

May 5, 2021

Estimating Covid-19 Fatalities in India

By: Murad Banaji

India's recorded Covid-19 fatalities are a gross underestimate. A more accurate assessment, from careful interpretation of available data and of international trends, is needed to better respond to the pandemic.

How many have died of Covid-19 in India? This is a very hard question to answer. The 220,000 recorded deaths (as of 3 May) are a terrible catastrophe in themselves. But multiple reports from around the country confirm that the recorded fatalities are not the whole story. International data also indicates that given India's population and the extent of disease spread we should expect many more deaths.

Even in the midst of this calamitous second wave, counting the dead is important. Low recorded Covid-19 mortality during the first surge fed [government narratives on the successful handling of the epidemic](#), and these, in turn, might explain the complacency which preceded the devastating current surge. Arriving at credible estimates of the epidemic's true toll could be key to mitigating further disaster.

Why do we need to correctly count all covid deaths?

To spread doom and gloom for the sake of it? No.

Coz it's an important tool for epidemiologists? Not only that.

To deal with a problem we must first understand how bad it is..

And because it's vital for our collective memory.

— Varun C. MD, DM (@drcheruvarun) April 26, 2021

What we learn from reports in the media

Media reports have told us of big discrepancies in official data, of graveyards running out of space, of temporary cremation grounds, and of shortages of wood to fuel the burning pyres. We have seen page after page of obituaries from cities officially recording just a few deaths from the coronavirus.

We cannot reliably infer the scale of India's unrecorded coronavirus deaths from such reports. But they provide some valuable lessons.

The first lesson is that undercounting is widespread. Reports of fatality undercounting have emerged from West Bengal, Delhi, Tamil Nadu, Maharashtra, Madhya Pradesh, Gujarat, Telangana, Uttar Pradesh, Assam, Odisha, Kerala, Karnataka, Bihar, Haryana, and Chhattisgarh. These states make up around 80% of India's population.

Some news reports on COVID-19 death undercounting in India... 1/

West Bengal: 50% of deaths (late March to April) were omitted on account of comorbidities before a "reconciliation" in early May. (Modelling suggests undercounting continued after this.)<https://t.co/FQAIRmwBu3>

— Murad Banaji (@muradbanaji) July 18, 2020

Although most reports focus on urban areas, a few describe major undercounting in rural areas. Some states seem to be worse offenders than others, but almost certainly the problem is nationwide. Where such reports have not emerged, local media may simply have lacked the resources or motivation to pursue these stories.

The second lesson is that undercounting occurs in many ways. Again and again, we have seen officials leaving out [deceased patients with co-morbidities](#) — conditions that increase the risk of severe Covid-19 — from the toll. This practice appears to be very widespread, despite being in clear violation of Indian Council of Medical Research (ICMR) [guidelines](#). But apart from [censure of West Bengal](#) in April 2020, little action has ever been taken. Also in violation of ICMR guidelines is a widespread requirement that a deceased patient must have [COVID confirmed through testing](#) to make it to the official count. A likely consequence is that where testing is poor, death recording may also be weak.

According to data from the third national serosurvey, across India just 3% of infections were detected through testing in 2020.

Reporting of Covid-19 deaths could be also very rare in regions where awareness of the disease is low and access to healthcare minimal. Ground reports in local and national media show that rural Uttar Pradesh has been hit catastrophically in this surge. In [one village](#), a resident spoke of mysterious deaths following fever, cough, and breathing problems: “I am being informed about a death everyday in the village. There was no testing, so we don’t know what is the cause of those mysterious deaths.”

The third lesson is that the scale of underreporting is very variable and can be huge. Some reports describe modest levels of underreporting. But the scale can also be dizzying. In the first three weeks of April 2021, almost 1,000 bodies were disposed of with Covid-19 protocols in Bhopal, while government data listed only 50 Covid-19 deaths in the city. According to another report, on one day alone, [Kanpur reported five times as many funerals as normal](#), amounting to almost 400 extra funerals; but official figures listed three Covid-19 deaths on that day.

Shocking as these mismatches are, they are not without [international precedent](#). A [study from Zambia](#) which involved random testing of the deceased found ten times as many Covid-19 deaths as recorded. Another study from [Damascus, Syria](#) found that a tiny 1.25% of possible Covid-19 deaths had been recorded.

Seroprevalence data and variable disease surveillance

The news reports provide valuable insights, but it is hard to build a clear national picture from them. What other data could we turn to?

[Seroprevalence surveys](#), or “serosurveys” for short, sample a population to estimate how many have developed antibodies to SARS-CoV-2, the virus which causes Covid-19. They are not perfect; but give us an idea of how many have been infected.

The first thing we learn from serosurveys in India is that only a fraction of infections get recorded as cases. According to [data from the third national serosurvey](#), across India just 3% of infections were detected through testing in 2020.

Given that many SARS-CoV-2 infections are mild or asymptomatic, we expect many to be missed. But local estimates show wide variation: ranging from over 10% of infections detected in Delhi in late 2020, to under 1% in Bihar or the slums of Mumbai. Some of the variation might reflect the fact that youthful populations are less prone to severe infection. But we also see poor detection in states where testing has been low, unfocussed, or manipulated, [as reported in Bihar](#).

Where disease detection is poor, could death surveillance also be weak? Examining local death rates seems to confirm this suspicion.

At this point we might recall that the deceased who have not tested positive do not, in general, make it into official tolls. Where disease detection is poor, could death surveillance also be weak?

Examining local death rates seems to confirm this suspicion.

Highly variable death rates

In some regions, serosurvey data, along with the death toll, can be used to estimate the local “infection fatality rate” (IFR) of Covid-19: the fraction of all SARS-CoV-2 infections which have resulted in death. This is a crucial quantity – after all, we want to know our chances of dying if infected with the virus.

But using recorded deaths risks grossly underestimating death rates. Even the ICMR [acknowledged](#) these risks when it chose to ignore fatality data from districts reporting very few deaths in its estimates of national IFR after the first national serosurvey.

To remind ourselves of the limitations of official data, let us refer to IFR estimates based on recorded Covid-19 deaths as ‘naïve’. We find huge variability in naïve IFR estimates from different locations. For example, it would seem that Chhattisgarh’s naïve IFR was [more than ten times](#) greater than Bihar’s. This result is even more astounding since there are no demographic differences or other obvious factors which could account for such a divergence.

Abandoning the highly implausible position that people in Bihar were much more resilient to the disease than in Chhattisgarh, the obvious conclusion is that death surveillance was much weaker in Bihar.

|| India’s national serosurveys are consistent with a story of ‘missing’ rural deaths.

There are also major variations in apparent death rates between urban and rural areas. India’s national serosurveys are consistent with a story of ‘missing’ rural deaths. The period between the second and third national serosurveys (roughly from September to the end of 2020), saw very substantial rural spread. Even as the pandemic slowed in the cities and the country’s total case-load declined, [rural seroprevalence jumped](#) to 19% from 5.2%. Nationwide, total infections tripled during this period, but recorded deaths only doubled, pulling the country’s naïve IFR down to 0.05% from 0.08%.

Was this drop in IFR real? Or does it reflect weak rural death recording? Case studies from rural areas suggest the latter: for example, the *Dainik Bhaskar* newspaper recently reported a tragic surge in [Covid-19 deaths in Belkheda village in Madhya Pradesh](#), almost all of which went unrecorded. With the spotlight on larger urban areas, we have to wonder just how many deaths in smaller towns and villages went unreported during India’s first wave.

All-cause mortality data

With evidence pointing to huge variations in death surveillance between states and between urban and rural areas, can we hope for any national estimate of Covid-19 deaths?

All-cause mortality data tells us how many deaths occurred in some region in a given period and allows us to compare with the same period in previous years. [Excess deaths](#) over and above expected numbers might give us clues about uncounted deaths from the coronavirus. Internationally, all-cause [mortality data](#) has provided vital information on the true toll of the pandemic.

Unfortunately, all-cause mortality data in India is extremely patchy. In many states, only a fraction of deaths are registered, and this fraction could well diminish during a pandemic. Even where death registration is high, the data may not be shared or may lack detail.

|| Even in a city with high death registration a large number of Covid-19 deaths – perhaps even half of the total – could go uncounted.

Where excess mortality data is available it can be hard to interpret. Kerala is a case in point. The main surge began later in the state than in most of India. By the end of 2020, Kerala had recorded a relatively modest 3,000 Covid-19 deaths. There were, however, reports of significant undercounting based on a [volunteer effort tracking local news reports and obituaries](#) and comparing these with official figures.

And yet, [remarkably](#), in 2020 the state recorded around 4% fewer deaths than expected from the previous five years’ data. Could death registration have dropped in Kerala, or did the lockdown reduce some kinds of mortality? The story is incomplete, but in Kerala’s case, all-cause mortality data gives us little insight into how many died of Covid-19. It cannot, after all, be a negative number.

In Mumbai, on the other hand, [death registrations tell a different story](#). Despite significant drops in infant deaths and accidental deaths, 2020 saw roughly twice as many excess deaths as total recorded Covid-19 deaths. We cannot be sure how many of the excess deaths were from Covid-19; but the data indicates that even in a city with high death registration a large number of Covid-19 deaths – perhaps even half of the total – could go uncounted.

Could survey data tell us the true toll?

Survey data could overcome some of the weaknesses in death registration. The [sample registration system](#) is a large national survey that provides annual estimates of mortality. It could give crucial insights into the toll of the pandemic nationwide. But the latest available data is from 2018.

|| The Consumer Pyramids Household Survey reportedly showed almost twice as many deaths as expected in the sample population during May-August 2020.

There are other surveys. Some data from the [Consumer Pyramids Household Survey](#) reportedly showed [almost twice as many deaths](#) as expected in the sample population during May-August 2020. This is a huge rise, but we cannot be sure if the surveyed households were representative of the population as a whole, and it is not clear what fraction of the additional deaths could be attributed to Covid-19.

In the future, survey data could provide a more complete picture of the true impact of the pandemic. [Verbal autopsies](#) could help in estimating how many deaths were caused by Covid-19 nationally. Of course, there would need to be a political appetite for properly mapping the impact of the pandemic, something which is currently absent.

Expectations from international data

How many coronavirus deaths would we *expect* to have occurred in India? Predictions must take India's youthful population into account. We know that all else being equal, an infected 20-year-old is much less likely to die than an infected 80-year-old. Fortunately, there is now a considerable volume of international data on what fraction of infections is likely to result in death in different age groups. In particular, there are meta-analyses which use data from multiple studies to estimate how infection fatality rates vary with age.

Let us use India's [projected age structure](#) for the year 2021 alongside two well-regarded meta-analyses, by [O'Driscoll *et al*](#) and [Levin *et al*](#). Data from the first paper indicates that in India around 0.25% of infections (one in 400) should result in death, while the second paper suggests that about 0.4% (one in 250) infections should result in death.

|| Even the lower estimate implies that 80% of Covid-19 deaths in 2020 went 'missing'.

What do these look like in absolute numbers? The third national serosurvey gave an estimate of around 300 million infections during 2020. Instead of the recorded 1.5 lakh fatalities, we would — depending on which analysis we use — expect around 7.6 lakh or 12 lakh Covid-19 deaths during 2020. Even the lower estimate implies that 80% of Covid-19 deaths in 2020 went 'missing'.

There are some important caveats. The estimates assume that all groups are equally likely to be infected. But the spread amongst India's elderly could have been lower than amongst younger people, in part because poverty is associated with wider spread, and life expectancy [decreases sharply](#) as poverty increases.

Interestingly, the third national serosurvey did *not* find [reduced infection levels in the over-60s](#). But we should not read too much into this without more granular data on the oldest groups most vulnerable to severe disease.

IFR predictions from international data also assume that factors like the extent of co-morbidities and access to healthcare are comparable between India and the countries from where the data comes. Could other factors be at play? Are Indians naturally protected from severe Covid-19 for some as yet unknown reason, which makes the international data less relevant to India?

What could Mumbai's data tell us about the national situation?

Rather than international data we can try to use Mumbai's data to infer India's Covid-19 fatality rate.

A careful analysis of 2020 data finds a [median estimate for the IFR of Covid-19](#) of 0.23% in Mumbai. Taking uncertainties into account, plausible estimates of Mumbai's 2020 IFR range from 0.15% to 0.33%. The median estimate is close to expectations from the meta-analysis of [O'Driscoll *et al*](#), while the upper estimate of 0.33% is close to expectations from [Levin *et al*](#).

|| In Mumbai, at least, we do not find strong evidence of some natural protective force at play.

In other words, the city's 2020 death toll was roughly as expected from the meta-analyses, given the extent of disease and Mumbai's age structure. In Mumbai, at least, we do not find strong evidence of some natural protective force at play.

Moreover, the city's age structure is not very different from the national one. If Mumbai's IFR estimates apply nationally, we would expect India to have seen between 5 lakh and 11 lakh Covid-19 deaths during 2020. Even the *minimum* estimate would imply that more than two-thirds of deaths were missed last year.

Summary

Both international data and Mumbai's data give estimates for expected Covid-19 deaths in India ranging between three and eight times the officially recorded death toll. These estimates overlap with [those of Bhramar Mukherjee](#), professor of biostatistics and epidemiology at the University of Michigan, who was quoted in a [CNN report](#) as estimating that India's "*covid fatalities could be underreported by a factor of between two and five*".

If indeed, India's true Covid-19 death toll is close to the recorded toll, then something astonishing is occurring: Indians are, on average, *much* less vulnerable to severe disease than others around the world. But not all Indians; rather coincidentally, this 'resilience' is highest in parts of the country where disease surveillance is weakest.

The proponents of exotic theories to explain the low recorded death toll often dismiss reports of undercounting as isolated aberrations. It is disappointing that they rarely advocate careful surveying to check if the premise of low mortality is indeed correct. If Covid-19 swept through rural Bihar and almost no one died, then surely the whole country needs to know this and understand why.

Finally, what of the second wave? [Will there be fewer deaths this time around?](#) It seems unlikely. At the time of writing, a catastrophe is unfolding. Deaths are mounting fast. Health systems are overwhelmed, and avoidable deaths are occurring as people struggle – and fail – to get oxygen or hospital beds. All the signs are that the death toll will be high, and recorded deaths even less reliable than during the first wave.

Until the political will exists for detailed and careful surveying, there can be no confident answer to the question of how many have died of Covid-19 in India. The immediate crisis will pass, but it may be years before we know the true scale of the devastation.