The toll of COVID-19 on Mumbai – more pieces of the puzzle

(Murad Banaji, 12/09/2020)

On September 9, *The Indian Express* obtained data from the Brihanmumbai Municipal Corporation (BMC) on excess deaths during Mumbai's COVID-19 epidemic. We now know that Mumbai recorded 13,058 more deaths between March 1 and July 31, 2020 than in the same period of 2019. This amounts to 6,663 deaths *over and above* the 6,395 COVID-19 deaths reported during this period. May, June and July saw the bulk of the additional deaths.

To provide some perspective, London's total excess deaths during the first wave of its COVID epidemic were about 20% higher than its recorded COVID-19 deaths; Mumbai's excess deaths are over 100% higher than its recorded COVID-19 deaths. How many of the apparently non-COVID deaths could actually be from COVID?

There are no easy answers to this question. During lockdown certain kinds of deaths – e.g., from road and rail accidents – are expected to fall, and indeed the report by the Indian Express tells us that there were about 800 fewer deaths of this kind during March-July 2020 compared to the same period in 2019. At the same time, we would expect other deaths to rise, for example of non-COVID patients struggling to access healthcare. There may also have been deaths recorded in the city of residents who came from surrounding areas to seek treatment.

Thus the true number of COVID-19 deaths in the city during March to July could be lower or higher than the 13,000 or so excess deaths reported in this period. Nevertheless, despite the uncertainties, we can see this new data as a piece in a larger puzzle which brings us closer to understanding how Mumbai's epidemic has progressed. To go forward, we first need to rewind.

The early trajectory of Mumbai's epidemic

In the early days of the epidemic, until the middle of April, Mumbai's case and fatality data followed a very standard pattern. There was exponential growth in cases and deaths, with deaths following cases after a delay. (This delay is hard to infer accurately from Mumbai's data – rather the data is roughly consistent with a range of delays.) However, in mid-April both cases and deaths slowed, no doubt partly as a consequence of lockdown. As I noted at the time, what was surprising was the *extent* to which fatalities slowed – they were now rising at a much slower pace than we would expect from the rise in cases. And there were no compelling reasons to believe this was a consequence of rapidly improving detection of infections.

Officials implied that the city was seeing a genuine drop in infection fatality rate (IFR), namely the proportion of those getting COVID-19 who were dying, and attributed it to "strategic interventions". But what seems most likely from the data we now have is that the lower than expected fatalities resulted from a combination of two effects.

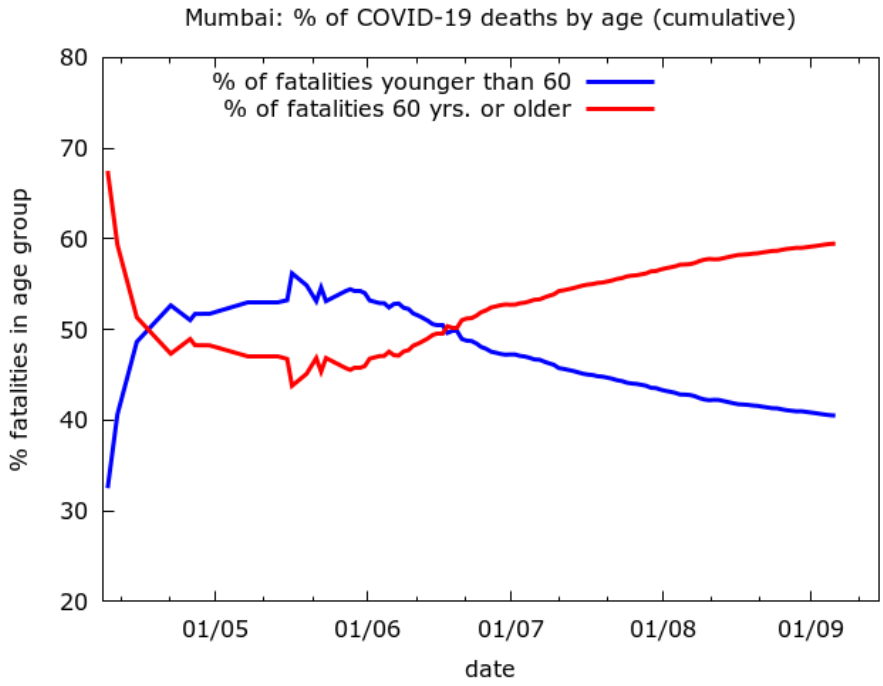
1) Deaths being increasingly undercounted. That this was occurring was later confirmed by the addition of around 1700 "old" fatalities into Mumbai's count during the latter half of June – over 800 of these on a single day. The new excess deaths data add force to the claim that this reconciliation of Mumbai's COVID-19 fatality data was incomplete.
2) A genuine drop in IFR caused by disease spreading in a younger population. We know that older people are very much more likely to die of COVID-19 than younger people: so if disease shifts to a younger population fatality rates will, on the whole, fall.

Can we disentangle these two effects and ask what has “really” been happening with COVID-19 fatality in Mumbai? The answer could be important for other cities facing COVID-19 epidemics in India.

Uneven spread of COVID-19 in different age groups

A key piece of the puzzle was provided by data from Mumbai's serosurvey carried out in the first half of July. This data told us that Mumbai’s epidemic could be viewed as two interlinked epidemics, one occurring in slum areas and the other in non-slum areas. At the time of the serosurvey, residents tested in the slums were more than 3 times as likely to have been exposed to COVID-19 than residents of non-slum areas. We can infer that the slum epidemic was rapid, picking up in late March and early April, and winding down two months later – in late May and early June. Meanwhile the non-slum epidemic progressed much more slowly, at least after lockdown in late March.

Given the younger, often migrant worker, population of the slums, we would expect the shift of disease to/from the slums to be reflected in changing levels of infection – and hence deaths – in different age groups. And when we examine recorded COVID-19 fatalities by age over time, we find a striking picture.

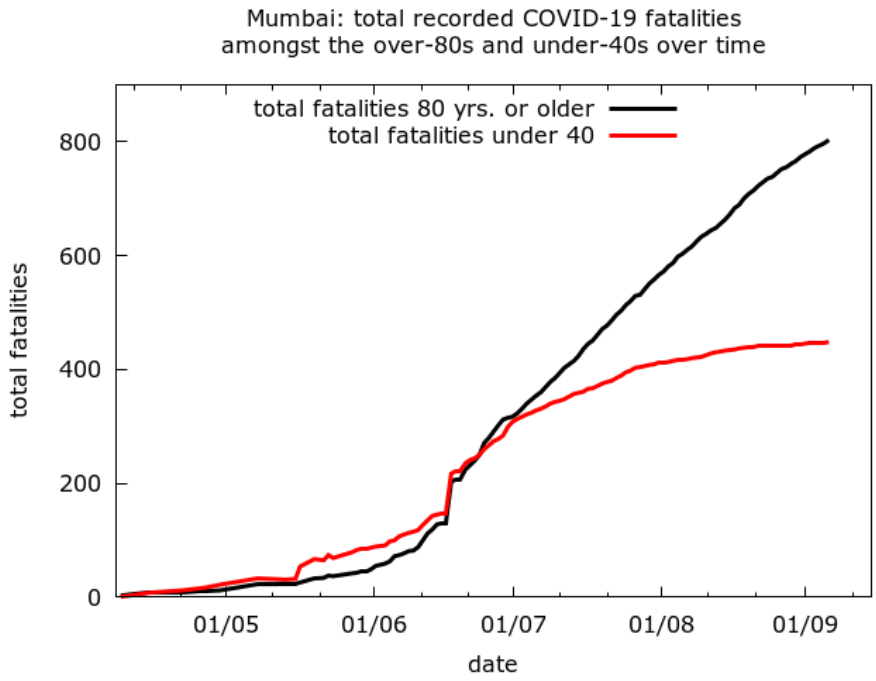


In early April we see fatalities skewed towards the elderly. The numbers are still small but out of Mumbai’s 64 recorded COVID-19 deaths by April 11, 60% had been in the over 60s. Remembering that there is a delay of perhaps 3 weeks between infection and death, these fatalities reflect disease spread in mid March. It is likely that much of this early spread was in housing societies with relatively older populations, following the import of cases into the city in late February and March.

However by the end of April the picture had changed, and there were now more fatalities in the under 60s. This pattern then persisted throughout May and early June, consistent with the rapid spread amongst the relatively young slum population, but slower spread in the housing societies.

But in June, things were changing again with a steady shift in fatalities back towards the older age groups. Again, bearing in mind the delay between infection and deaths, we see the likely cause of this second shift: in late May, although disease was slowing in some slum areas, it was still smouldering – perhaps even picking up – in non-slum areas.

Even more striking than the comparison between the under-60s and the over-60s is the comparison between the under-40s and the over-80s. The first group is numerically large, but at relatively low risk of severe disease; the second group is small, but at significant risk of dying of COVID. Available data suggests that 8% or more of the over 80s who contract COVID-19 will die.



In the early days, the total numbers of deaths in both groups were similar. Even the jumps in mid-June caused by addition of old deaths were of much the same size. But by July deaths in the younger group were slowing while deaths in the over 80s continued to rise. Again, we see likely signs that the disease had slowed in the slums, but was continuing at some pace in non-slum areas.

Disentangling uneven spread and undercounting

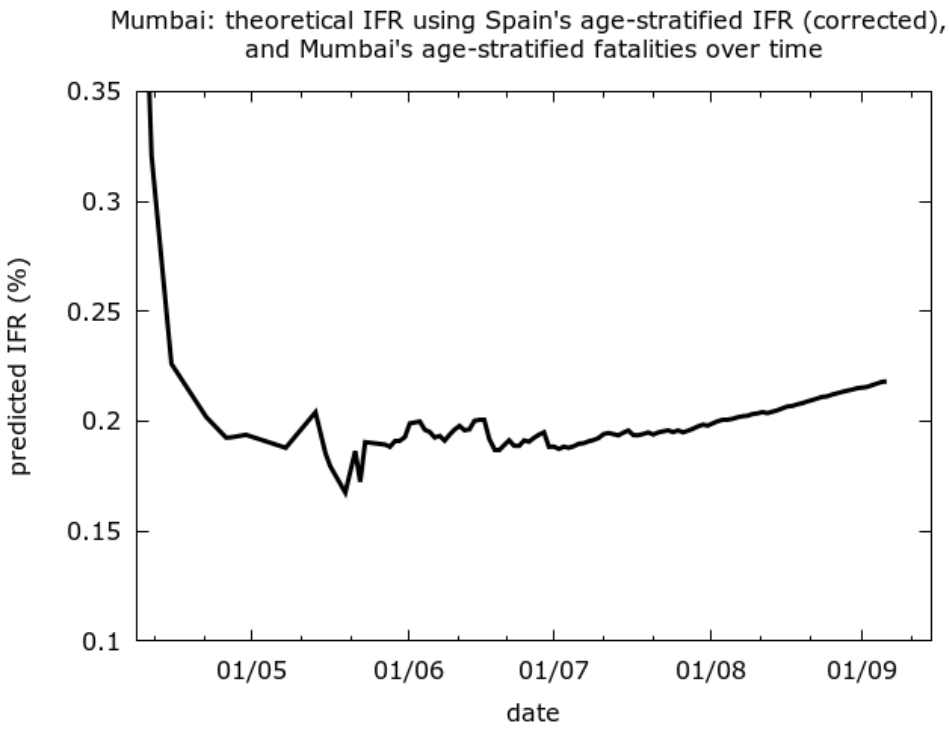
The analysis of Mumbai's fatalities by age is certainly consistent with the story of the epidemic in the city. But we can do a little better and use this data along with data on COVID-19 fatality rates to roughly predict likely changes in IFR in the city.

The simplest approach is as follows: we take the recorded deaths by age at a given moment; use these and IFR values for each age group to estimate disease prevalence in each age group (some time earlier); and divide the total deaths across age groups by the total estimated infections (some time earlier) to get an estimate for overall IFR.

Although conceptually simple, this explanation is slightly misleading, because the raw numbers of fatalities actually do not matter in this calculation – all that matters is that recorded fatalities

correctly capture the *proportion* of fatalities in each age group. A key caveat is that these estimates assume that fatality undercounting not varying significantly with age. More on this later.

As an example, let's use Spanish IFR values reported here, all multiplied by a factor of 1.5 in the light of reports of possible undercounting of Spanish fatalities. We get the picture below.



This experiment gives an estimated IFR value for Mumbai of 0.32% on April 11. This is very much in the range of IFR values that we arrive at for the early period through modelling in conjunction with later seroprevalence data. The estimated IFR then falls rapidly in the second half of April dropping below 0.20% before climbing back to around 0.22% in early September, consistent with increasing spread in older populations. Note that these are *cumulative* IFR values: current IFR based only on recent deaths would be higher.

This theoretical IFR pattern shadows the observed drop in fatalities seen in Mumbai in late April. But the estimates put IFR at around 0.19% at the time of Mumbai's serosurvey, quite a lot above the naive values of 0.11% to 0.12% inferred from the recorded deaths. The theoretical IFR suggests that we should have seen approximately 9,000 fatalities shortly after the serosurvey rather than the observed 5,500. Fastforwarding to the end of July when Mumbai had recorded 6350, the calculations suggest that another 4000 to 5000 COVID-19 deaths had gone undetected. These estimates correspond to about 40% of COVID-19 fatalities going unrecorded. They would imply that, in total, around 80% of Mumbai's excess deaths by the end of July could be attributed directly to COVID-19.

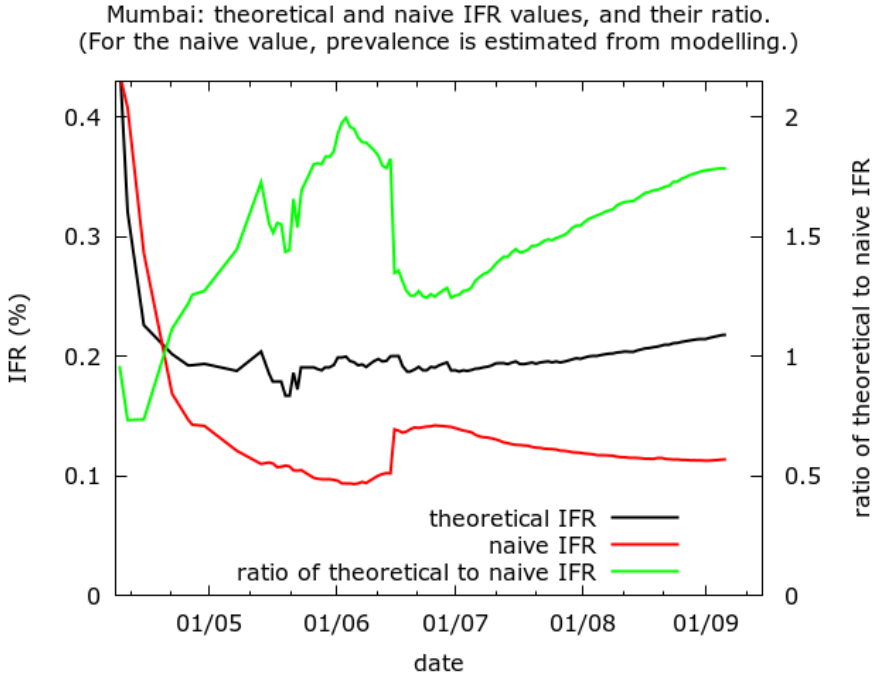
If we assume that COVID-19 fatalities are disproportionately missed in the elderly, then the IFR pattern in the picture above would change. The initial fall would be similar, but the subsequent rise would be steeper, and estimated IFR by early September would be higher. This could also push the current estimate for unrecorded COVID-19 deaths higher. As a sanity check, it would be extremely useful to have a breakdown of Mumbai's excess deaths by age.

Conclusions

The trajectory of COVID-19 deaths in Mumbai can be largely explained by a combination of two factors: disease spreading with different speeds in populations with different age structure; and fatality undercounting, whether systematic or inadvertent.

We now, finally, have a plausible story consistent with Mumbai's case, death and seroprevalence data. In April and early May, there was both increased undercounting and a genuine drop in fatality rate in the city as disease moved to the slums. Rough calculations assuming that undercounting is uniform across age groups suggest that today Mumbai's recorded COVID-19 death count should be multiplied by a factor of 1.6 to 1.8 to obtain the true death count.

The analysis also suggests that this multiplier – captured in the green line in the graph below – could be increasing. In particular, if we compare theoretical IFR to naive IFR estimated after the serosurvey using modelling, we find that the former is rising, while the latter is falling. This trend of apparently falling IFR could of course be a consequence of improved treatment. If excess deaths data for August becomes available it should shed some light on this.



One implication of the uneven spread is that the IFR for the city could rise as disease spreads increasingly in an older population. At a practical level, it is crucial that mitigation and containment efforts do not cease while there are still large numbers of elderly residents of the city who have not been infected and may be vulnerable to severe COVID.

A postscript. In May I had stated, about the sudden mid-April slowdown in Mumbai's daily COVID-19 deaths, "*one must consider carefully whether [disease spreading in the slums] should really cause such a sudden reduction in fatalities*." I'd gone on to suggest we should see a "*slight and more gradual slowing in fatalities*". With the data we now have both on fatality rates and the very rapid spread in the slums this view seems half correct! In fact, the slowing in fatalities caused by the combination of mitigation taking effect in non-slum areas and disease spreading in the slums

could be very significant. But, crucially, this was not the whole story and we now know that fatality undercounting also played a major part in the apparent slowing of fatalities.

Notes on data and simulations

The data on fatalities in Mumbai by age are taken from Mumbai's dashboards published daily by BMC. This age related data has been gathered here: https://github.com/muradbanaji/MumbaiData. Aside from the major June reconciliation, the data includes various minor anomalies including occasional *decreases in death counts* which could reflect changes in which deaths were classified as COVID, or in the age categorisation of deaths. The non-smoothness in several of the graphs above is a consequence of these anomalies.

The modelled naive IFR curve in the final graph is taken from the simulation using COVIDAGENT and using an initial IFR of 0.38% here. Parameter values are as given in that document.