Tracking CoVid19 India data
Methodology

Sarika Bhattacharyya, Suman Chakrabarty, Pinaki Chaudhuri,
Ujjwal Kumar Nandi, R. Shankar and Prateek Sharma

June 28, 2020

1 National and state wise data for reported numbers of infected, dead, recovered, active cases.

1. The confirmed cases are those who tested positive for the infection. The number of confirmed cases on any day, $C(t)$ is the total number of infected people who have been detected till that day.

   Note that the actual number of infected people would be more than those who have been detected.

2. The number recovered cases on any day, $R(t)$, is the total number of the confirmed cases who have recovered till that day.

3. The number of deceased cases on any day, $D(t)$, is the number of the confirmed cases who have unfortunately died.

4. The number of active cases on any day, $A(t)$, is the total number of confirmed cases who are still infected on that day. $A(t) = C(t) - R(t) - D(t)$.

2 Doubling time of reported cases of infection

If a quantity is growing as $C_0 2^{t/t_d}$, where C_0 is a constant, then it will double in every t_d days. The number of cases are growing in a similar fashion, but with a time dependent doubling time ($t_d(t)$). So we define the instantaneous doubling time of the number of confirmed cases as the inverse of the instantaneous growth rate (slope) of $\log_2(C(t))$ as follows:

$$\frac{1}{t_d(t)} \equiv \frac{d \log_2(C(t))}{dt} \tag{1}$$

In order to compute the time-dependent slope of $\log_2(C(t))$, we perform linear regression over a rolling time window of consecutive 5 days, that is slope at day N is computed over the time period from day $(N-2)$ to day $(N+2)$.

Based on the available state-wise data of reported cases, $C(t)$ is often not a smoothly increasing quantity. The number of daily reported cases can fluctuate significantly. There may not be any new reported case over a short time-span causing $C(t)$ to become constant during this period. Such temporal heterogeneity in real data would reflect into sudden spurious jumps in the temporally local rate (and doubling time) as defined above.

We would also like to stress that this is the doubling time of the *detected* cases and not the actual growth rate of the infection, since there may be many undetected cases. Thus apart from spurious features discussed above, the doubling time will decrease when a new cluster is detected and will increase again when it is contained. These oscillations are due to the containment process and do not reflect the progress of the disease.

3 National numbers and their link to states

The percentage contribution of a state to the all India value of a quantity is defined as follows. We denote the quantity by X (which is one of (C, R, D, A, T)), where T is the total number of samples tested. If X_n is the number for the n^{th} state, then the percentage contribution, x_n is,

$$x_n = \frac{1}{\sum_n X_n} X_n \times 100 \tag{2}$$

4 Test Positivity

The test positivity is the ratio of the number of tests with positive results to the total number of tests, namely $T^+(t)/T(t)$, where $T(t)$ is the total number of tests carried out till time t and $T^+(t)$ the number of positive results.

5 The case fatality ratio

The ratio of the number of deaths to the total number of actual cases, namely the probability that a person will die if she/he gets infected is called the Infection Fatality Ratio (IFR). It is difficult to to measure the IFR directly from the data since it is difficult to detect *all* the infected.

The ratio of the number of deaths to the total number of confirmed cases, namely the probability that a detected case will die, is called the Case Fatality Ratio (CFR). The CFR is also difficult to estimate when there are a large number of active cases since their fate is unkown. What can be measured are $CFR_1 \equiv D/C$ and $CFR_2 \equiv D/(R+D)$. These ratios will be equal to each other and the CFR at the end of the epidemic when almost all the confirmed cases have either recovered or are deceased.

During the epidemic, if the average time between detection of the case and death is smaller than the average time between detection of the case and declared recovery (as it is currently), then the actual CFR lies between the above two, $CFR_1 < CFR < CFR_2$. Note that this situation can change in the future.

The CFR for South Korea and Germany, which have very few active cases, are 2.3% and 4.8% respectively.

References

[1] https://ourworldindata.org/grapher/coronavirus-cfr