

## **Epidemic Landscape and Forecasting of SARS-CoV-2 in India**

Aravind Lathika Rajendrakumar MPH<sup>1\*</sup>, Anand Thakarakkattil Narayanan Nair MPH<sup>1\*</sup>, Charvi Nangia MSc<sup>1</sup>, Prabhal Kumar Chourasia MD<sup>2</sup>, Mehul Kumar Chourasia MPH<sup>1</sup>, Syed Mohammed Ghouse MTech<sup>1</sup>, Anu Sasidharan Nair MSW<sup>3</sup>, Arun B Nair MA<sup>4</sup>, Shaffi Fazaludeen Koya MBBS MPH<sup>5#</sup>

<sup>1</sup>Doctoral Researcher, School of Medicine, University of Dundee, UK

<sup>2</sup>Hospitalist, Mary Washington Hospital, Fredericksburg WV, USA

<sup>3</sup>Advisor, Health System Research India Initiative, Kerala, India

<sup>4</sup>Trustee, Health System Research India Initiative, Kerala, India

<sup>5</sup>Research Analyst and Doctoral Researcher, School of Public Health, Boston University, MA, USA

\*Contributed equally

# Corresponding Author

### BACKGROUND

India was one of the countries to institute strict measures for SARS-CoV-2 control in early phase. Since, then, the epidemic growth trajectory was slow before registering an explosion of cases due to local cluster transmissions.

### METHODS

We estimated growth rate and doubling time of SARS-CoV-2 for India and high burden states using crowd sourced time series data. Further, we also estimated Basic Reproductive Number (R<sub>0</sub>) and time dependent reproductive number (R<sub>t</sub>) using serial intervals from the data. We compared the R<sub>0</sub> estimated from five different methods and R<sub>0</sub> from SB was further used in analysis. We modified standard SIR models to SIRD model to accommodate deaths using R<sub>0</sub> with the Sequential Bayesian method (SBM) for simulation in SIRD models.

### RESULTS

On an average, 2.8 individuals were infected by an index case. The mean serial interval was 3.9 days. The R<sub>0</sub> estimated from different methods ranged from 1.43 to 1.85. The mean time to recovery was 14 ± 5.3 days. Daily epidemic growth rate of India was 0.16 [95%CI; 0.14, 0.17] with a doubling time of 4.30 days [95%CI; 3.96, 4.70]. From the SIRD model, it can be deduced that the peak of SARS-CoV-2 in India will be around mid-July to early August 2020 with around 12.5% of population likely to be infected at the peak time.

### CONCLUSIONS

The pattern of spread of SARS-CoV-2 in India is suggestive of community transmission. There is a need to increase fund for infectious disease research and epidemiologic studies. All the current gains may be reversed rapidly if air travel and social mixing resumes rapidly. For the time being, these must be resumed only in a phased manner, and should be back to normal levels only after we are prepared to deal with the disease with efficient tools like vaccine or a medicine.

**NOTE:** This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

## Background

SARS-CoV-2 originated in Hubei province of China and quickly spread to several countries including Japan and South Korea.<sup>1</sup> The source of the infection was later traced back to wet market in Wuhan, capital of Hubei.<sup>2</sup> A phylogenetic analysis of virus in Italy confirmed likely import of disease from China.<sup>3</sup> Rapid travelling modes such as air transport further accelerated the epidemic by introducing new cases in to more countries.<sup>4</sup> As of April 13 2020 there were 1,922,891 reported cases of SARS-CoV-2 globally along with a huge death toll almost touching 120,000.<sup>5</sup> There is increasing evidence that the virus may have jumped organisms to reach humans through pangolins or similar animals acting as intermediates.<sup>6</sup>

Emerging and re-emerging infections of this scale disrupt health system functioning and cause massive losses to economy. Mathematical modelling has been used widely to understand the spread of disease in populations.<sup>7</sup> One of the aims of the modelling is to estimate parameters that are critical to the spread of diseases such as basic reproductive number or  $R_0$  and incubation period.<sup>8</sup> This is quite useful in the scenario of early spread of disease to plan future strategies.

India was one of the countries to institute strict measures for SARS-CoV-2 control in early phase.<sup>9</sup> The first case of SARS-CoV-2 in India was reported on January 30 2020 from a student airlifted from Wuhan and at the same time another 800 suspected individuals kept under observation.<sup>10</sup> Since then, the epidemic growth trajectory was slow before registering an explosion of cases due to local cluster transmissions. In this paper, we describe the current trend of SARS-CoV-2 transmission in in India and estimate basic parameters such as basic reproductive number  $R_0$  and time reproductive number  $R_t$ , doubling time and future trend for India from real world datasets. Our findings will help to understand the effectiveness of government response and recalibrate suppression strategies for the epidemic.

## Methodology

We used crowd sourced time series data available from the internet to estimate country specific parameters for the epidemic.<sup>11</sup> As of now, this is the best available database for information on SARS-CoV-2 in India in public domain. The most reliable information in this database is regarding the daily reported incidence of COVID cases. Also, it contains aggregated information on total confirmed cases, total death and total recovery. We did not use variables with limited information. The data was scraped on to R software on 12<sup>th</sup> April 2020, cleaned and reshaped for the current analysis.<sup>12</sup> Package ggplot2 was mainly used to create figures along with features from base R.<sup>13</sup> We defined serial interval as the time difference in diagnosis of SARS-CoV-2 in infectee and infector.<sup>14</sup> We assumed serial interval and generation time to be same. Growth rate was computed from this incidence curve by fitting a log- linear model<sup>15</sup> and basic reproduction number ( $R_0$ ) was obtained from previously calculated serial interval. We also calculated time dependent reproductive number ( $R_t$ ) to show the change in infectivity over time. State-wise growth was measured using the slope from a linear model. To estimate an  $R_0$  for projection purposes, in addition to log linear, we also estimated  $R_0$  from other methods such as Exponential growth (EG), Maximum likelihood estimation (ML) and Sequential Bayesian Method (SB) and Estimation of time dependent

reproduction numbers (TD).<sup>16</sup> We selected Sequential Bayesian Method (SB) to account for stochasticity in incidence curves as the priors change with time and are drawn from posterior distribution from time dependent informative priors.<sup>16</sup>

Conceptually, susceptible individuals become infectious (i.e., move from the susceptible compartment to the infectious compartment), and then ultimately recover from the infection (i.e., move from the infectious compartment to the recovered compartment). The rates at which they move from one compartment to another depend on the proportion of the population in each of these compartments, as well as the transmission and recovery rates associated with the disease. We modified standard SIR models to accommodate an additional compartment for deaths by assuming transition probabilities, the susceptible-infectious-recovered/Death (SIRD) model. Our model can be represented by the following ordinary differential equation

$$\begin{aligned}\frac{dS}{dt} &= -\beta SI \\ \frac{dI}{dt} &= \beta SI - (1 - \alpha)\gamma I - \alpha\rho I \\ \frac{dR}{dt} &= (1 - \alpha)\gamma I \\ \frac{dD}{dt} &= \alpha\rho I\end{aligned}$$

where  $\beta$  is the disease transmission rate,  $\gamma$  is the recovery rate,  $\alpha$  is the case fatality rate and  $\rho$  rate at which death occurs.  $\beta$  is the product of contact rate and transmission probability. As these parameters could not be estimated directly from the data, we back calculated  $\beta$  from  $R_0$  which was computed previously using Sequential Bayesian method.  $\gamma$  was calculated from infectious period - (1/ infectious period). Case fatality rate ( $\alpha$ ) can be defined as deaths out of confirmed cases.  $\rho$  was estimated as inverse of time to death. We did not have  $\rho$  from our data and hence used estimated range of 2-8 weeks from the WHO report.<sup>17</sup> We used 5 weeks as the value of  $\rho$  for our study. The values of the parameters are as follows:  $R_0=1.85$ ,  $b=0.13$ ,  $\gamma=0.07$ ,  $\alpha=0.035$  and  $\rho=0.028$ .

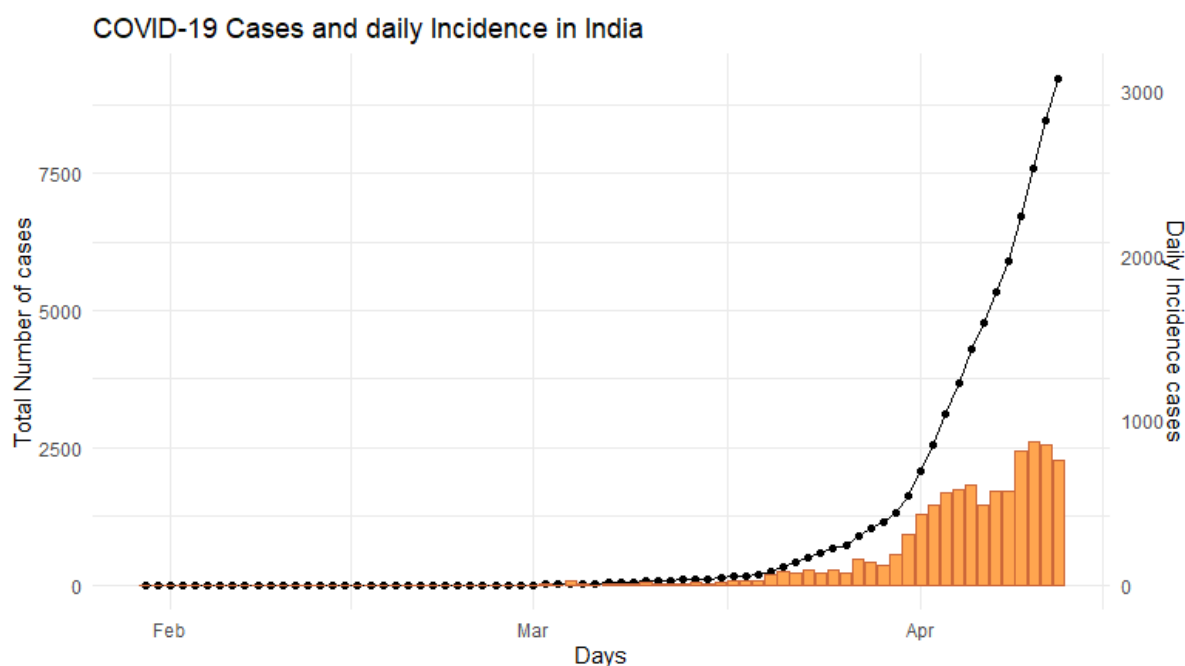
## Ethics Approval

We used anonymised data available in the public domain for this analysis.

## Results

The first case of SARS-CoV-2 was reported in Kerala state. Since then as of 12 April 2020, there are 9212 confirmed cases, 248 deaths and 1226 recovered in India from SARS-CoV-2. Maharashtra report the highest number of cases (1982) and north eastern states have the lowest reported cases. The trend of the epidemic is shown in Figure 1.

*Fig 1: Daily and cumulative incidence of SARS-COV-2 cases in India*



Daily growth rate of India fitted from a log incidence over time is 0.16 [95%CI; 0.14, 0.17]. The doubling time for the epidemic is 4.30 days [95%CI; 3.96, 4.70]. The growth rate and doubling time for the 5 states with highest burden are given in Table 1 and the rates for the states with available data are given in Supplementary Figure 3. Kerala, which had the highest number of cases have effectively contained the outbreak by prolonging the doubling time, 19.48 days [95% CI;10.35,164.60] and the present growth rate is 0.03 [95%CI; 0.004, 0.066].

Table 1: Growth rate and doubling curve from log-linear fit for 5 states with highest burden of SARS COV-2 in India

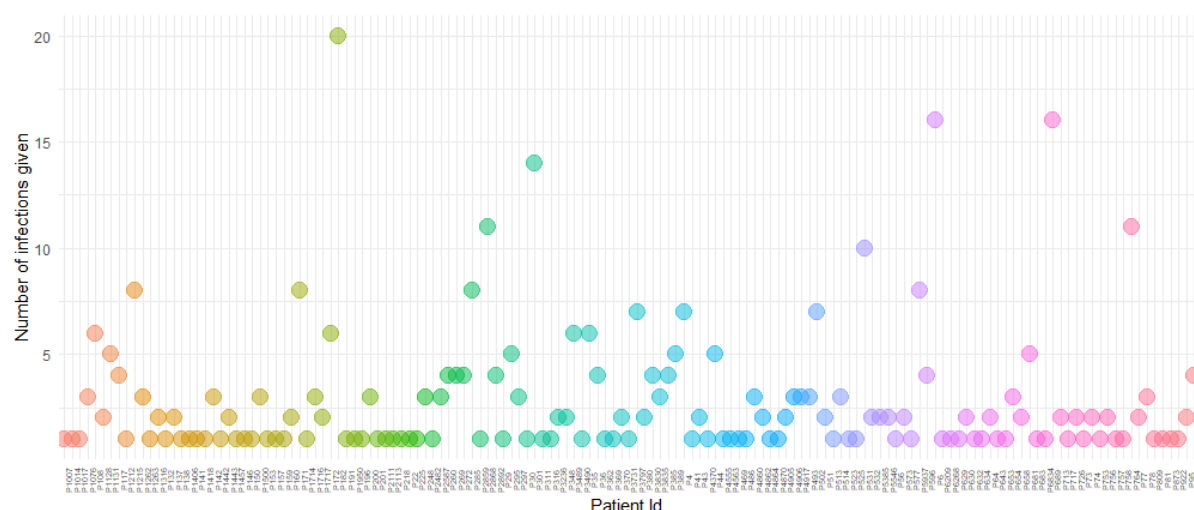
State	Growth rate (95% CI)	Doubling time in days (95% CI)
Tamil Nadu	0.173 (0.13,0.20)	4.00 (3.32,5.03)
Delhi	0.145 (0.12,0.17)	4.78 (4.05,5.83)
Maharashtra	0.144 (0.12, 0.16)	4.81 (4.23, 5.58)
Rajasthan	0.131 (0.11,0.15)	5.27(4.61,6.17)
Gujarat	0.119(0.07,0.17)	5.77(4.13,9.59)

The nature of the epidemic is changing as the transmission is now more prominent with in local communities as opposed to imported cases from individuals with travel history from China, Middle East and Europe. The type of transmission for most cases were not available. However, those individuals whose source of infection is yet to be determined can be assumed to be of local origin as travel history of SARS COV-2 patients is well tracked and documented in India (Supplementary Figure 2).

There were 145 index cases who were identified to have transmitted the disease. On an average, 2.8 individuals were infected by an index case and the range varied from 1-20

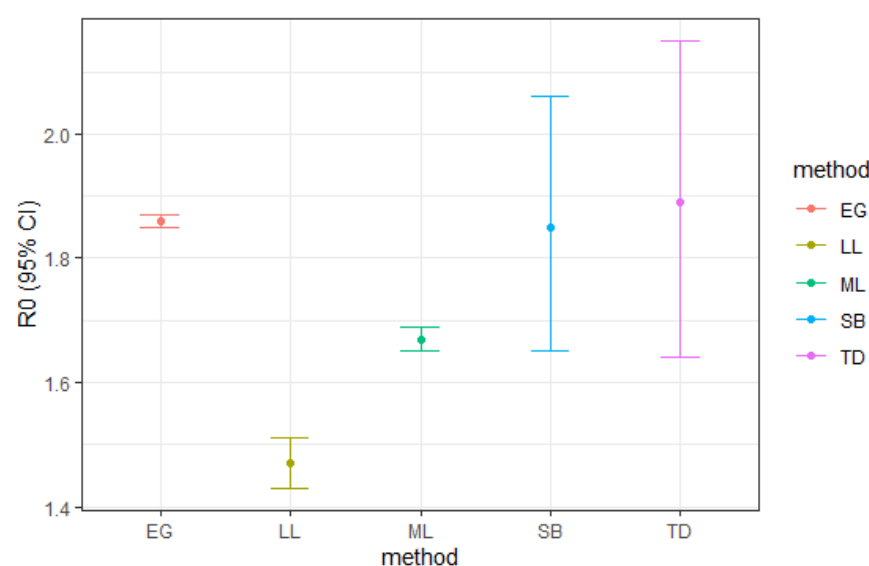
individuals (Figure2). We had information of 413 pairs of primary and secondary cases to generate serial intervals for computing Basic Reproductive Number or  $R_0$ . Serial interval ranged from 0-19 days as the primary and secondary cases reported to the health facility on the same date. The mean of serial interval was 3.9 days with a standard deviation of 2.85 computed from pair of 324 pairs with non-zero values

Figure 2. Number of Secondary cases resulting from contact with each index case



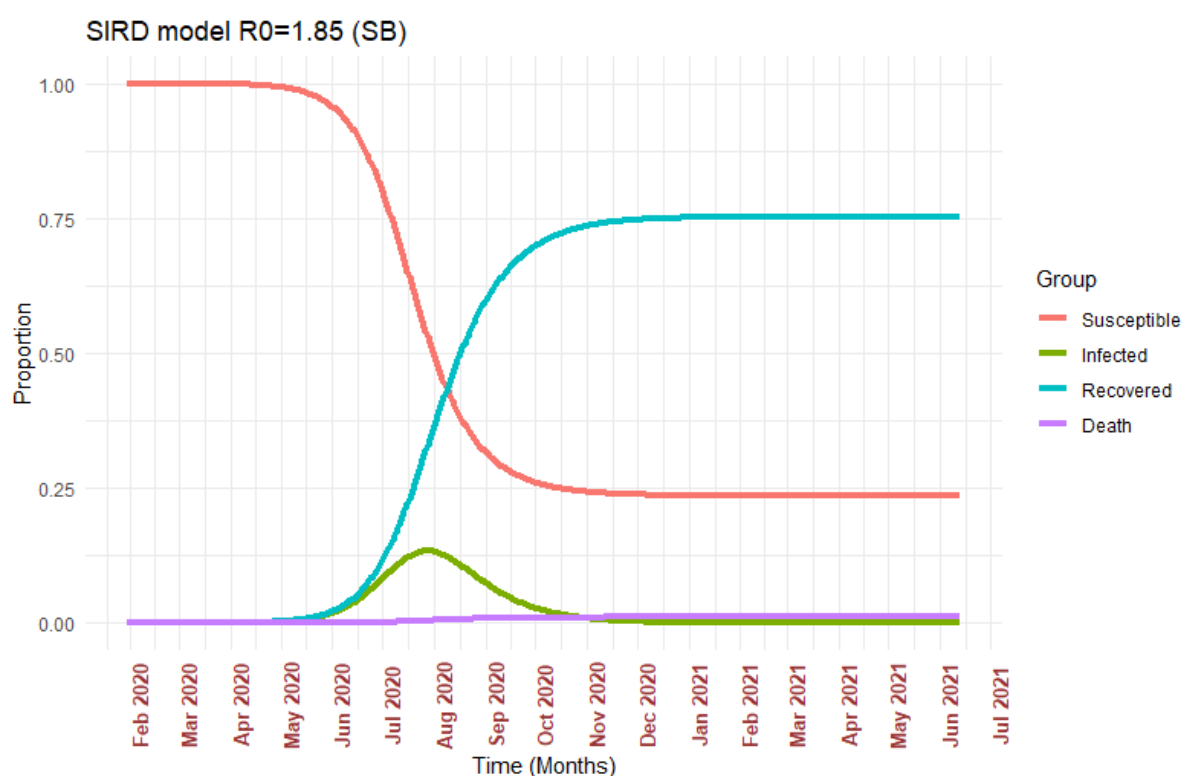
The lowest estimated  $R_0$  was from log linear model 1.47 [95% CI, 1.43, 1.51] and highest from time dependent model 1.89 [95% CI, 1.64-2.15] (Figure 3). The time varying reproductive number ( $R_t$ ) and SI distribution curve across the analysis time is provided in Supplementary Figure 1. The recovery time was available only for 133 individuals. The mean time of recovery was 14 days (SD 5.3) ranging from 5 to 25 days.

Figure 3: Basic reproductive number ( $R_0$ ) calculated from different estimation methods



From the compartmental models it can be deduced that the peak of SARS-CoV-2 in India will be around mid-July to early August 2020. Around 12.5% of the susceptible individuals are likely to be infected during the peak. The  $R_0$  estimated from different methods ranged from 1.43 in log- linear method to 1.85 in exponential growth methods. Since the estimates were similar between exponential, Time dependent and Sequential Bayesian methods we went forward with the Sequential Bayesian to create the SIRD model (Figure 4).

Figure 4. SIRD models based on  $R_0$  derived from Sequential Bayesian Method (SBM)



## Discussion

We estimated several parameters of SARS-CoV-2 infection in India, from crowd sourced data and used minimal set of assumptions in our forecasts. Models are useful in identifying the transmission dynamics if parameter inputs are based on real world data in early phase of epidemic. Social distancing and travel ban to reduce community transmission seem to have impact on infection transmission.

There is an urgent need to estimate setting specific parameters for future planning. For instance, we did not have information on death rate and recovery rate and used estimates from published studies in the SIRD model. We also could not compute incubation period as the database contained information only on linked cases and their admissions. Studies have shown that this could be approximately between 5-8 days.<sup>18,19</sup> Also, it is difficult to separate the serial interval time from generation time overlaps as the information of cases are only available from reporting to the health authorities. Studies have suggested that SARS-Cov-2 can spread before the symptomatic phase which precludes full understanding of the transmission pattern and underlying characteristics.

The  $R_0$  estimated across many studies may not reflect actual transmission potential of virus as these were measured in dynamic cohorts with some sort of mitigation measures. Estimates from studies among closed population such as in Diamond Princess and Wuhan may be more reliable in this regard.<sup>20</sup> Riou et al. estimated a reliable range for  $R_0$  which was between 1.4 and 3.8 by simulating Wuhan epidemic incidence trajectories in a cluster environment.<sup>21</sup> We estimated the range of  $R_0$  in India to be between 1.42 and 1.84 by comparing estimates from 5 different methods. It can be assumed that the real  $R_0$  in India is close to 1.85. The upper limit of uncertainty of  $R_0$  in our analysis exceeds 2 which may be indicative of spread of infection at local level and is comparable to European scenario.<sup>22</sup> This is much lower than  $R_0$  in most countries and may be indicative of the effectiveness of measures taken at policy level such as lock down, cancellation of flights, social distancing and sealing of identified hotspots.

India has done well to contain population level spread to a great extent and to keep low mortality rates. India has one of the lowest attack rate of 0.47 at this stage of the epidemic which is several times lower than developed countries.<sup>23</sup> However, the epidemic is on the rise in India and creates new challenges such as optimal use of resources including swabs and testing kits besides identifying the high-risk groups. A recent study by Prem et al suggests that mitigation measures may be helpful only in delaying the infection from reaching its peak.<sup>24</sup> Thus, it implies that the measures would likely prevent immediate flooding of hospitals preventing people from obtaining adequate care. The epidemic is yet to reach its peak in India which we expect to be around mid-July and early August. Our models suggest that 12.5% of the susceptible individuals are likely to be infected during the peak. This means that there is still around 2- 3 months to prepare for the worst phase of the epidemic.

Control measures of SARS-CoV-2 are less effective in populations with high density.<sup>20</sup> India has a huge population and potential of SARS-CoV-2 for sustained transmission is well established.<sup>21,25</sup> We relied solely on estimates from compartmental model which generates estimates from the infection parameters than depending on forecasts based on trend. Our findings are based on simulations and further measures at policy level can alter the trajectory of the disease. Lipsitch et al recommends a strategy combining epidemiological studies, testing and lab studies as more effective strategy in generating evidence for suppressing the epidemic.<sup>26</sup>

Our analysis has certain limitations. Firstly, there may be underreporting of cases and lack of validated data for research purposes available to the public. However, we compared the incidence from our dataset to the aggregated numbers elsewhere like European Disease Prevention and Control<sup>23</sup> and Worldometer<sup>5</sup> and the figures were similar. Another issue in estimating the parameters is the presence of super spreaders and asymptomatic cases. The estimated proportion of asymptomatic cases can be as high as 10% in the population.<sup>27</sup> The virus can be spread around during asymptomatic stage and delays from onset to reporting or treatment can be as high as 11 days.<sup>28</sup> Nonetheless, pooling estimates across several studies will aid in computing values that are closer to the true estimates.

In conclusion, our study provides information on India specific estimates of SARS-CoV-2 transmission parameters using real world data for the first time and shows that measures taken till date have been effective in reducing the spread of disease. However, the rising



incidence and pattern of spread is suggestive of community transmission and is likely to increase cases in the future. The availability of individual level data is critical to assess the effectiveness of ongoing measures and plan future strategies. There is a need to increasingly fund infectious disease research and epidemiologic studies and make that data available in the public domain. Future studies could focus on studying genetic variations in asymptomatic individuals and those index cases with higher disease transmission rates. There needs to be a renewed thinking of health systems particularly regarding emergency preparedness and optimal utilization of scarce resources. Epidemiologists need to be considered in decision making and developing disease control strategies. Furthermore, there needs to be concerted action at global level to contain the virus including joint research and developing vaccines and drugs. All the current gains may be reversed rapidly if air travel and social mixing resumes rapidly. For the time being, these must be resumed only in a phased manner, and should be back to normal levels only after we are prepared to deal with the disease with efficient tools like vaccine or a medicine.

### **Acknowledgement**

We acknowledge Dr Mike Loneragan, University of Dundee, UK for reviewing and providing critical inputs for the manuscripts. We thank the team managing COVID19 open data source which was used for this analysis.

### **Funding Statement**

None

### **Conflict of Interest**

Nothing to declare

### **References**

1. Munster VJ, Koopmans M, van Doremalen N, van Riel D, de Wit E. A Novel Coronavirus Emerging in China — Key Questions for Impact Assessment. *N Engl J Med*. 2020;382:692–4.
2. Shereen MA, Khan S, Kazmi A, Bashir N, Siddique R. COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. Vol. 24, *Journal of Advanced Research*. Elsevier B.V.; 2020. p. 91–8.
3. Giovanetti M, Benvenuto D, Angeletti S, Ciccozzi M. The first two cases of 2019-nCoV in Italy: Where they come from? *J Med Vir*. 2020.202;92:518–21.
4. Arima Y, Shimada T, Suzuki M, Suzuki T, Kobayashi Y, Tsuchihashi Y, et al. Severe Acute Respiratory Syndrome Coronavirus 2 Infection among Returnees to Japan from Wuhan, China, 2020. *Emerg Infect Dis*. 2020.
5. Worldometer. Coronavirus Cases [Internet]. Worldometer. 2020 [cited 2020 Apr 10]. p. 1–22. Available from: <https://www.worldometers.info/coronavirus/coronavirus-cases/#daily-cases>
6. Zhang C, Zheng W, Huang X, Bell EW, Zhou X, Zhang Y. Protein Structure and Sequence Reanalysis of 2019-nCoV Genome Refutes Snakes as Its Intermediate Host and the Unique

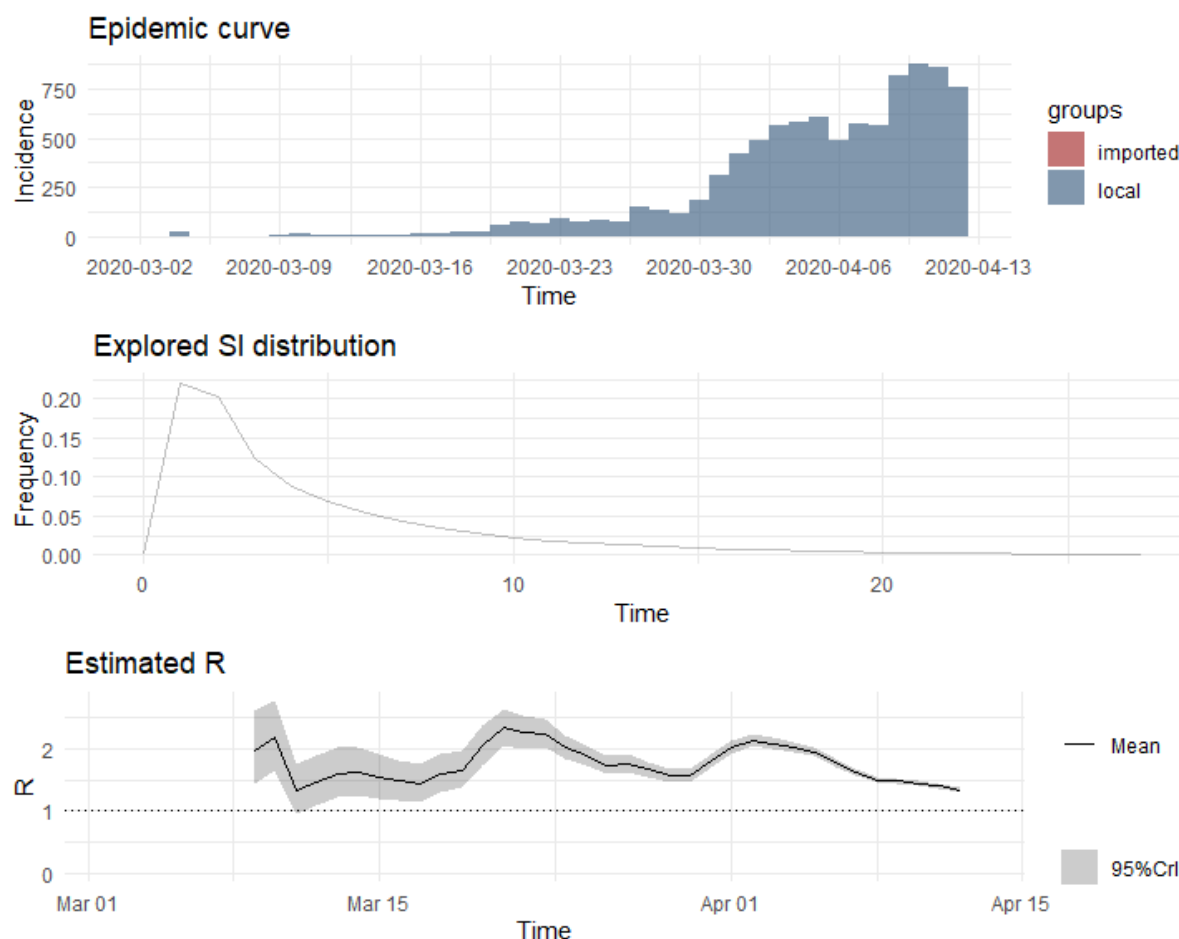


Similarity between Its Spike Protein Insertions and HIV-1. J Proteome Res. 2020 Mar 24;N Engl J Med 2020.

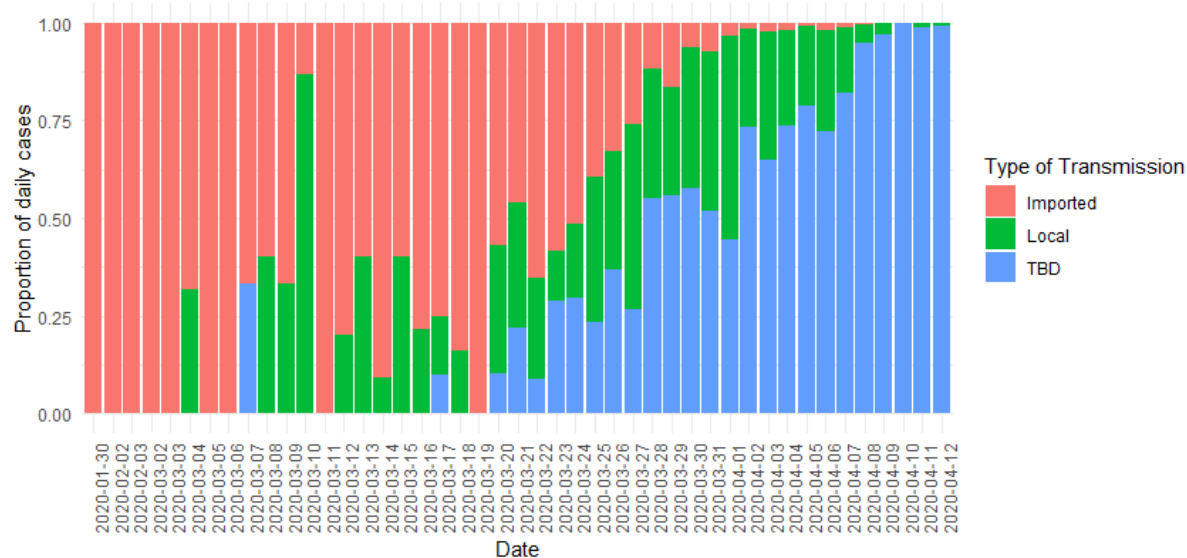
7. Siettos CI, Russo L. Mathematical modeling of infectious disease dynamics. Vol. 4, Virulence. Taylor and Francis Inc.; 2013. p. 295–306.
8. Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. Lancet Infect Dis. 2020;0(0).
9. Coronavirus Disease (COVID-19). Available from: <https://www.who.int/india/emergencies/novel-coronavirus-2019>
10. Coronavirus: India confirms first case, with 800 others under observation [Internet]. [cited 2020 Apr 10]. Available from: <https://www.telegraph.co.uk/global-health/science-and-disease/coronavirus-india-confirms-first-case-800-others-observation/>
11. COVID-19 Tracker | India. Available from: <https://www.covid19india.org/>
12. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria; 2014.
13. Wickham H. ggplot2: Elegant Graphics for Data Analysis Using the Grammar of Graphics. Springer-Verlag New York; 2016. Available from: <https://ggplot2.tidyverse.org>.
14. Moghadas S, Milwid R. Glossary of Terms for Infectious Disease Modelling A Proposal for Consistent Language.
15. Jombart T, Cori A, Kamvar ZN, SchumakerDirk. Package 'epitrix.'; Available from: <https://cran.r-project.org/web/packages/epitrix/epitrix.pdf>
16. Obadia T, Haneef R, Boëlle PY. The R0 package: A toolbox to estimate reproduction numbers for epidemic outbreaks. BMC Med Inform Decis Mak. 2012 Dec 18 [cited 2020 Apr 13];12(1):147. Available from: <http://bmcmmedinformdecismak.biomedcentral.com/articles/10.1186/1472-6947-12-147>
17. WHO. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). Feb 2020. <https://www.who.int/docs/defaultsource/coronaviruse/who-china-jointmission-on-covid-19-final-report.pdf> (accessed April 13, 2020).
18. Backer JA, Klinkenberg D, Wallinga J. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020. Eurosurveillance.2020;25(5).
19. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. Ann Intern Med. 2020.
20. Rocklöv J, Sjödin H. High population densities catalyze the spread of COVID-19. J Travel Med. 2020.
21. Riou J, Althaus CL. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. Euro Surveill [Internet]. 2020;25(4). Available from: [www.eurosurveillance.org](http://www.eurosurveillance.org)
22. Yuan J, Li M, Lv G, Lu ZK. Monitoring Transmissibility and Mortality of COVID-19 in Europe. Int J Infect Dis. 2020 Mar;
23. European Centre for Disease Prevention and Control. COVID-19. [cited 2020 Apr 12]. Available from: <https://qap.ecdc.europa.eu/public/extensions/COVID-19/COVID-19.html>

24. Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Heal.* 2020 Mar;0(0).
25. Endo A, Abbott S, Kucharski AJ, Funk S. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Res.* 2020;5:67.
26. Lipsitch M, Swerdlow DL, Finelli L. Defining the Epidemiology of Covid-19 — Studies Needed. *N Engl J Med.* 2020;382:1194–6.
27. Hellewell J, Abbott S, Phd J, Russell TW, Munday JD, Kucharski AJ, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Glob Heal.* 2020;8:e488–96.
28. Li DQ, Guan X, Wu P, Wang X, Cowling DB, Yang B, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *N Engl J Med* 2020. 2020;382:1199–207.

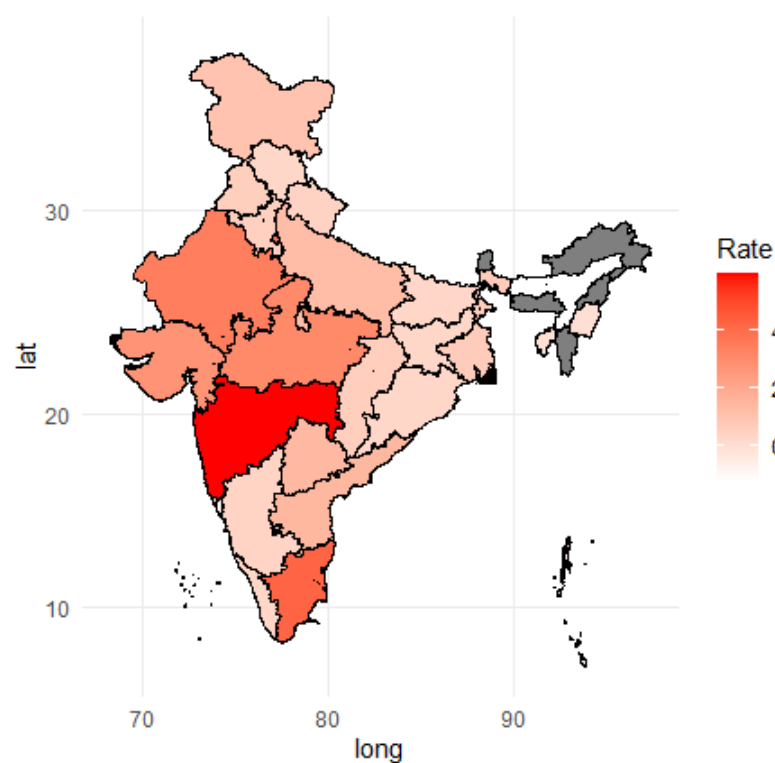
Supplementary Figure 1. Epidemic curve, SI distribution and time varying reproductive number ( $R_t$ )



Supplementary Figure 2: *Daily incidence and type of SARS COV-2 transmission in India*



Supplementary Figure 3: Growth rate in each state in India fitted with linear model



Grey colour indicates non availability of data