Notes on Delhi's COVID-19 seroprevalence surveys: estimating prevalence, infection detection, and naive infection fatality rate

(Murad Banaji, 13/11/20)

Residents of Delhi would probably like to know the answers to a number of questions. For example:

1. How many have had COVID in the city?
2. What drove the second surge in cases in late August and September, and what is driving the huge third surge (currently ongoing in mid-November). When might is wind down?
3. What is the city's infection fatality rate (IFR) – how many of those infected are dying – and is this proportion rising or falling?
4. How has the disease differentially impacted different age-groups and people living in different kinds of housing?

Unfortunately, very limited data is available in Delhi bulletins to shed light on such questions. (This stands in contrast to Mumbai's detailed daily COVID bulletins, although there too, there are some gaps.) Could the city's four serosurveys from July, August, September, and October fill in some of the gaps?

The surveys are certainly a valuable addition to the data, but attempts to get deeper insights into Delhi's COVID-19 epidemic using serosurvey data are hindered by poor planning and transparency. Here are some of the issues:

- There appears to be no publicly available technical documentation on the serosurveys. This means there is very little detail about sampling. From what has been reported, it seems that both sampling methodologies and test-kits changed between the surveys. Thus the prevalence values obtained in different surveys are not easy to compare – even biases may not be consistent across the surveys.
- As we know from Mumbai and Pune, the nature of housing can be a key predictor of seroprevalence. Reports from Delhi's first two surveys (early July and early August) did not seem to distinguish between dwellers of different kinds of housing – e.g., slums or non-slums. Unrepresentative sampling in different strata of housing could have biased the results. In the third survey (early September), five different kinds of dwelling were treated separately, but again detail is not available on how residents of different dwellings were sampled.
- There were differences reported in seropositivity by gender, age and (in the third serosurvey) by housing type. However, it is unclear if the sampled populations correctly reflected the age, housing, and gender profile of the city. On the other hand, the seropositivity values reported do not seem to have been adjusted for sampling weights. So one cannot rule out possible biases associated with sampling a population which did not correctly reflect the demographics of the city in terms of age, gender, or residence.
- So far, almost no detail has been made available about the results of the fourth survey.

Some of the problems with the conduct of the serosurveys, transparency, and the way the results were shared (or not shared) are highlighted in comments on the serosurveys by a bench of Delhi High Court judges. According to the Delhi government, this dispute resulted in a delay to the fourth serosurvey, which was, however, eventually completed in the third week of October.

Other issues with Delhi's COVID data

Another factor complicating analysis of Delhi's COVID epidemic was major fatality undercounting during April and May followed by reconciliation in June. This means that the early trajectory of recorded fatalities does not reflect early spread in the city and cannot be used to infer early dynamics of the epidemic. In fact, Delhi's fatality data has raised questions more recently too.

Accurate and granular fatality data would have helped to understand the scale and detail of the first wave: which groups were hit, and which remained vulnerable to infection. In the case of Mumbai, knowing the age profile of recorded COVID-19 fatalities provided extremely valuable insights into progression of the disease in the city.

Although not a problem *per se*, Delhi has also seen a widely varying level and profile of testing over time. Test numbers, and the proportion of rapid antigen tests in the total have seen several major changes. This means that test positivity is a metric which is hard to interpret, and every increase or decrease in cases needs to be interpreted with some care.

The serosurveys

The headline figures from the serosurveys have been reported as follows: the first survey was carried out from June 27 to July 10, involved 21,387 samples, and found a seropositivity of about 23%. The second was carried out from August 1 to 7, involved 15,239 samples, and found a seropositivity of about 29%. The third was carried out from Sept 1 to 15, involved 17,197 samples, and found a seropositivity of about 25%. The fourth survey was carried out from Oct 15 to Oct 21, 15,015 people were tested and found a seropositivity of around 25%.

The Kavach test-kit was used in the first two surveys. However, in the third serosurvey, a different test-kit was used: "Erbalisa" (reported here). Values of sensitivity and specificity are given as 92.1%/97.7% for Kavach, and 99.12%/99.33% for Erbalisa in the Hindustan Times (although it is unclear if there are in fact two different kits with the same name – different values of sensitivity and specificity are reported here). No information is currently available on the test-kit used in the fourth survey.

Adding to the confusion, reports gave values of seropositivity either corrected or uncorrected for sensitivity and specificity, and there appeared to be errors in the corrected values reported after the first survey. In any case, with measured values of seropositivity between 20% and 30% in all three surveys, correction for sensitivity and specificity makes a relatively small difference to the raw values.

Inferring prevalence taking into account waning sensitivity

Delhi's serosurvey data only makes sense if we assume "waning antibodies": older infections are less likely to be detected by antibody tests. Since this phenomenon seems to depend at least in part on the antibody tests used, we refer to it as waning test sensitivity. We can try to take account of waning sensitivity and infer prevalence in the city following an approach developed by Buss *et al* to analyse Brazilian data, and previously used on Mumbai's data here. This approach was developed for a different antibody test – the Abbott test – but it can be applied to any test.

It is worth abbreviating and distinguishing the following quantities associated with a serosurvey:

1. **MSP**: the measured sample seroprevalence (let's say after correction for reported test sensitivity and specificity, although as noted above this does not make a great deal of difference in this case).
2. **MPP**: the measured population seroprevalence: the seroprevalence that would have been obtained if the entire population had been tested using the given test. Estimating this from the MSP would involve correcting for possible sampling biases such as overrepresentation of certain age groups.
3. **TPP**: the "true" population seroprevalence: the seroprevalence that would be observed in the sampled population if the entire population were tested, *and*, *moreover,* tests were 100% sensitive to prior infection (after seroconversion) and this sensitivity did not wane over time.

We assume that MSP reasonably well approximates MPP, but because of a lack of detail on sampling in Delhi's serosurveys, we still treat MPP in the four surveys as a random variable. The idea is that TPP can then be estimated from MPP by taking into account the rate of decline of test sensitivity as follows.

Since we have no basis to assume otherwise, we assume that properties of the named tests, Kavach and Erbalisa, and whatever test was used in the fourth serosurvey, are similar with regard to waning sensitivity.

We require two parameters associated with the waning antibodies and the tests: p, the proportion of those who serorevert; and α, the monthly attenuation factor. To understand these parameters, suppose that 1000 people are infected on a given date, and – for the sake of argument – all are found seropositive by the anitbody tests some time later which we'll call day 0. Suppose that p=0.95 and α=0.7 – what would this mean? p=0.95 would mean that of the 1000 infected individuals, about 50 will never serorevert and will have measurable antibodies (with this test) for all time; α=0.7 would mean that of the remaining 950, about 30% will have seroreverted a month after day 0, another 30% of the remainder will have seroreverted by the end of the next month, and so forth.

Lost fraction. If we have estimates for MPP and TPP at a given moment, then we can estimate the fraction of true seropositives lost to waning sensitivity, termed the "lost fraction": (TPP-MPP)/TPP. This quantity can be used to compare computed data with data from experiments where samples of recovered COVID patients are tested for antibodies some time later.

In order to get a picture of the epidemic which makes sense as whole, it is a good idea also to track what prevalence estimates tell us about:

- **Infection detection** (ID) – here taken to be the ratio of cases on a given date to estimated TPP one week later. The delay is chosen on the assumption that cases are typically recorded at a point in the infection cycle about one week before seroconversion.
- **Naive infection fatality rate** (IFR), here taken as the ratio of recorded COVID-19 fatalities on a given date to estimated TPP four days earlier. The delay is chosen on the assumption that deaths are typically recorded at a point in the infection cycle about four days after seroconversion.

Both cumulative and recent values of ID and naive IFR can be computed. Cumulative values are described above; recent values take the change in cases/deaths between surveys (appropriately delayed) as the numerator, and the change in estimated TPP between surveys as the denominator.

In calculations of ID and naive IFR, Delhi's population is taken to be 19 million.

Initial experiments

The four serosurvey mid-points (July 4, August 4, September 8, October 18) are approximately 1 month apart and we round the MSP values obtained to 23%, 29%, 25% and 25%. We take time-steps of 34 to 35 days, and set the first survey midpoint to be July 1 (it is assumed that MSP on July 1 would also have been 23%), and set the final survey midpoint to be October 12 (it is assumed that MSP on October 12 would also have been 25%). We augment the data by introducing two additional hypothetical serosurveys with midpoints on April 23 and May 27. This gives a total of 6 time-points – and six surveys which we'll number 1, 2, **3**, **4**, **5**, **6** – to work with (actual surveys in bold).

There is, of course, high uncertainty in what MPP in the hypothetical serosurveys might have been, and also some uncertainty in the estimates of MPP from MSP in the three actual serosurveys. We will eventually deal with these uncertainties by putting distributions on the values of MPP at each survey. First, to get a feeling for the process, let's consider four examples.

In the following examples we fix p=0.95 and the estimates of MPP for surveys 1 to 6, varying only α. The value of α given is the monthly attenuation factor – in calculations this needs to be scaled to take into account the fact that the time-steps are a little more than a month long.

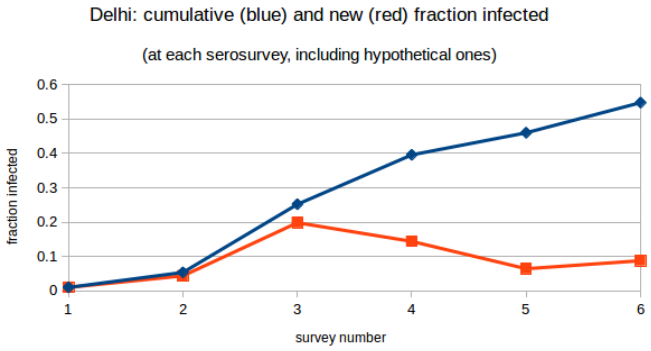
Case and fatality data are taken from https://www.covid19india.org/.

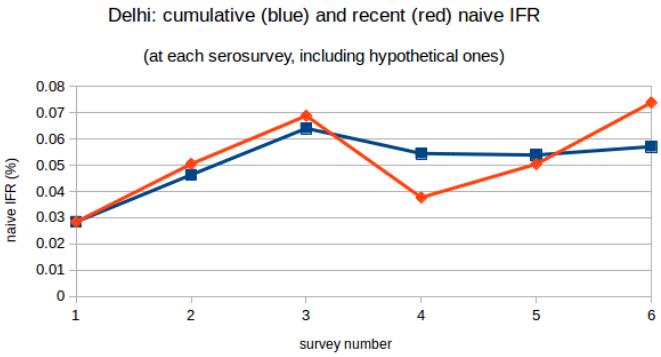
Experiment 1: rapid waning

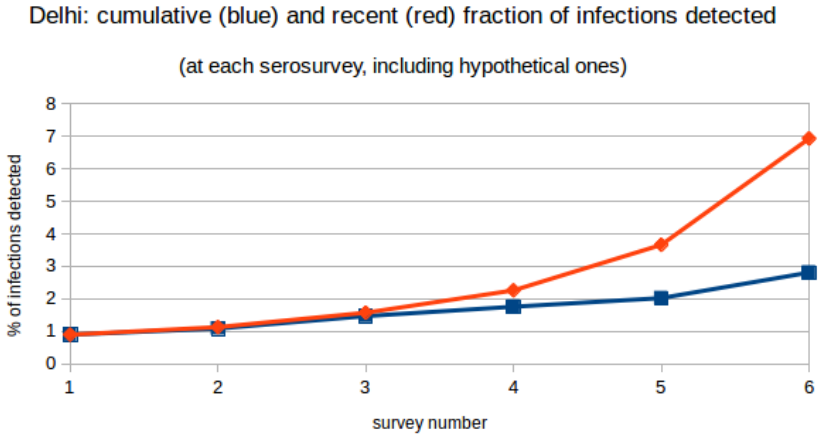
α=0.65.

date	MPP %	cases (one week earlier)	recorded deaths (4 days later)	TPP % (lost fraction)	Naive IFR % (new)	ID % (new)
23/04	1	1707	54	1 (0)	0.028 (0.028)	0.90 (0.90)
27/05	5	11088	473	5 (0.068)	0.046 (0.050)	1.09 (1.13)
01/07	23	70390	3067	25 (0.087)	0.064 (0.069)	1.47 (1.57)
04/08	29	132275	4098	40 (0.27)	0.054 (0.038)	1.76 (2.26)
08/09	25	177060	4715	46 (0.46)	0.054 (0.050)	2.02 (3.66)
12/10	25	292560	5946	55 (0.54)	0.057 (0.074)	2.81 (6.93)

The experiment can be summarised in 3 graphs. There had been a slight resurgence by mid-October, with mid-October TPP about 55%. Naive IFR had risen steadily over the past two cycles, reaching above 0.07% by mid-October. Note that the very low early value of naive IFR followed by an overshoot is consistent with systematic early fatality undercounting followed by reconciliation. Infection detection has risen steadily over the past several cycles.



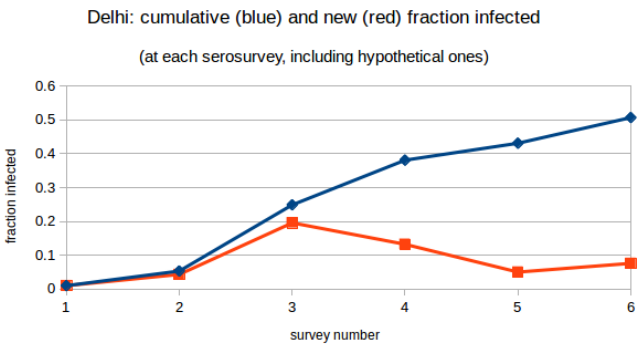


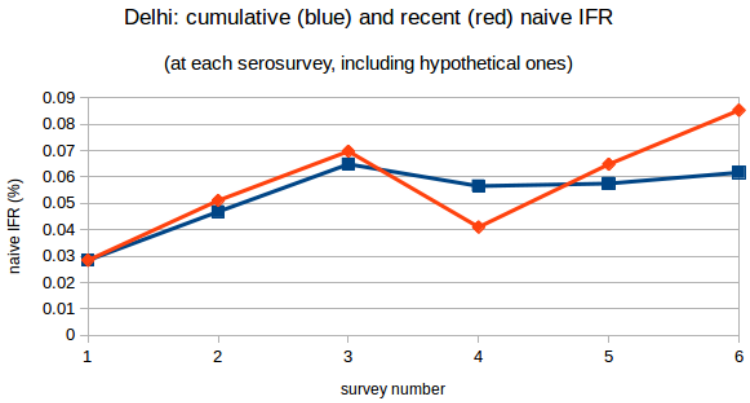


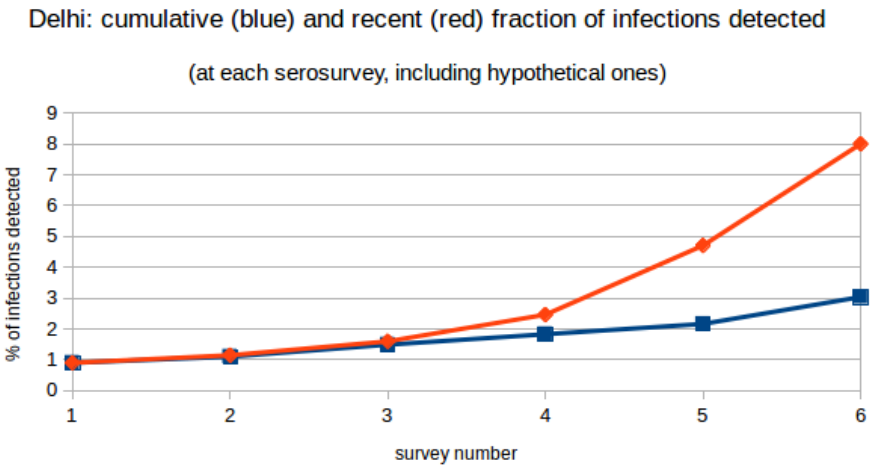
Experiment 2: medium waning

α=0.7.

date	MPP %	cases (one week earlier)	recorded deaths (4 days later)	TPP % (lost fraction)	naive IFR % (new)	Case detection % (new)
23/04	1	1707	54	1 (0)	0.028 (0.028)	0.90 (0.90)
27/05	5	11088	473	5 (0.059)	0.047 (0.051)	1.10 (1.14)
01/07	23	70390	3067	25 (0.076)	0.064 (0.070)	1.49 (1.59)
04/08	29	132275	4098	38 (0.24)	0.056 (0.041)	1.83 (2.46)
08/09	25	177060	4715	43 (0.42)	0.058 (0.064)	2.16 (4.71)
12/10	25	292560	5946	51 (0.51)	0.062 (0.085)	3.03 (8.00)





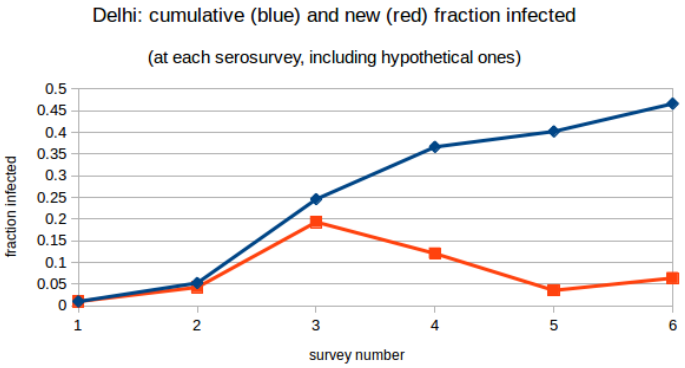


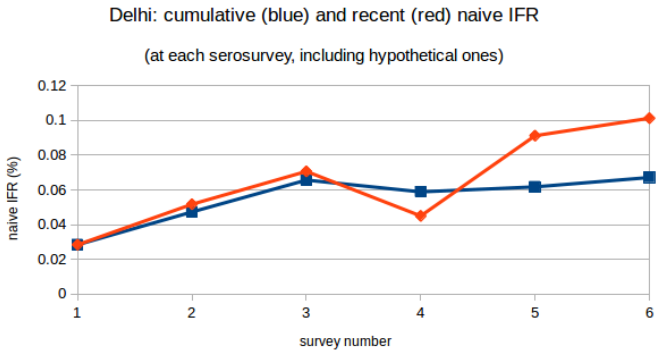
Again, there had been only a slight resurgence by mid-October with mid-October TPP at about 51%. There has been a more significant rise in infection detection and IFR over the past two cycles than in the previous example. Current IFR is at over 0.08%

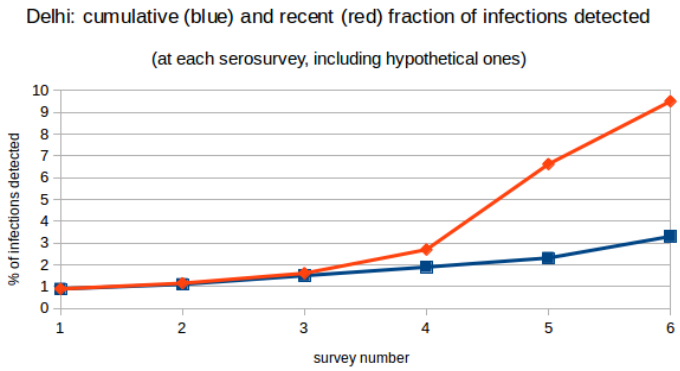
Experiment 3: fairly slow waning

α=0.75

date	MPP %	cases (one week earlier)	recorded deaths (4 days later)	TPP % (lost fraction)	naive IFR % (new)	case detection % (new)
23/04	1	1707	54	1 (0)	0.028 (0.028)	0.90 (0.90)
27/05	5	11088	473	5 (0.050)	0.047 (0.052)	1.11 (1.16)
01/07	23	70390	3067	25 (0.064)	0.066 (0.071)	1.51 (1.62)
04/08	29	132275	4098	37 (0.21)	0.059 (0.045)	1.90 (2.70)
08/09	25	177060	4715	40 (0.38)	0.062 (0.091)	2.32 (6.62)
12/10	25	292560	5946	47 (0.46)	0.067 (0.10)	3.30 (9.50)





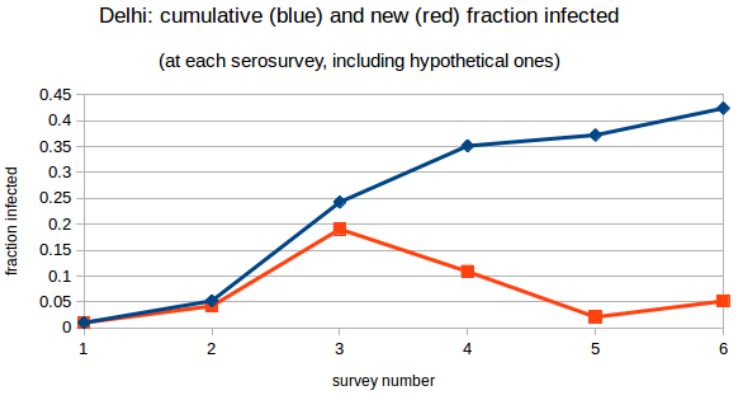


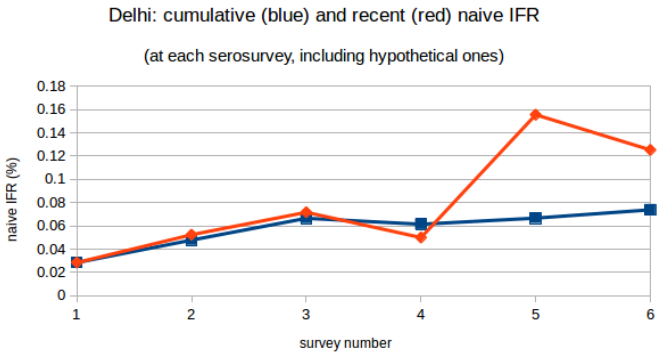
With these parameters, there is again a noticeable resurgence, and mid-October TPP is about 47%. Infection detection and naive IFR have both been rising sharply over the past two cycles.

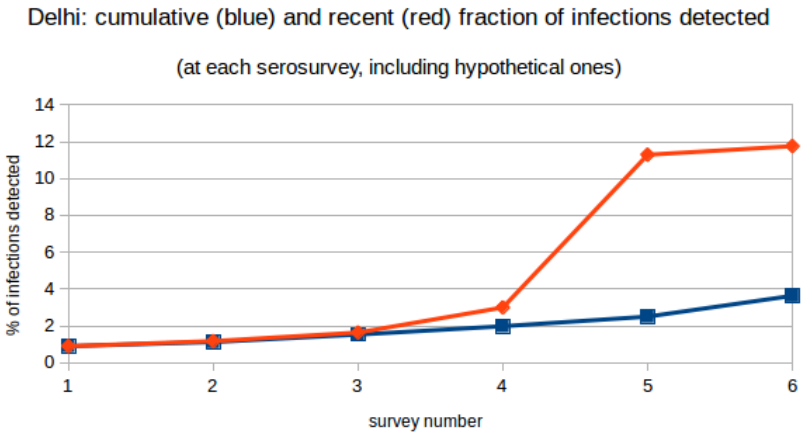
Experiment 4: slow waning

α=0.8.

date	MPP %	cases (one week earlier)	recorded deaths (4 days later)	TPP % (lost fraction)	naive IFR % (new)	case detection % (new)
23/04	1	1707	54	1 (0)	0.028 (0.028)	0.90 (0.90)
27/05	5	11088	473	5 (0.041)	0.048 (0.052)	1.12 (1.17)
01/07	23	70390	3067	24 (0.052)	0.067 (0.072)	1.53 (1.64)
04/08	29	132275	4098	35 (0.17)	0.061 (0.050)	1.98 (3.00)
08/09	25	177060	4715	37 (0.33)	0.067 (0.16)	2.50 (11.3)
12/10	25	292560	5946	42 (0.41)	0.074 (0.13)	3.63 (11.8)







With these parameters, there has again been a resurgence, and mid-October TPP is about 42%. Infection detection has risen very significantly over the past two cycles, primarily between surveys 5 and 6. Naive IFR rose very significantly between surveys 4 and 5 and then dropped between surveys 5 and 6, but remains at levels much higher than seen upto August.

At all parameters explored above, there had been a resurgence in cases – i.e., new infections added between September and October outnumbered new infections added between August and September. Another feature constant across these experiments was a rise in IFR across the last two cycles. The rise has been more or less steady, and more or less dramatic. The same holds for infection detection.

The slower test sensitivity wanes, the greater the recent rise in IFR and infection detection has been.

What is the value of α?

Since α, the monthly attenuation in sensitivity of the tests to prior infection, is such a crucial parameter, do we have any indication of its value? We can put an approximate upper-bound on α: if p=0.95 (close to the value obtained from Brazilian data), and with the other parameter values used in the experiments above, setting α to about 0.84 gives rise to unrealistic levels of case detection in the period between the August and September surveys (greater than 25%). Setting α=0.87 gives rise to no change in TPP between the August and September serosurveys. This can be regarded as an approximate upper bound on α if we trust the MPP values used.

There is also a little data we can use to say more. Recall that the "lost fraction" tells us how much lower MPP is than TPP. In the experiments above, at survey 4 (early August), the lost fraction varies from 0.17 (i.e., MPP 17% lower than TPP – 17% of prior infections "lost") when α=0.8, to 0.27 (i.e., 27% of prior infections lost) when α=0.65. In the August serosurvey a sample of recovered COVID-19 patients were tested for IgG antibodies to SARS-CoV2, and it was found that 79 of the 257 recovered patients (30%) were seronegative (presumably using the Kavach test). This suggests that waning is quite rapid and values of α may be closer to 0.65 than to 0.8; but we should, of course, be cautious about drawing conclusions especially as crucial information is missing about *when* the 79 seronegative individuals had previously tested positive for COVID.

In the October survey (survey 6), 43.5% of recovered COVID-19 patients who were tested for antibodies were found to be seronegative. The sample size is not given, and the test used in this survey is not mentioned in the reports; so we should be cautious about comparing with the previous value. But it is reassuring that at least we see an increase in this number consistent with the decreasing sensitivity of the tests to prior infection over time.

Monte Carlo experiments

We can carry out more complete Monte Carlo experiments, putting distributions on α, *p*, and MPP values. We put beta distributions on α, *p* and the MPP values for suveys 3, 4, 5 and 6 (the actual surveys); and uniform distributions using guessed intervals on the MPP values for the first two (hypothetical) surveys. We reject experiments which give unrealistic data in any of the following forms:

- a change in TPP (between any pair of surveys) of less than 0.1%, or
- a recent naive IFR outside the interval [0.01%, 0.3%], or
- a recent ID of greater than 30%.

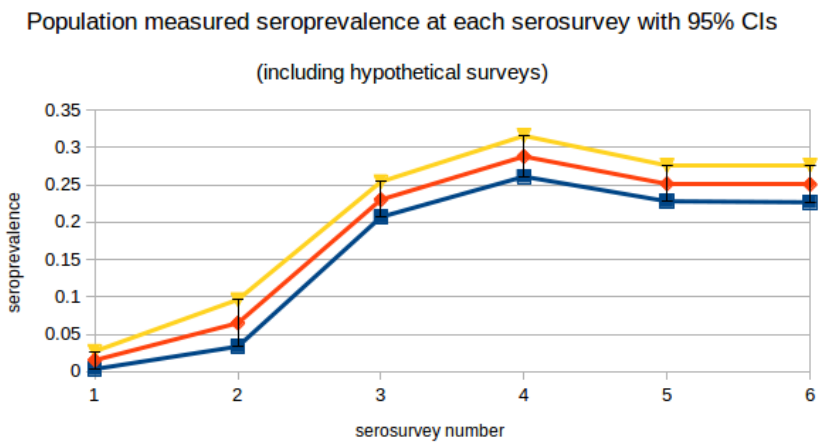
These rules mean that the actual mean values of parameters in the experiments may not match the priors.

We use the following priors in the experiments to follow:

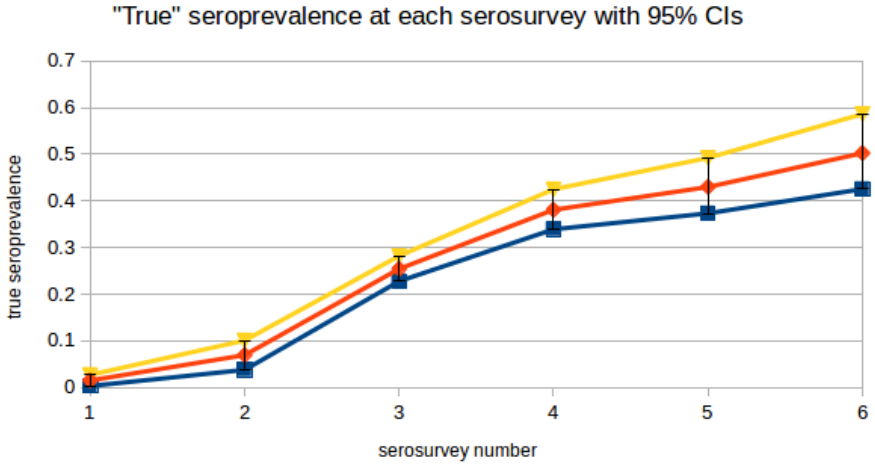
parameter	**distribution**	**distribution parameters**
α (monthly)	beta	mean: 0.725, second shape: 60, leading to SD of 0.0488
p	beta	mean: 0.95, second shape: 2 leading to SD of 0.1237
MPP 1 (23/04)	uniform	on interval [0.002,0.03]
MPP 2 (27/05)	uniform	on interval [0.03,0.1]
MPP 3 (01/07)	beta	mean: 0.23, second shape: 200, leading to SD of 0.0143
MPP 4 (04/08)	beta	mean: 0.29, second shape: 200, leading to SD of 0.0173
MPP 5 (08/09)	beta	mean: 0.25, second shape: 200, leading to SD of 0.0153
MPP 6 (12/10)	beta	mean: 0.25, second shape: 200, leading to SD of 0.0153

Running 10,000 experiments gives the following trajectories for MPP, TPP, change in TPP at each step, cumulative naive IFR, most recent naive IFR, cumulative infection detection, and most recent infection detection.

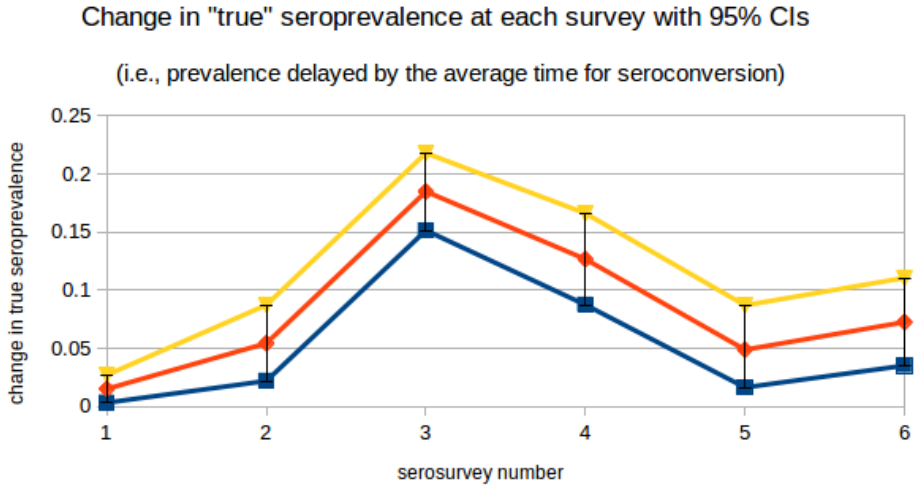
MPP. Values of MPP closely reflect the priors.



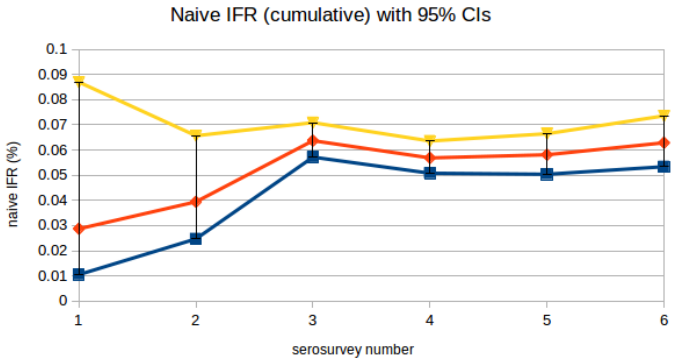
TPP. We find a mean of 43% (95% CI: 38-49%) true seroprevalence in early September and 50% (95% CI: 43-59%) true seroprevalence by October 12 reflecting levels of prior infection a few weeks earlier.

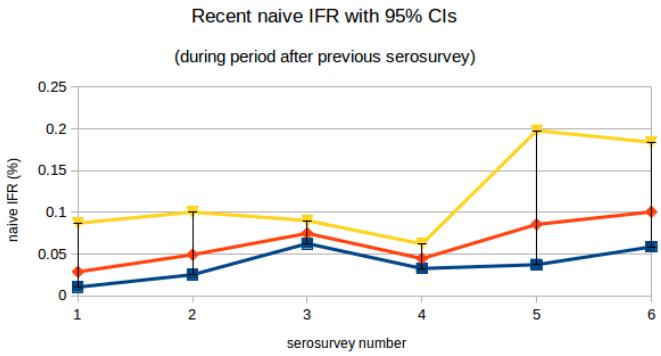


Change in TPP. The priors used are consistent with the claim that the latest surge is real, and not simply a function of some change in testing. About 80% of experiments give the most recent increase in TPP (i.e., mid-September to mid-October) as greater than the previous increase.

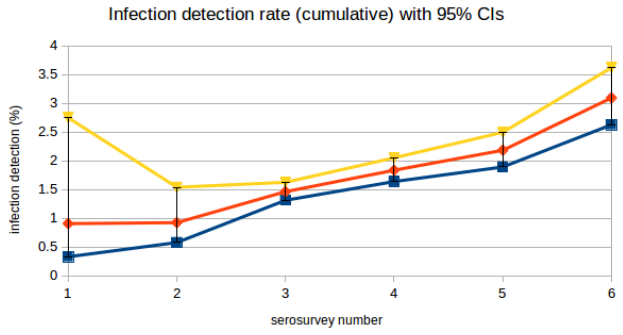


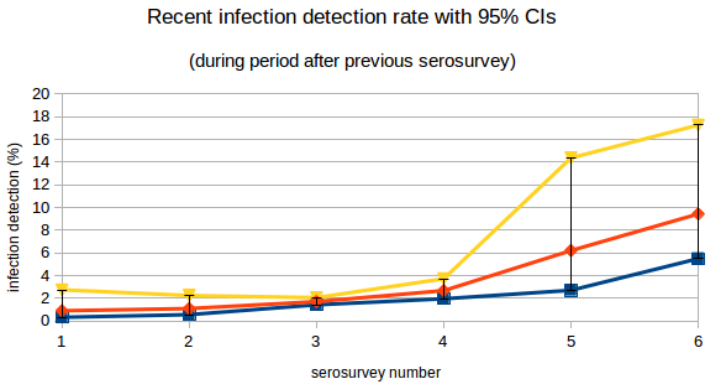
Cumulative and recent naive IFR. The mid-October cumulative value of naive IFR is about 0.062% (95% CI: 0.053-0.073%). However, there is a high degree of uncertainty on the most recent values, which could be considerably higher. About 96% of simulations show an increase in naive (cumulative) IFR between surveys 5 and 6.





Infection detection. There is a very clear increase in ID over time. The mid-October cumulative value is 3.1% (95% CI: 2.6-3.6%), but the most recent values are considerably higher. As with IFR, there is much higher uncertainty on the two most recent values: 6.0% (95% CI: 2.7-13.9%) for early September and 9.2% (95% CI: 5.4-16.7%) for early October. 100% of experiments give an increase in ID between surveys 4 and 5, and then again between surveys 5 and 6, although the scale of the estimated increase varies greatly.





If we take the estimated mid-October infection detection of 9.2% to have remained steady, then with about 1.6 lakh cases in the month to mid-November, about 1.7 million individuals – roughlly 9% of the city's population – would have been infected between mid-October and mid-November. This seems somewhat high, and we can speculate that detection was actually higher, or has further improved since the October serosurvey. But, however we look at it, the city is experiencing a remarkable surge in infection so late in its epidemic.

Conclusions

What can we conclude from the analysis? First, there are probably interesting and important stories occurring in Delhi connected with uneven spread and changes in fatality rate, but these are hard to tease out because of uncertainties around the seroprevalence data, and unavailability of granular case and fatality data.

All conclusions rely on the priors on parameters such as monthly attenuation and measured population seroprevalence. We find a 95% CI on true population seroprevalence in Delhi in early October of 43% to 59% – this would reflect prevalence a few weeks earlier, say, late September. The lower value corresponds to relatively slow spread, increasing naive IFR, rapidly increasing infection detection, and fairly slowly waning test sensitivity. The higher value corresponds to relatively rapid spread, relatively low naive IFR and rapidly waning test sensitivity to prior infection.

The August-September surge in cases probably reflected a real surge in infections, but of smaller magnitude than the earlier surge which occurred before July. This surge was not yet visible in seroprevalence data from early September, but became visible in seroprevalence data from October. The size of the surge in terms of cases has definitely been magnified by higher infection detection than in June.

Increasing infection detection could be partly a consequence of increased testing, but could also be a consequence of disease having shifted towards a population where testing is more likely to occur (more elderly/symptomatic patients, plus better access to resources), as seen in Mumbai. As an aside, test positivity does not tell a clear story. There was a big fall during July, but after that things are complicated by the use of rapid antigen tests. Without more granular data, it is hard to be certain what is leading to the improved case detection.

Delhi's low naive IFR remains a puzzle – even assuming levels of undercounting of around 50-60% (i.e., at least one in two deaths missed), it means that Delhi's true IFR seems to be lower than we are seeing in other cities such as Mumbai, but there is now strong evidence that it is rising. Low but rising IFR would be consistent with a scenario where the elderly were initially shielded from infection. This would be possible to check if fatality data by age were available.