Notes on Bihar's COVID epidemic

(Murad Banaji 27/10/2020. Updated on 30/10/20 to take into account reports from Bihar's districts in the second national seroprevalence survey.)

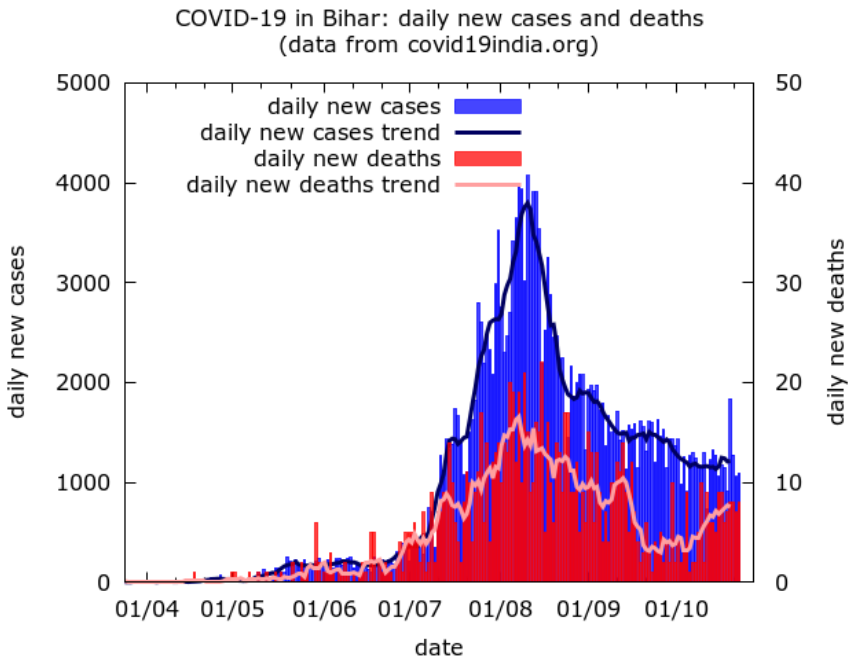
Let's start with a contrast. To date (late October)

- **Bihar**, with an estimated population of 120M, has had 2.1 lakh recorded COVID cases and just 1K recorded COVID deaths.
- **Maharashtra**, with a similar estimated population has had 16.4 lakh recorded COVID cases and 43K recorded COVID deaths.

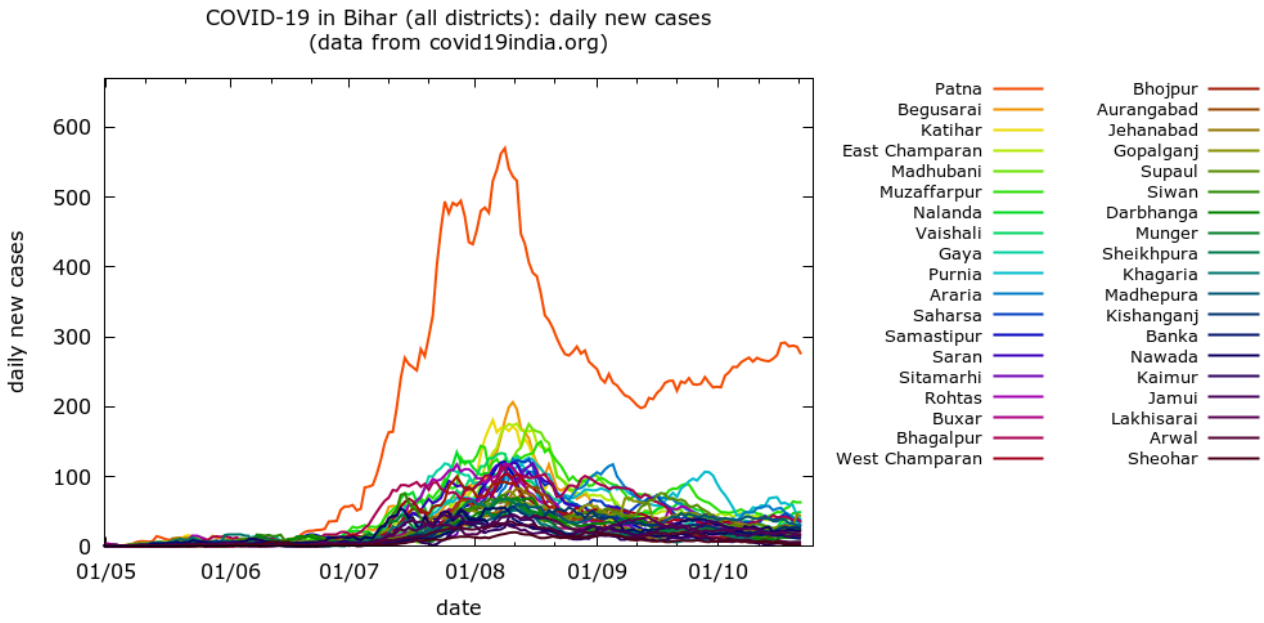
What explains such huge differences? Has Bihar's data been manipulated in the run-up to an election, as has been alleged? We'll see that there are strong signs the data has been massaged. But there are also other interesting factors in play. It may be important that Bihar is a young and predominantly rural state. Demographic data, preliminary seroprevalence data, and granular case, fatality and testing data at https://www.covid19india.org/ could help to clarify things.

A continuing epidemic

First, Bihar's COVID-19 epidemic is certainly not over. Daily cases rose from about 500 in the first week of July and peaked at about 3,800 (weekly average) in the second week of August. But the subsequent decline effectively halted by the end of September and daily cases were steady at about 1,200 through most of October. Daily recorded COVID-19 deaths peaked at around the same time as cases, came down, and have recently been rising again.

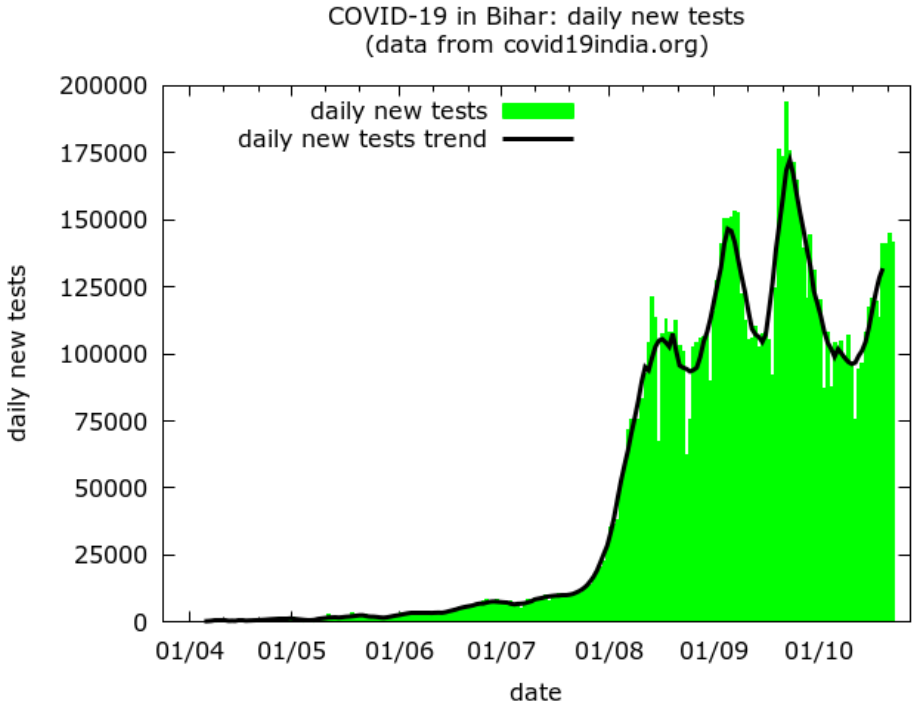


When we look at individual districts the picture is more messy – as we might expect. Not all districts have seen a clear peak in cases, while several have had more than one apparent peak. Cases in Patna district, by far the biggest contributor to Bihar's case and death data, peaked in early August, dropped sharply, but have been rising gradually through much of September and October, even as some other districts have wound down.



A huge increase in testing magnified the peak in cases

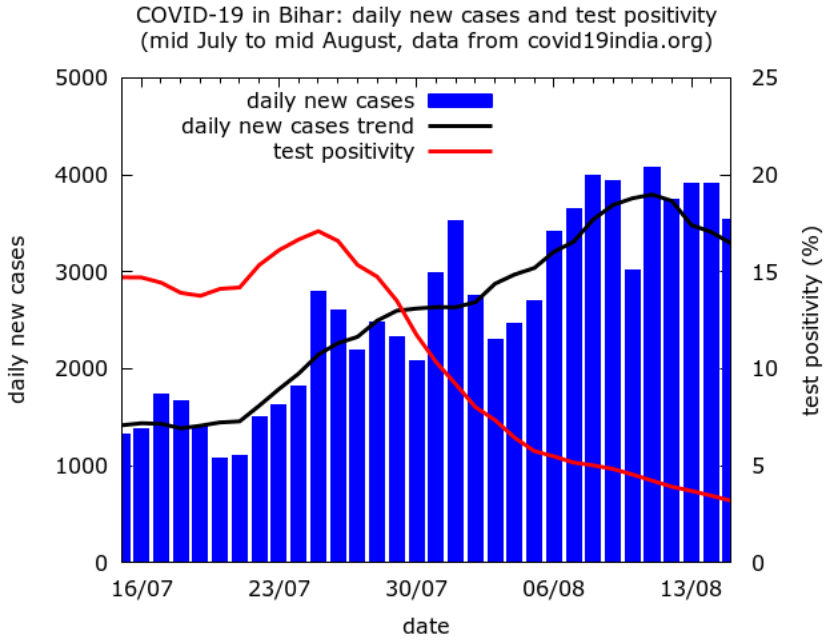
Bihar's pronounced peak in cases was exaggerated by changes in testing. There was a very sharp rise in tests in late July and early August: the total rose more than ten-fold from about 10,000 per day near July 22 to over a lakh per day one month later. Testing has stayed at a high level since then, although with some fluctuations.



It seems that a very large proportion of the new tests were rapid antigen tests (RATs). This percentage is not regularly shared, but by August 11 it was already at 86%, and at the last available

count (Sept 21) about 92% of all of Bihar's tests were RATs. What the ramp-up in testing through July and early August did was to increase the apparent speed at which Bihar's epidemic was growing – detection was almost certainly improving even as infections were also rising. If we look at fatalities instead of cases, there is still a clear rise and fall, but less dramatic.

The rise in testing and, crucially, the fact that most tests were relatively low sensitivity RATs led to a dramatic fall in test positivity (here defined as the ratio of cases to total tests) – this dropped from over 15% in late July to under 2% a month later, and is currently at around 1%. For two weeks during late July and early August test positivity was falling sharply, even as cases were rising sharply.



Unclear targeting of tests

What has the testing strategy been? There have been reports of doctors struggling to meet targets for rapid testing without any clear messaging about who should be tested. We also know that RATs miss a significant fraction of infections. This does not mean that they cannot play an important role in disease surveillance; but they are also convenient from the point of view of making test positivity fall!

The total number of tests for each district in Bihar is available at https://www.covid19india.org/ and this data allows one to examine the testing strategy a little more closely. Note, however that the total reported for all districts is about 20% less than the total for the state as a whole, namely about 2 million tests are currently missing from the district totals.

Adding to the suspicion that testing was ramped up in order to make test positivity fall, there is only a weak correlation between cases in a district and testing in that district, resulting in wide fluctuations in test positivity between districts. For example, as of October 24, Patna had seen 16% of the state's cases and 25% of the state's deaths, but only 4% of the state's tests. A strategy of

testing widely in developing hotspots would most likely lead to a stronger correlation between tests and cases.

Testing alone cannot explain Bihar's low number of cases and deaths

Could the big increase in testing have helped to control Bihar's epidemic? Despite an unclear testing strategy it seems possible that the improved detection allowed some people with COVID to isolate early and thus slowed the spread. Even a modest increase in detection would break some chains of transmission and help control the epidemic.

Although it probably helped, the increase in testing does not explain the state's relatively low total count of cases and deaths. Delhi's per capita testing, for example, is almost twice that of Bihar's, and yet Delhi's cases and deaths per million are many times higher than Bihar's.

Logically, there are two possible explanations for Bihar's low per capita case and death count: relatively low spread of infection, and poor disease surveillance.

Has Bihar seen much lower disease spread than the national average? Bihar's preliminary serosurvey data suggests the opposite. In late August, an estimated 16% of the population in six surveyed districts had developed antibodies to SARS-CoV2, considerably higher than the national average of 6.6% from the August-September serosurvey. In the surveyed districts just 24000 cases had been detected from an estimated 3.4 million infections, a very poor infection detection of about 0.7%. These districts had recorded only 72 deaths by the end of August giving a naive infection fatality rate of just 0.002%.

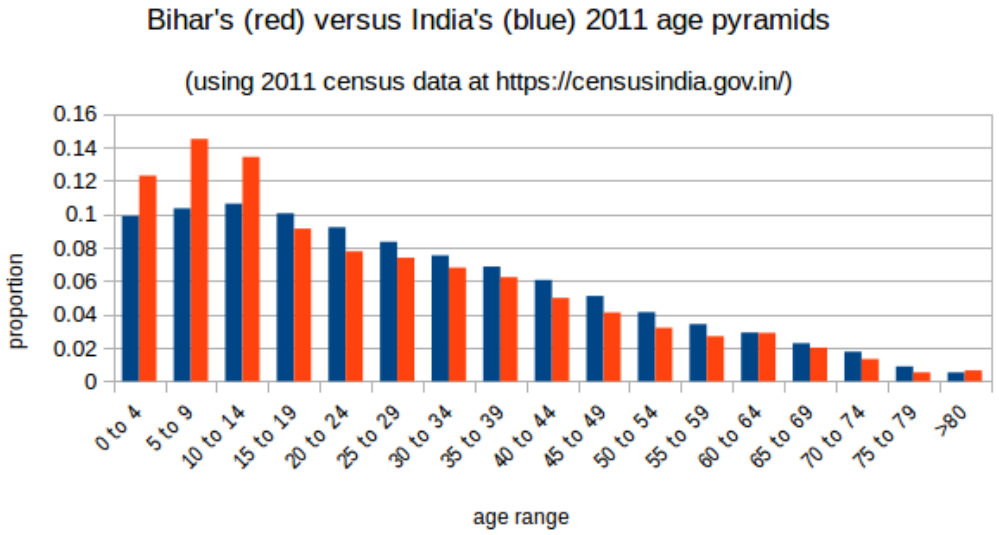
By contrast a serosurvey carried out in 10 districts of Chhattisgarh in September found about 5.5% of samples had antibodies to SARS-CoV2, which, after correction for the population sizes of the districts, gave and estimated seroprevalence of 6.7% in these districts. This means that over 6% of infections had been detected in testing. Moreover, there had been around 600 fatalities by September 25, giving a naive IFR in these districts of around 0.06%, 30 times higher than that in Bihar.

The question arises: do we really believe that COVID-19 is 30 times more deadly in Chhattisgarh than in Bihar? Or is Chhattisgarh detecting and reporting its COVID-19 fatalities more transparently? The data, although preliminary, strongly suggests that Bihar's "success story" is founded on extremely poor disease and death surveillance.

Another way of approaching the data is to attempt to infer prevalence from fatalities. In order to do this we first need an estimate of infection fatality rate (IFR) in the state.

Estimating IFR and prevalence suggests high fatality underreporting

Bihar is a young state. At the time of the 2011 census, 37.3% of the state's population was 0-14 years old, as opposed to 27.2% in Maharashtra and 29.5% nationwide. On the other hand 7.0% were over 60 as against 9.3% in Maharashtra and 8.0% nationwide. As we know, age makes a very large difference to expected COVID-19 fatality rates.



Using Bihar's 2011 age pyramid, we can calculate the state's expected COVID-19 IFR using various sources of age-adjusted IFR for COVID. We find a range of expected values for Bihar's IFR from about 0.14% to 0.28%. The lower end is derived from French data as reported here, and the upper end from the age-dependent IFR formula in the systematic review of Levin *et al* here. By comparison, using the same age-adjusted values gives an IFR range of 0.21% to 0.41% for Maharashtra – 50% higher than Bihar's. For India as a whole we get a range of 0.18% to 0.35% using the same sources.

Note that these estimates are very rough and rely on partial data. The population will have aged somewhat since 2011, pushing expected IFR up; meanwhile, a factor pushing expected IFR down would be lower prevalence amongst the elderly, as can be inferred, for example, from Mumbai's serosurvey.

In order to estimate prevalence using IFR estimates we need either to take fatalities at face value, or to make some assumptions about fatality underreporting. The outcome of this process for Bihar and Maharashtra is summarised in the table below. Case and death data is from https://www.covid19india.org/

	Bihar	**Maharashtra**
Estimated 2020 population	121M	123M
Recorded COVID-19 cases (October 24)	2,11,443	16,38,961
Recorded COVID-19 deaths (October 24)	1,042	43,152
IFR estimates (lower, mid-point, upper)	0.14%, 0.21%, 0.28%	0.21%, 0.31%, 0.41%
Assumed fatality undercounting	67%	50%
Estimated prevalence (based on upper, mid-point, lower IFR values respectively)	1.1M, 1.5M, 2.2M	21M, 28M, 41M
% of infections detected (based on lower, mid-point, and upper IFR values)	9.5%, 14%, 20%	4.0%, 5.9%, 7.8%

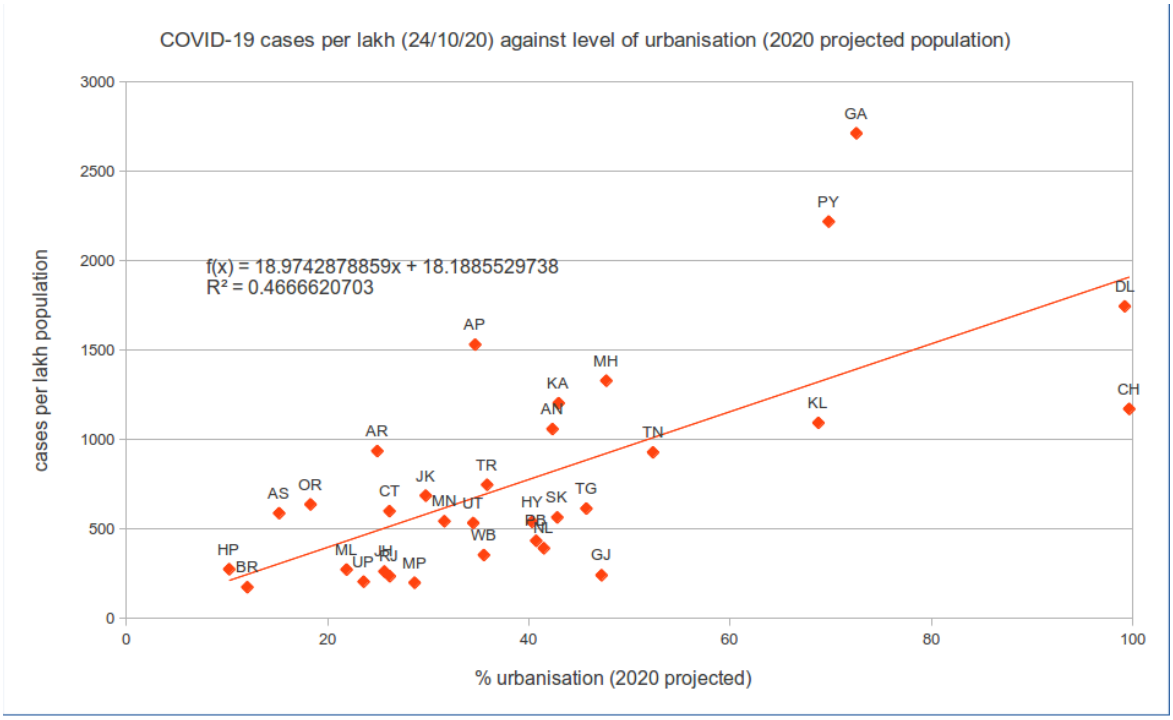
Let’s focus on the estimates which we arrive at by using mid-point estimates of IFR. In the case of Bihar, if we assume that about 1 in 3 COVID deaths have been counted, and take IFR to be 0.21% this would imply that by early October there had been about 1.5M infections in the state. In other words, only about 1.2% of the population would have had COVID, much lower than the national estimate from the August-September serosurvey. But this estimate is entirely inconsistent with the 16% prevalence recorded in the serosurvey. *Another sign that the vast majority of Bihar’s deaths may be missing from official figures.*

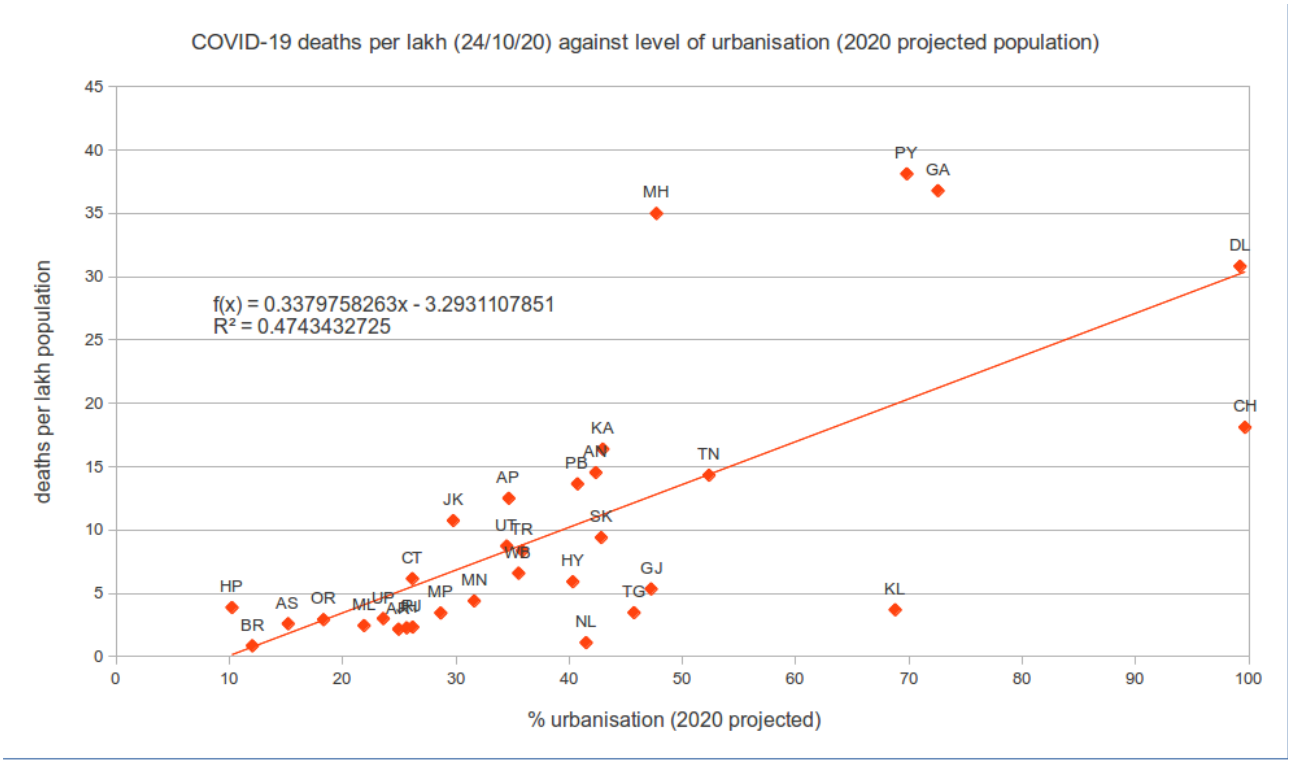
By contrast, if we suppose that Maharashtra, with its better death surveillance, has counted about 1 in 2 of its fatalities, then taking IFR to be 0.31%, there would have been 28M infections in the state by early October. This would mean that about 23% of people in the state have had COVID, much higher than the estimated national figure. This would also mean that 5.9% of infections have been detected, which is plausible.

The urban-rural divide

Other than age, a second factor stands out in Bihar. According to 2020 projections Bihar's population is 88% rural (Maharashtra 52%, Chhattisgarh 74%, India 66%). This makes Bihar nearly the most rural state in the country, second only to Himachal Pradesh.

The level of urbanisation certainly seems to matter to the level of cases and deaths reported. If we look across Indian states and territories, we find that urbanisation (in the census definition) is a significant predictor of the number of COVID-19 cases and deaths in a region. In fact, the relationship, shown in the next two plots, is surprisingly pronounced given different epidemic timings, and probably very big differences between states in disease surveillance, mitigation measures, and fatality undercounting.





(In the plots above 2020 projected urbanisation levels are taken from here, case and death data is from https://www.covid19india.org/, and states/regions are coded as follows: JK = Jammu & Kashmir, HP = Himachal Pradesh, PB = Punjab, CH = Chandigarh, UT = Uttarakhand, HY = Haryana, DL = NCT Of Delhi, RJ = Rajasthan, UP = Uttar Pradesh, BR = Bihar, SK = Sikkim, AR = Arunachal Pradesh, NL = Nagaland, MN = Manipur, TR = Tripura, ML = Meghalaya, AS = Assam, WB = West Bengal, JH = Jharkhand, OR = ODISHA, CT = Chhattisgarh, MP = Madhya Pradesh, GJ = Gujarat, MH = Maharashtra, AP = Andhra Pradesh, KA = Karnataka, GA = Goa, KL = Kerala, TN = Tamil Nadu, PY = Puducherry, AN = Andaman & Nicobar Islands, and TG = Telangana. Mizoram which has reported no deaths, Dadra and Nagar Haveli and Daman and Diu which has reported 2 deaths, and Lakshadweep the only territory to report no cases so far, are omitted in order to avoid three small territories with possibly special COVID stories to overly influence the data. For completeness the plots with these three territories included are given in an Appendix at the end.)

Once we take into account the urban-rural divide, Bihar no longer appears as an outlier in India's COVID-19 data. In fact, both its cases and deaths lie close to the lines given by regression analysis for the country as a whole. Even within the state the urban-rural divide seems clear: Patna, Bihar's most urban district, with 5.6% of Bihar's population has reported 16% of its cases and 25% of its deaths.

Maharashtra, on the other hand, stands as something of an outlier in national data, particularly when it comes to fatalities. We should remember that Maharashtra's numbers include data from several long and harsh city epidemics. It is also possible that fatality undercounting in Maharashtra is below the national average, given relatively good indicators for death surveillance, including the large number of institutions reporting medical certification of cause of death (MCCD).

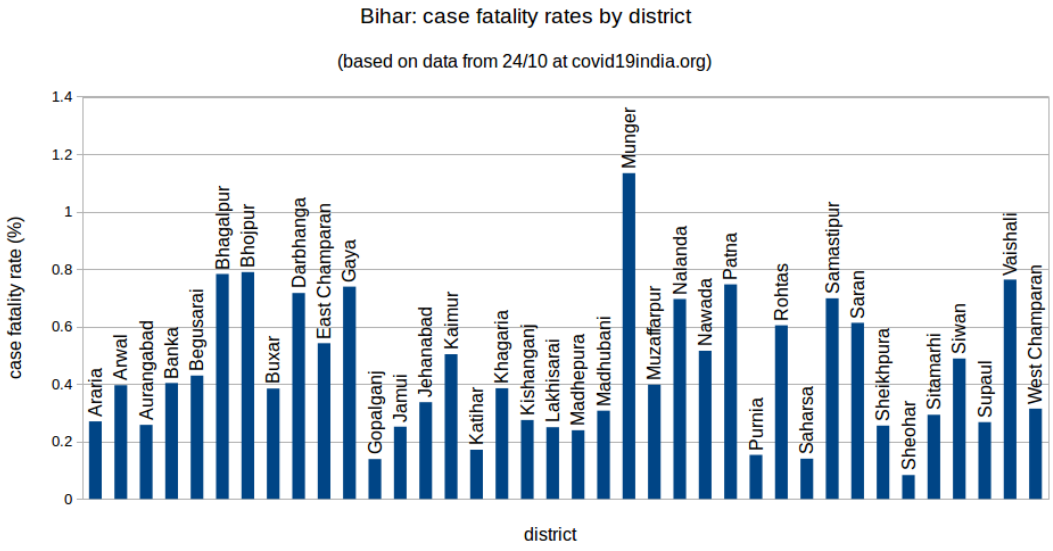
While the relationship between urbanisation and cases/deaths is strong, it is important to stress that we don't fully understand what drives it. In particular, to what extent is the spread of disease in rural areas is slowed by lower population density and different modes of interaction, and to what extent are fewer recorded cases and deaths a consequence of poorer surveillance? Again, Bihar's serosurvey data strong suggests that the answer is **poor surveillance**.

Missed fatalities?

As we saw above, either Bihar's IFR is much lower than values estimated from international and national data for some as yet unknown reason, or the vast majority of its COVID-19 fatalities are missing.

How could a very large percentage of COVID-19 deaths have never showed up in official figures? We know that large numbers of deaths can be missed nationwide: for example, the WHO estimates about 4.4 lakh TB deaths in India every year, but only about one fifth of these get recorded as such. It is also worth noting that, as of 2018, only about 14% of registered deaths were medically certified in Bihar.

If we look at district-wise data, case fatality rates (CFRs - the ratio of confirmed deaths to confirmed cases) fluctuate widely by district.



This is not just a consequence of fluctuations in case detection on account of the wide variation in testing between districts as discussed above. For example, Bhagalpur recorded 8307 cases from 281K tests, and had 65 confirmed deaths; but Gopalganj recorded 5055 cases from 210K tests, but had only 7 confirmed deaths. Test positivity in these two districts is comparable, but case fatality is not. Across all districts, there is only a very weak correlation between case fatality rates and test positivity rates, suggesting that the variation in CFRs between districts could also be about uneven death reporting.

Summary

Bihar's COVID-19 epidemic has risen and fallen, and then stabilised. A significant number of cases are still being reported, and it is currently unclear if this number is diminishing. The state's big increase in testing, with a high volume of rapid antigen tests, created the appearance of a dramatic rise in infections and fall in test positivity. The testing strategy was unclear, but it could have played some part in controlling the epidemic.

Has Bihar's data been manipulated to manufacture a success story? Very likely. There are signs of poorly targeted testing, possibly aimed at reducing test positivity rather than better disease surveillance. There are also indications that death recording is not uniform across districts, with wide variation in reported deaths between districts despite similar case numbers. Most tellingly, seroprevalence data indicates that a very small fraction of infections have been detected, and probably a very high fraction of deaths have not been recorded as such.

Bihar's low numbers of cases and deaths are explained by some combination of

- relatively low spread in a predominantly rural population,
- low IFR and more mild cases as a consequence of a young population, and
- poor detection of infections and high fatality undercounting.

Seroprevalence data strongly suggests that the last of these is the dominant effect. We need to be cautious, of course, because only six of 38 districts were surveyed. With regard to fatalities, even if excess mortality data were to become available, it is possible that COVID-19 deaths would not show up clearly above the noise, given an estimated 60K deaths per month in Bihar.

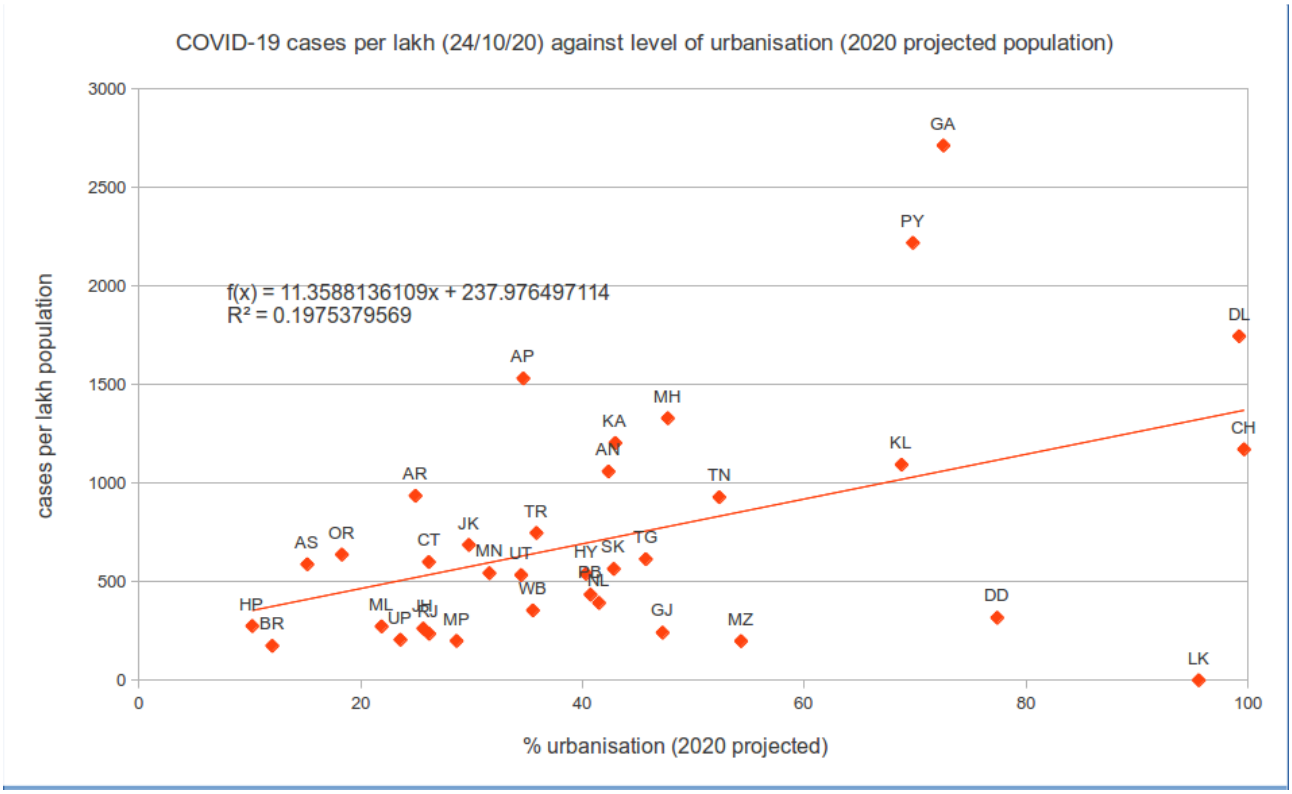
But with these caveats in mind, the data strongly suggests that Bihar's "success story" is founded on extremely poor disease and death surveillance.

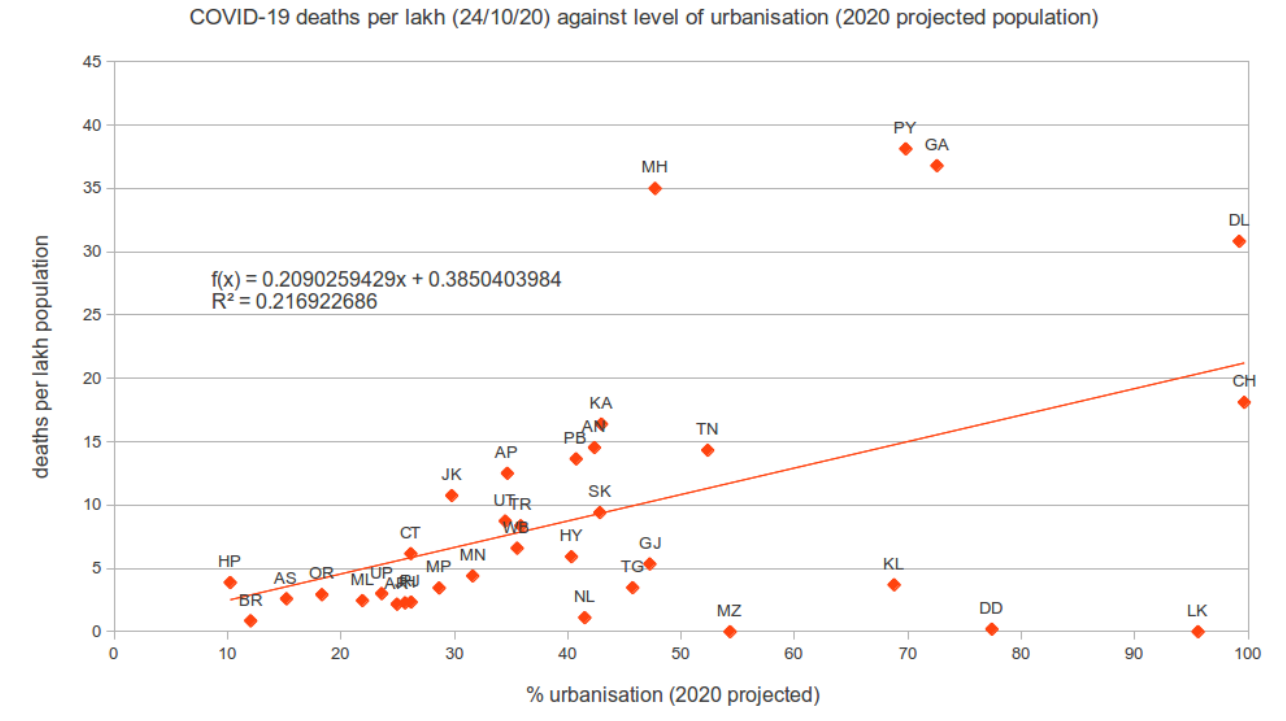
Aside from this, Bihar's data highlights a strong connection between urbanisation and low COVID numbers which warrants further exploration. It suggests that this relationship may be more about weak surveillance than low spread.

Could Bihar see a resurgence? With prior infection levels probably still below 25%, the majority are still susceptible. There is no reason to believe that there cannot be a resurgence driven, for example, by the easing of restrictions, festival season, and huge election rallies.

Appendix

The relationship between urbanisation and cases/deaths with all territories included is shown in the two plots below.





If we include only territories with estimated 2020 population greater than 1.5M, then we obtain the following plots. The relationship is even more pronounced in this case.