalities in Period 2 (namely, between July 15, 2020 and Jan 7, 2021). Suppose 5% of these Period 2 fatalities (285) were from the slums with the remaining 5413 from nonslum areas. Using the IFR values of 0.076% and 0.263% in [5], we would then infer $285/0.00076 = 3.75 \times 10^5$ new slum recoveries in Period 2, and $5413/0.00263 = 2.06 \times 10^6$ new nonslum recoveries in this period. This would amount to an additional 5.5% of slum dwellers infected, and an additional 34% of nonslum dwellers infected between July 8 and the end of year, giving end-of-year estimates of infection rate of 60% in slums, 50% in nonslum areas, and 55% city-wide.

In this calculation, the assumption that only 5% of Period 2 fatalities were from the slums is crucial. If, instead, we assumed 20% of Period 2 fatalities were from the slums, we'd get an infection rate of 61% overall (slums: 76%, nonslums: 44%) by the end of 2020. If 35% of Period 2 fatalities were from the slums, we get an infection rate of 67% (slums: 93%, nonslums: 37%) by the end of 2020.

Note, however, that there are a lot of values fixed in these estimates, including the initial estimates of slum and nonslum prevalence, the fatality delay, and the naïve IFR values. We clearly need a more systematic way of exploring such estimates, where quantities are allowed to vary over plausible ranges.

6. Monte-Carlo experiments

We can explore values of infection rate and IFR more systematically with the help of some Monte Carlo experiments. The idea is to put probability distributions on quantities where there is uncertainty (a full list is given below), and then sample repeatedly from these distributions to compute a distribution on end-of-year infection rate and IFR. The code used is available on gitthub [19].

All probability distributions used are uniform. The implications of this choice are discussed later. We use the following ranges for various parameters:

1. 2020 COVID-19 fatalities: 11,000 to 25,000. This range, discussed in Section 3 above, allows COVID-19 deaths to range, roughly, between the recorded value and a value roughly 3,000 greater than the excess deaths in the city. As the range is so wide, we assume that ignoring the fatality delay makes only a marginal difference.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.08.21255101; this version posted April 10, 2021. 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-ND 4.0 International license .

2. Slum prevalence at the time of the first serosurvey: 45% to 62%. Note that this range is somewhat wider than the bounds in [5] to allow for possible biases as discussed in Section 4.

3. Nonslum prevalence at the time of the first serosurvey: 12% to 20%. Again, this range is somewhat wider than the bounds in [5].

4. Percentage of Period 1 recorded COVID-19 fatalities from the slums: 0.95*52% to 1.05*52%. In other words, a 5% variation is allowed around the measured 52% (see Section 5).

5. Percentage of Period 2 recorded COVID-19 fatalities from the slums: 5% to 35%. This range is discussed in Appendix C.

6. The fatality delay: 0 days to 27 days. An integer uniform distribution was used. The high upper limit is to account for possibly lengthy delays in fatality reporting. The fact that the median estimate is higher than the 5 to 6 days calculated from [5], can be read as the assumption that there is likely some right censoring in the naïve IFR estimates given there.

7. Naïve IFR values in the slums and nonslum areas during Period 1 are not independent random variables: they are calculated from slum/nonslum prevalence and the number of Period 1 recorded COVID-19 fatalities from the slums/nonslums (see Section 5). The latter, in turn also depend also on the fatality delay.

8. Naïve IFR values during Period 2: these vary between 90% and 110% of the Period 1 values. Once a Period 1 value for, say, slum naïve IFR is chosen and fixed, we then choose the Period 2 slum naïve IFR value from between 90% and 110% of the Period 1 value. This is to allow for:

   - improved fatality recording after the first serosurvey (there was acknowledgement of fatality under-counting and addition of over 1,700 old fatalities in June [9]).
   - decreased IFR after the first wave passed and the health system in the city was no longer very overstretched.

The results of carrying out 100,000 experiments are shown in Table 1. We obtain median (and mean) estimates of infection rate of 60% by the end of 2020, and of IFR of 0.23% by the end of 2020. The median values correspond to an infection rate in the slums of around 75%, and in nonslum areas of around 44%, indicating that by the end of the year spread remained quite uneven between slums and nonslum areas. This is despite the fact that nonslum areas were quite heavily affected during the city's second wave in September-October 2020.

Histograms of the frequency of different infection rate and IFR estimates obtained in the experiments are shown in Figure 1. Note the fairly wide 95% CIs arising from the choice to put uniform distributions on key parameters.

	infection rate	IFR	naïve slum IFR	naïve nonslum IFR
mean	60%	0.23%	0.084%	0.29%
median	60%	0.23%	0.084%	0.29%
range	(42%, 85%)	(0.10%, 0.44%)	(0.058%, 0.12%)	(0.18%, 0.47%)
95% CI	(50%, 71%)	(0.15%, 0.33%)	(0.068%, 0.10%)	(0.22%, 0.38%)

Table 1: Estimates of infection rate and IFR in Mumbai by the end of 2020. The values are based on 100,000 experiments as described in the text.

Given the flat priors, we should interpret values with low probability as requiring a conjunction of circumstances. For example, to obtain an end-of-year infection rate below 50%, the first serosurvey must have overestimated prevalence, but in addition there must have been relatively few subsequent infections. This could be, for example, if improved fatality reporting during Period 2 led to a rise in naïve IFR values.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.08.21255101; this version posted April 10, 2021. 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-ND 4.0 International license .

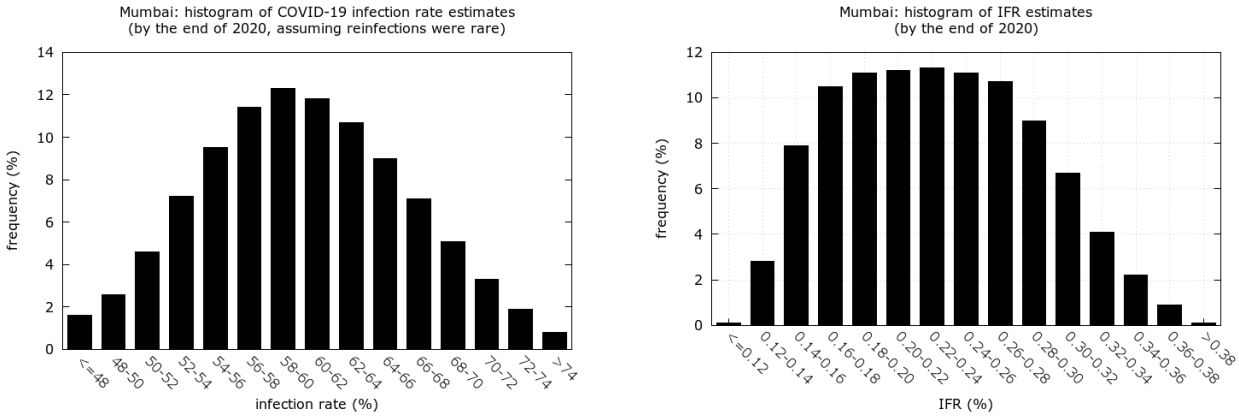

Figure 1: Histograms of the relative frequency of infection rate (*left*) and IFR (*right*) estimates from 100,000 experiments as described in the text.

Similarly, to get IFR values below 0.15% we would require a conjunction of little fatality underrecording, and a high infection rate. Higher-end estimates of infection rate often correspond to slum infection rates close to or above 100% – not, in itself, impossible, but implying that reinfections during 2020 were much more common than generally assumed. If we are reluctant to believe this, then we should discount IFR estimates much below 0.15%.

Overall, claims that the city's infection rate at the end of 2020 probably lies in the interval $(50\%, 71\%)$, and that the city's IFR during 2020 probably lies in the interval $(0.15\%, 0.33\%)$, can be regarded as statements about a balance of probabilities.

7. Comparison with estimates from meta-analyses

How do the IFR estimates above compare with predictions using published age-stratified data? Using Mumbai's 2011 age pyramid (see Appendix A), meta-analyses in O'Driscoll *et al* [3] and Levin *et al* [4] give estimated IFR values for Mumbai of 0.21% and 0.35% respectively. Note that these lie close to our median estimate (0.23%), and to the upper end of the 95% CI on IFR (0.33%) respectively. But there are a number of caveats when discussing IFR values predicted from meta-analyses.

Is the 2011 age pyramid still valid? Mumbai's age pyramid has likely changed since 2011. The national trend is of an aging population, and this would tend to increase predicted IFR values. On the other hand, Mumbai is a city of migrants; and continuing migration into the city likely pulls the median age of the city down. We don't know how these effects add up, and an up-to-date pyramid is not available.

Uneven spread by gender and age. The IFR estimates obtained from [3] and [4] assume equal spread across age groups and genders. But the data indicates lower spread amongst the over-60s, and possibly lower spread amongst men, both of which would tend to reduce predicted IFR. Treating gender first, both main serosurveys showed consistently higher seroprevalence in females [20, 6] which reached significance in the slums during the first survey. Regarding age, limited demographic information in [5] (Supplementary information, Table e3) indicates that over 60s formed about 16% of the nonslum population but only about 7% of the slum population in the areas surveyed. On the other hand, median estimates of slum/nonslum infection rates by the end of 2020 are 74% and 44% respectively. In addition, both serosurveys found significantly lower prevalence amongst the over 60s in nonslum areas. We thus expect a strong decrease in the infection rate amongst the elderly, first because the majority live in nonslum areas, and secondly because even within these areas there may have been some shielding.

How does gender affect the estimates using [3]? Using either gender-specific values or general values in [3] gives the same predicted IFR of 0.21% (assuming equal spread across genders). This is coincidental:

medRxiv preprint doi: https://doi.org/10.1101/2021.04.08.21255101; this version posted April 10, 2021. 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-ND 4.0 International license .

the greater proportion of men in Mumbai, tending to increase predicted IFR, is offset by the lower proportion of men in the highest age groups most liable to severe infection (see Table A.2).

How should the predicted IFR be calculated from [4]? The value of 0.35% computed from data in [4] uses mid-point values of the age intervals in the metaregression formula $\log_{10}(\text{IFR}) = 3.27 + 0.0524(\text{age})$. It also uses the value of 82.5 to estimate fatalities for over 80s. This latter could lead to underestimation: for example, replacing 82.5 with 84, increases the predicted IFR to 0.37%.

8. Conclusions

Estimates of Mumbai's COVID-19 IFR by the end of 2020 are broadly consistent with the results of meta-analyses. The median estimate of 0.23% can be seen as arising from an infection rate of 60%, and an assumption of 18,000 COVID-19 deaths by the end of 2020. These numbers arise if the first serosurvey accurately measured prior infection in the city, naïve IFR in the slums/nonslums did not change much in Period 2, and about 64% of all excess deaths not officially recognised as COVID-19 deaths were actually COVID-19 deaths.

While there are many uncertainties, the analysis tells us that naïve IFR at the time of the first serosurvey (roughly 0.12%) is highly unlikely to approximate true IFR in the city during 2020. Similarly, IFR values of 0.4% or higher require a conjuction of circumstances including significant overestimation of prevalence in the first serosurvey.

We should understand the assumptions which might lead to values near the ends of confidence intervals. If most of the city's excess deaths were COVID-19 deaths, and prevalence was slightly overestimated at the first serosurvey, we easily obtain IFR values of over 0.3%. If death recording was fairly accurate, and a large part of the city's excess mortality was non-COVID, we can easily obtain IFR values of 0.18% or below.

Finally, whatever it's true value by the end of 2020, IFR in the city is not static – it likely changed considerably over time as disease moved between different subpopulations of the city [9]. Disease spread remains lower in older groups, and so IFR could rise; more or less deadly variants could begin to circulate pushing it up or down; and vaccination and reinfections could add complicating factors into the mix.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.08.21255101; this version posted April 10, 2021. 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-ND 4.0 International license .

Appendix A. Notes on data used

Data on cases and deaths are taken from Mumbai's daily reports from the municipal corporation. These are of two kinds: a brief report posted daily on twitter; and a more detailed bulletin posted daily by the municipal corporation on the web [21]. Cumulative case and fatality data at any point is taken from the brief reports. The data is archived at [22].

The reliance on news sources for some kinds of information reflects the fact that data is often shared with news portals but not shared publicly. The use of some estimation procedures is necessitated by the fact that data such as a breakdown of cases and deaths between slum and nonslum areas is not routinely shared; if such data were to become available, some estimates could be replaced by measured values.

2011 demographic data used for predicted IFR values, are shown in Table A.2. Table A.3 contains data on mortality from the Vital Statistics report [10], accessed in March 2021. Table A.4 is reproduced from [16], and gives 2011 data on the slum/nonslum populations in each ward, and the city as a whole.

Appendix B. Mumbai's second serosurvey

Mumbai's second serosurvey (approximate mid-point: August 23, 2020) was carried out in the same three wards as the first, and reported seroprevalence values of 45% in the slums and 17.5% in nonslum areas [6]. This gives around 32% seroprevalence citywide. These values appear to be adjusted for population characteristics, but not for sensitivity or specificity of the tests.

The apparent drop in seroprevalence between first and second city serosurveys is consistent with waning sensitivity of the assay used (the "Abbott Diagnostics Architect N-protein based test") to prior infection as measured in Muecksh *et al* [7]. Calculations based on the two serosurveys, using techniques and values for the speed and frequency of seroreversion from [24] were carried out [25], and gave estimates of 28% prevalence in nonslum areas and 78% prevalence in the slums by the time of the second survey.

These estimates are somewhat higher than values predicted by tracking fatalities as in this paper, suggesting that the waning sensitivity of the tests may not have been as rapid in Mumbai as estimated from Brazilian data. Nevertheless, we can broadly conclude that the measured seroprevalence values in the first and second surveys are not inconsistent with each other.

Appendix C. The fraction of Period 2 recorded fatalities coming from Mumbai's slums

Using the procedure in [16], we find that around 16% of Mumbai's COVID-19 cases between July 8 and the end of 2020 came from the slums. On the other hand, by the time of the first serosurvey, detection in the slums was much lower than in the nonslum areas, with an estimated 0.8% of slum infections identified in testing as against about 5.9% in nonslum areas [26]. If detection in the nonslum areas remained more than 7 times higher than in the slums after the first serosurvey then we can calculate, using naïve IFR estimates in slum and nonslum areas in [5], that about 29% of recorded fatalities after the first serosurvey would have occurred in the slums.

However, it is quite likely that detection of infections in the slums was particularly poor during the massive slum surge in April and May and improved (proportionally) more significantly than detection in the nonslum areas during Period 2. If, say, slum detection tripled during Period 2, while nonslum detection only doubled during, then around 21% of Period 2 recorded fatalities would have occurred in the slums.

This appears a plausible scenario, but given the uncertainties, we allow a wide range: between 5% to 35% of Period 2 COVID-19 deaths occurred in the slums. The lower figure corresponds to detection in the slums achieving parity with detection in nonslum areas during Period 2. The upper figure corresponds to detection in nonslum areas being almost 10 times higher than in the slums during Period 2.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.08.21255101; this version posted April 10, 2021. 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-ND 4.0 International license .

	all	male	female
All ages	12442373	6715931	5726442
0-4	837269	436703	400566
5-9	909385	477656	431729
10-14	998448	529186	469262
15-19	1100379	613217	487162
20-24	1333546	761998	571548
25-29	1310338	728325	582013
30-34	1109979	609191	500788
35-39	1015414	546897	468517
40-44	866960	466417	400543
45-49	753234	400590	352644
50-54	611020	331428	279592
55-59	481470	256815	224655
60-64	390770	200440	190330
65-69	260538	131235	129303
70-74	183134	90112	93022
75-79	110548	53075	57473
80+	121329	55159	66170

Table A.2: Mumbai's 2011 population by age and gender from [23].

year	2015	2016	2017	2018	2019	2020
estimated mid-year population	1,26,43,252	1,26,89,644	1,27,36,036	1,27,82,429	1,28,28,821	1,28,75,213
births	1,74,902	1,52,952	1,55,386	1,51,187	1,48,898	1,20,188
deaths	94,706	86,642	89,037	88,852	91,223	1,11,942
infant deaths	4575	3998	4079	3723	3430	2649

Table A.3: Mumbai's estimated mid-year populations, births, all cause mortality, and infant mortality during 2015–2020, as reported in [10].

medRxiv preprint doi: https://doi.org/10.1101/2021.04.08.21255101; this version posted April 10, 2021. 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-ND 4.0 International license .

ward	total	slums	nonslums	% in slums
A	185014	22282	162732	12%
B	127290	12711	114579	10%
C	166161	16571	149590	10%
D	346866	34699	312167	10%
E	393286	124194	269092	32%
G/S	377749	124306	253443	33%
H/W	307581	82552	225029	27%
K/W	748688	215678	533010	29%
R/C	562162	172849	389313	31%
T	341463	85560	255903	25%
F/S	360972	180128	180844	50%
F/N	529034	238128	290906	45%
G/N	599039	361674	237365	60%
P/S	463507	230829	232678	50%
R/S	691229	414395	276834	60%
M/W	411893	164992	246901	40%
N	622853	249229	373624	40%
S	743783	408442	335341	55%
R/N	431368	281151	150217	65%
K/E	823885	572818	251067	70%
H/E	557239	388923	168316	70%
P/N	941366	708247	233119	75%
L	902225	758108	144117	84%
M/E	807720	685994	121726	85%
total	12442373	6534460	5907913	53%

Table A.4: Total population in Mumbai's wards at the 2011 census, broken down into those residing in slums and those not residing in slums.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.08.21255101; this version posted April 10, 2021. 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-ND 4.0 International license .

References

[1] P. Pulla. What counts as a COVID-19 death? *BMJ*, 370:m2859, 2020.

[2] A. Karlinsky and D. Kobak. The World Mortality Dataset: Tracking excess mortality across countries during the COVID-19 pandemic. Available at https://www.medrxiv.org/content/10.1101/2021.01.27.21250604v1.full.

[3] M. O'Driscoll, G. R. Dos Santos, L. Wang *et al.* Age-specific mortality and immunity patterns of SARS-CoV-2. *Nature*, 590:140–145, 2021.

[4] A. T. Levin, W. P. Hanage, N. Owusu-Boaitey *et al.* Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications. *European J. Epidemiology*, 35:1123–1138, 2020.

[5] A. Malani, D. Shah, G. Kang *et al.* Seroprevalence of SARS-CoV-2 in slums versus non-slums in Mumbai, India. *The Lancet Global Health*, 9(2):e110–111, 2021.

[6] U. Kolthur-Seetharam, D. Shah, J. Shastri *et al.* SARS-CoV2 Serological Survey in Mumbai by NITI-BMC-TIFR: Preliminary Report of Round-2. Available at https://www.tifr.res.in/TSN/article/Mumbai-Serosurvey%20Technical%20report-NITI_BMC-Round-2%20for%20TIFR%20website.pdf.

[7] F. Muecksch, H. Wise, B. Batchelor *et al.* Longitudinal Serological Analysis and Neutralizing Antibody Levels in Coronavirus Disease 2019 Convalescent Patients. *J Infect. Dis.*, 223(3):389–398, 2021.

[8] J. Zhao, Q. Yuan, H. Wang *et al.* Antibody responses to SARS-CoV-2 in patients of novel coronavirus disease 2019. *Clin. Infect. Dis.*, 71(16):2027–2034, 2020.

[9] M. Banaji. Mumbai Had 13k 'Excess' Deaths in March-July. How Does Its COVID Story Change? *The Wire Science*, September 10, 2020, available at https://science.thewire.in/the-sciences/covid-19-mumbai-all-cause-mortality-data-ifr-bmc-seroprevalence-survey/.

[10] Municipal Corporation of Greater Mumbai. Vital Statistics report for 2020. retrieved on March 20, 2021, from https://portal.mcgm.gov.in/irj/portal/anonymous/qlvitalstatsreport?guest_user=english.

[11] M. Sood. Lockdown effect: Deaths on Mumbai's railway tracks down by 65% in 2020. *Hindustan Times*, Jan 19, 2021, available at https://www.hindustantimes.com/cities/mumbai-news/lockdown-effect-deaths-on-mumbai-s-railway-tracks-down-by-65-in-2020-101611079137368.html.

[12] S. Sen. Fatalities on Mumbai roads plunge 36%, accidents 39% in 11 months of 2020. *Times of India*, Jan 14, 2021, available at https://timesofindia.indiatimes.com/city/mumbai/fatalities-on-city-roads-plunge-36-accidents-39-in-11-months-of-2020/articleshow/80257334.cms.

[13] Rukmini S. How India Could Fill In The Blanks On Excess Mortality. *IndiaSpend*, June 17, 2020, available at https://www.indiaspend.com/how-india-could-fill-in-the-blanks-on-excess-mortality/.

[14] T. Barnagarwala. Covid outbreak: 64,349 deaths in Mumbai this year. *The Indian Express*, September 9, 2020, available at https://indianexpress.com/article/cities/mumbai/covid-outbreak-64349-deaths-in-mumbai-this-year-6587188/.

[15] K. Purohit. Meraj Shaikh's Lost Body & Other Non-Covid Victims. *Article 14*, August 14, 2020, available at https://www.article-14.com/post/meraj-shaikh-s-lost-body-mumbai-s-non-covid-victims.

[16] M. Banaji. Technical report: inferring COVID-19 cases from slums and nonslum areas in Mumbai. Dec. 2020, available at https://github.com/muradbanaji/MumbaiData/blob/master/SlumNonslum.pdf.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.08.21255101; this version posted April 10, 2021. 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-ND 4.0 International license .

[17] S. Dasgupta. ICMR denies sero survey hushed up data showing up to 48% Covid reach in hotspots in May. *The Print*, September 21, 2020, available at https://theprint.in/health/icmr-denies-sero-survey-hushed-up-data-showing-up-to-48-covid-reach-in-hotspots-in-may/507511/.

[18] M. Shaikh. Serosurvey finds 75% test positive for Covid-19 antibodies in Mumbai slums. *India Today*, November 30, 2020, available at https://www.indiatoday.in/coronavirus-outbreak/story/serosurvey-mumbai-slum-antibodies-coronavirus-1745303-2020-11-30.

[19] M. Banaji. Code to generate IFR estimates for Mumbai using Monte Carlo experiments. Available at https://github.com/muradbanaji/COVIDAGENT/blob/master/MumbaiIFR_MC.cc.

[20] U. Kolthur-Seetharam, D. Shah, J. Shastri *et al.* Technical details: SARS-CoV2 Serological Survey in Mumbai by NITI-BMC-TIFR. Available at https://www.tifr.res.in/TSN/article/Mumbai-Serosurvey%20Technical%20report-NITI.pdf.

[21] Municipal Corporation of Greater Mumbai. BMC Covid-19 Response War Room Dashboard. posted daily at https://stopcoronavirus.mcgm.gov.in/assets/docs/Dashboard.pdf.

[22] M. Banaji. Archive of Mumbai's recorded case and fatality data in daily bulletins. Available at https://github.com/muradbanaji/MumbaiData/blob/master/MumbaiByAgeBulletins.csv.

[23] Office of the Registrar General & Census Commissioner, India Ministry of Home Affairs, Government of India. C-14 CITY: POPULATION IN FIVE YEAR AGE-GROUP BY RESIDENCE AND SEX - 2011. Available at http://www.censusindia.gov.in/2011census/C-series/DDWCT-0000C-14.xls.

[24] L. F Buss, C. A. Prete Jr, C. M. M. Abrahim *et al.* COVID-19 herd immunity in the Brazilian Amazon. preprint at https://www.medrxiv.org/content/10.1101/2020.09.16.20194787v1.full.pdf.

[25] M. Banaji. What does preliminary data from Mumbai's second serosurvey tell us, and does it change Mumbai's COVID story? October 5, 2020, available at http://maths.mdx.ac.uk/research/modelling-the-covid-19-pandemic/mumbaisecondserosurveyfull/.

[26] M. Banaji. Notes on Mumbai's COVID-19 epidemic and the September resurgence. September 28, 2020, available at http://maths.mdx.ac.uk/research/modelling-the-covid-19-pandemic/mumbairesurgencesuppl/.