

Characterization of the second wave of COVID-19 in India

Rajesh Ranjan^{1,*}, Aryan Sharma² and Mahendra K. Verma²

¹Department of Aerospace Engineering, Indian Institute of Technology, Kanpur 208 016, India

²Department of Physics, Indian Institute of Technology, Kanpur 208 016, India

The second wave of COVID-19, which began in India around 11 February 2021, has hit the country hard with daily cases reaching nearly triple the first peak value as on 19 April 2021. The epidemic evolution in India is complex due to regional inhomogeneities and the spread of several coronavirus mutants. In this study, we characterize the virus spread in the ongoing second wave in India and its states until 19 April 2021, and also examine the dynamic evolution of the epidemic from the beginning of the outbreak. Variations in the effective reproduction number (R_t) are taken as quantifiable measures of virus transmissibility. R_t value for every state, including those with large rural populations, is greater than the self-sustaining threshold of 1. An exponential fit on recent data also shows that the infection rate is much higher than in the first wave. Subsequently, characteristics of COVID-19 spread are analysed region-wise, by estimating test positivity rates (TPRs) and case fatality rates (CFRs). Very high TPR values for several states present an alarming situation. CFR values are lower than those in the first wave, but are recently showing signs of increase as the healthcare system is being over-stretched with the surge in infections. Preliminary estimates with a classical epidemiological model suggest that the peak for the second wave could occur around mid-May 2021, with daily count exceeding 0.4 million. The study strongly suggests that an effective administrative intervention is needed to arrest the rapid growth of the epidemic.

Keywords: Coronavirus, COVID-19, epidemic evolution, reproduction number, second wave.

MORE than one year since coronavirus disease 2019 (COVID-19) was declared a pandemic on 11 March 2020, by the World Health Organization (WHO), the deadly SARS-CoV-2 virus continues to disrupt public life across the world. Although the lockdown norms are relatively relaxed in most countries, social life is still far from normal. Recently, multiple vaccines developed by Oxford–AstraZeneca (Covishield/Vaxzevria), Pfizer–BioNTech (Comirnaty), Moderna, Johnson & Johnson (Janssen), Bharat Biotech (Covaxin), Gamaleya Research Institute

of Epidemiology and Microbiology (Sputnik V), etc. have been approved in several countries and are given on priority basis to susceptible populations and those with co-morbidities. However, the production and distribution of vaccines at a massive scale to cover a very large population remain a formidable challenge. Meanwhile, in order to arrest the spread of the virus during the vaccination drive, intervention measures such as wearing masks, social distancing guidelines, partial lockdowns and restricted store hours are still in place in most countries.

While nations are taking extensive measures to accelerate the vaccination drive in order to control the pandemic at the earliest, a public health challenge has appeared due to mutations of the SARS-CoV-2 virus which make it highly contagious. For example, the SARS-CoV-2 lineage B.1.1.7, which was first detected in the United Kingdom in November 2020, is estimated to be 40–80% more transmissible than the wild-type SARS-CoV-2 (refs 1, 2). Similarly, strains detected in South Africa (B.1.351), Brazil (P.1), and India (B.1.617) are also significantly more contagious than the variants in early 2020 (ref. 3). There is no clear evidence on the severity of the new mutations¹; however the challenge is to prepare for health response, especially when the number of infections is exceedingly large.

In order to understand the impact of spread of the mutated coronavirus, we show in Figure 1 *a* and *b* the temporal variations of daily COVID-19 cases and deaths in

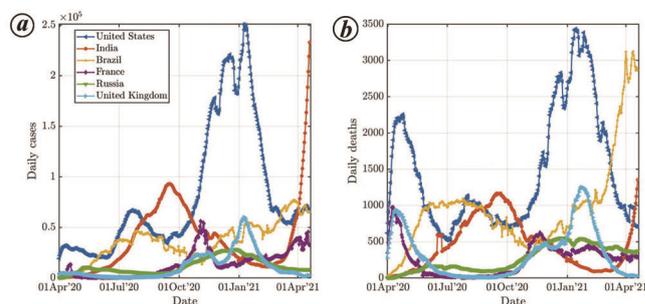


Figure 1. Comparison of COVID-19 spread in India with five other severely affected countries. Smoothened data between 1 April 2020 and 19 April 2021 as retrieved from ourworldindata.org are used for the plots. Variations in (a) daily number of new infections and (b) new deaths are shown. Curves clearly show the first and second waves for the US, UK and India. Peaks in second wave for all these countries are higher than those in first wave.

*For correspondence. (e-mail: rajeshr@iitk.ac.in)

the top six countries respectively. These countries are most affected in terms of the maximum number of cumulative cases as of 19 April 2021. Among the countries shown, the United States has maximum cumulative cases with a total of about 32.5 million, followed by India that has 15.6 million cases. If we compare the recent epidemic growth in these countries as shown in the figure, the daily number of infections in India is several times more than that in the US. The plot also shows the peaks of the first and second waves in different countries where they were observed. The US, for example, had the first peak in mid-July 2020, following which the cases subsided. However, the infection counts started increasing again in October and a much larger peak was observed in the second wave in December 2020, with daily cases of up to 0.25 million. In the UK, the first and second peaks were separated by only a couple of months, with the second wave largely attributed to a more infectious mutant⁴. In both countries, the peak of the second wave was much higher than in the first wave. For Russia, the infection curve is relatively shallow, with the number of daily cases never crossing 30,000. In the curve for Brazil, there are large fluctuations which may be due to insufficient testing; however, the number of deaths per infection (typically called case fatality rate (CFR)) is very large, as can be inferred from Figure 1 *b*. The curve for the death cases typically follows the infections. Note, however, that the CFR is very high at the beginning of the pandemic (around April 2020) for the US and European countries⁵.

For India, the first and second waves are separated by about five months. The peak of the first wave was in September 2020, with daily cases of around 0.1 million. The daily cases decreased until mid-February, after which it exhibited a sharp increase. The end of the first wave was likely a result of a combination of factors – effective implementation of Government interventions, increase in awareness, and, most importantly, the experience gained by medical professionals in treating the disease over the

initial months. On 19 April 2021, the number of new cases was about 0.3 million, which is already triple the first peak value. The sudden surge in the number of cases after a relatively long ‘cooling’ time is baffling, although it may be attributed to a highly infectious double-mutant variant of SARS-CoV-2 (B.1.617 lineage), complacent behaviour of the population and relaxation of interventions⁶. The number of daily deaths is also rising recently, but CFR is low compared to the first wave; this aspect will be further discussed later. Note that the study on B.1.617 mutant is limited and there is no clear evidence on whether the mutant virus is less severe than its predecessor.

The present study focuses on the ongoing second wave in India and therefore, we look at the regional distribution of COVID-19 spread. Figure 2 *a* and *b* shows the daily number of reported infections in 16 key states in India, in linear and log scales respectively. In Figure 2 *a*, we note that almost all the states show a surge in the number of cases since 11 February 2021. Further, the slopes of the growth curves are very steep in the second wave compared to the first wave. The daily number of cases in Maharashtra, which also leads in the daily as well as cumulative infections, increased from 652 on 11 February 2021 to about 63,000 in two months (as on 11 April 2021). This rapid growth is also observed in the other states (Figure 2 *b*), albeit the number of daily cases is fewer than that in Maharashtra as on 19 April 2021. The growth curve in the second wave can be further divided into relatively slow and fast growth phases, as shown by green and orange shadows in Figure 2. In the first region, until the first week of March, all the states except Maharashtra exhibit a slow increase in the number of cases. However, in the second region, most of the states show a sudden spurt in the number of infections propelling India’s total daily count to about 0.3 million.

At the outset, the second wave in India looks much more precarious than the first wave as on 19 April 2021. Although India is the ‘fastest country’ in the world to administer COVID-19 vaccine doses⁷, the fraction of vaccinated people is relatively small due to the very large population of the country. Further, India is yet to administer vaccines to those less than 45 years of age (except susceptible populations like doctors, frontline workers, etc.), many of whom are getting severely infected by the mutated coronavirus. Lastly, unlike the first wave, the current spread of the virus mutants has reached remote locations (discussed later), where healthcare services are not sufficiently adequate. The purpose of this study is to characterize the ongoing second wave so as to create awareness about the present grim situation, and also sensitize the public as well as policymakers about the critical need for social distancing and other interventions to arrest further growth. We use several parameters such as testing rates, death rates as well as vaccination data in different states to describe the ongoing situation. Further, we also use available data to make informed projections

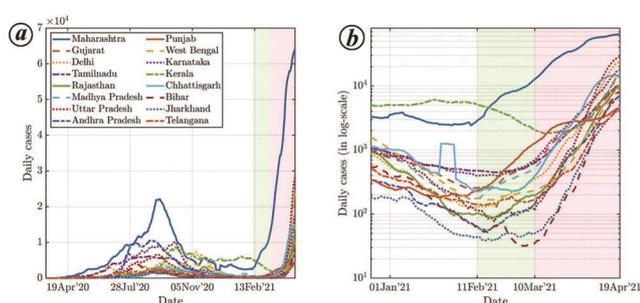


Figure 2. Comparison of COVID-19 spread in most impacted 16 states of India as on 19 April 2021. Time-series infection data are taken from covid19india.org and smoothed by taking seven-day averaging. *a*, Variations in daily infections from the beginning of the pandemic. *b*, Zoomed view focusing on the second wave. Green and orange shadows exhibit relatively slow and subsequent fast growth of daily cases in the second wave. The top three states in terms of new infections are Maharashtra, Uttar Pradesh and Delhi.

of the epidemic growth based on a widely used epidemiological model.

Materials and methods

The present study is based on COVID-19 time-series data from the beginning of the pandemic to 19 April 2021. COVID-19 data for countries around the globe and Indian states were taken from repositories maintained by *ourworldindata* (OWID⁸) and *covid19india.org* (COVID19INDIA⁹) respectively. OWID compiles data from the European Centre for Disease Prevention and Control (ECDC). COVID19INDIA curates data from several sources, including the Indian Government data, state bulletins, and official handles.

In order to identify the first and second waves, we studied the effective reproduction number (R_t)¹⁰ as a marker for the decrease or surge in infections. R_t provides real-time feedback on the spread of pandemic as a value >1 indicates an increase in infection; thus the goal is to implement social interventions to bring down R_t below 1 and close to 0 as much as possible. Time-varying R_t can be calculated using time-series data of the infections and generation time distribution^{11,12}. We used the approach developed by Thompson *et al.*¹³ for the estimation of effective reproduction numbers using the R-package EpiEstim. A MATLAB implementation of this package, developed by Batista¹⁴, has been used in this study. The generation time distribution requires serial interval as an input, which indicates the time between the onset of symptoms of a primary case and the onset of symptoms of secondary cases. Several studies have estimated serial interval for COVID-19 using data primarily from the first wave^{15–18}. It is expected that if the virus characteristics (infectiousness, incubation period) change significantly due to mutations, the serial interval will also change. However, a thorough study of the recent virus mutations is not available in the literature. Therefore, we used the available COVID-19 parameters as estimated by Nishiura *et al.*¹⁸. The mean serial interval was taken as 4.7 days (95% confidence interval (CI): 3.7, 6.0) with the standard deviation (SD) as 2.9 days (95% CI: 1.9, 4.9).

The available dataset for the second wave has also been used for predictions employing the popular compartmental susceptible–infected–recovered (SIR) model¹⁹. The incidence data between 27 February and 17 April 2021 have been utilized to obtain best-fit parameters used in the model. Note that the pandemic is still in the exponential-growth phase, as further expanded in the next section. Therefore, there is large uncertainty in the estimation of the peak of the daily cases, as well as the duration of the second wave. Details about the computation of underlying parameters and implementation of SIR model are available in the literature^{20,21} and are briefly summarized in Appendix 1.

Results

We first analysed the available time-series data to quantify the virus transmissibility in India as well as in the different states. Figure 3 shows the variations in daily confirmed cases as well as effective reproduction numbers in India. The effective reproduction number trend broadly follows the infection rate variation. In the first wave, the R_t value decreased from about 1.37 (95% CI: 1.25–1.52) on 17 April 2020 to 1.09 (95% CI: 1.07–1.11) on 10 September 2020. It went below the self-sustaining threshold of 1 for the first time on 23 September 2020 and remained so for the next five months, except for a minor flare up on 29 November 2020 for a couple of days. After this relatively long quiet interval, R_t started rising on 19 February 2021, which can be considered as an indicator of the arrival date of the second wave in India. R_t has been increasing since then with minor fluctuations; as on 19 April 2021, it has reached approximately 1.37 (95% CI: 1.28–1.47).

Next, we describe the regional variations of effective reproduction numbers. These variations are expected to be dissimilar commensurate with trends of new infections as shown in Figure 2. Figure 4 displays these variations for the nine most impacted states as of 19 April 2021, while Table 1 lists the most recent values of R_t for all the states. As evident in the figure, the incidence curves for these states show that the current daily cases are already higher than that of the first peak, except for Andhra Pradesh. The R_t curve crossed the threshold first in Maharashtra and about a week later in the other states.

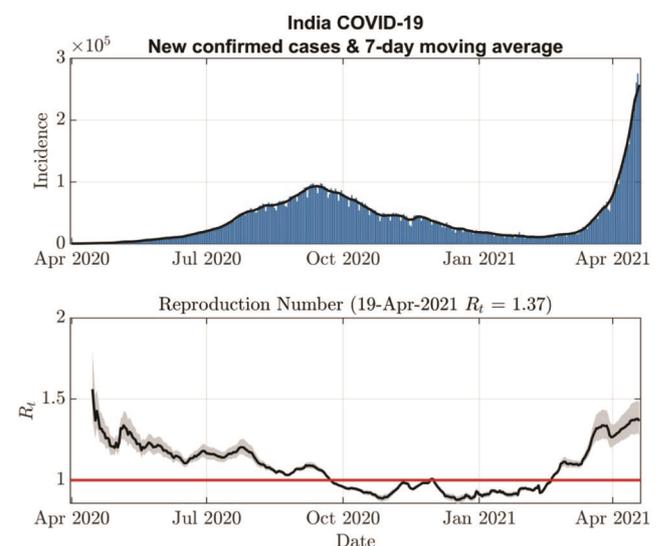


Figure 3. Temporal variations of the effective reproduction number (R_t) along with incidence (daily infections) data. The 95% CI of R_t variation is also shown in grey shadow. R_t was oscillating about 1.2 during the growth phase of the first wave and went below 1.0 for the first time around the end of September 2020, marking the decline of the pandemic. R_t started rising again in February 2021 and had a value of 1.37 on 19 April 2021.

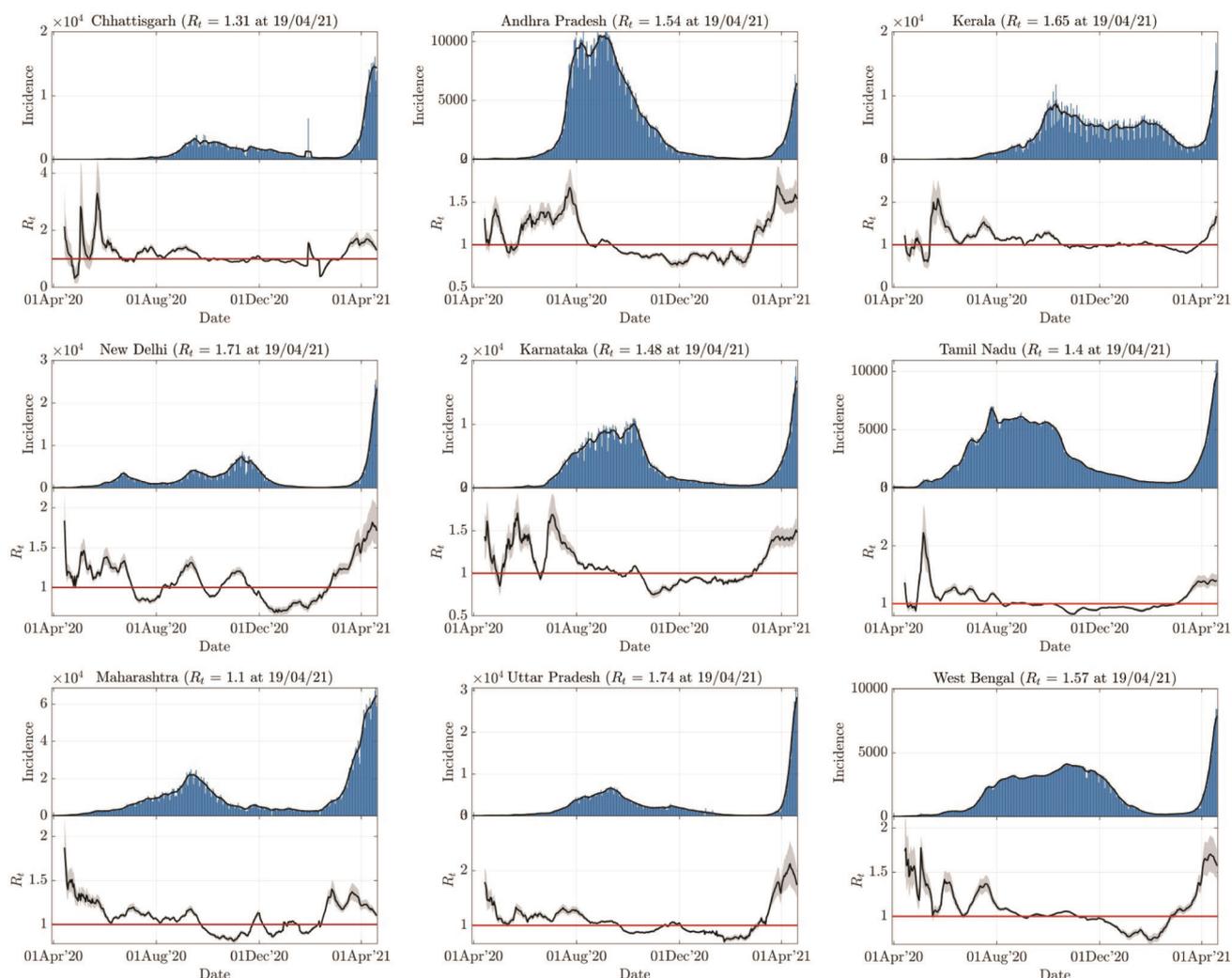


Figure 4. Temporal variations of the effective reproduction number (R_t) along with incidence (daily infections) data for the most impacted states. Signs of the second wave were first noted in Maharashtra, where R_t crossed the threshold of 1.0 for the first time this year. All states had $R_t > 1.0$ as on 19 April 2021.

However, very recently there has been a slight decline in the R_t value in Maharashtra (and also in Chhattisgarh and Punjab, which were highly impacted in the beginning of this wave). The most populated state, i.e. Uttar Pradesh, which was one of the least impacted during the first wave, is currently in the high growth phase with R_t value of 1.74 (Table 1). Similarly, West Bengal and Bihar also have high R_t values (Table 1). This is of concern as there is a large rural population in these states and the health-care system is inadequate for such a scale of the pandemic. For reference, Uttar Pradesh, Bihar and West Bengal together constitute 32% of India's population in which about 78%, 88% and 68% people respectively, live in rural areas in these states.

For a better perspective, we compare the R_t value of India with those of other most impacted countries as on 19 April 2021 (Table 2). Except for the UK, where the pandemic is on a decline ($R_t \sim 0.90$), the other four coun-

tries have R_t close to the threshold value. The fraction of the vaccinated population is also listed in the table. Among all countries, the US and UK have the highest vaccinations per 100 persons; these countries are possibly least susceptible to another wave provided the vaccines can give protection against change in virus phenotype due to mutations²². India, despite running one of the largest vaccination efforts in the world⁷ with two made-in-India vaccines – Covaxin and Covishield – has a relatively low share of vaccinated people per capita due to a very large population (Table 2). Further, the population covered so far is above 45 years of age, except for healthcare professionals and frontline workers. The vaccination drive needs to be enhanced to include the younger population, say from 18 years of age, as several reports suggest that the mutated virus may have higher virulence towards the younger population than that in the first wave²³. As of 19 April 2021, India has approved other vaccines to be

Table 1. Regional characteristics of the epidemic spread in Indian states and Union Territories as on 19 April 2021^a

Region	Cumulative		Daily		R_t	Vaccinations (%) ^b	Population (in crores)
	CFR (%)	TPR (%)	CFR (%)	TPR (%)			
Andhra Pradesh	0.77	6.16	0.45	15.79	1.53	9.1	5.2
Arunachal Pradesh	0.33	4.05	0.00	5.59	2.52	12.0	0.15
Assam	0.51	2.86	0.51	2.09	1.88	4.9	3.4
Bihar	0.54	1.31	0.55	8.98	1.84	4.9	12
Chandigarh	1.21	9.45	0.65	20.36	1.21	12.7	0.118
Chhattisgarh	1.09	8.45	1.26	28.42	1.20	17.4	2.9
Delhi	1.41	5.38	1.01	26.12	1.72	13.4	2
Goa	1.32	11.46	1.81	34.48	1.43	16.5	0.154
Gujarat	1.32	2.58	1.03	7.20	1.48	15.6	6.8
Himachal Pradesh	1.52	5.61	0.77	30.04	1.36	18.2	0.73
Haryana	0.95	5.28	0.48	21.20	1.50	10.9	2.9
Jammu and Kashmir	1.39	2.18	0.40	4.45	1.35	13.0	1.3
Jharkhand	0.90	2.59	1.07	11.27	1.54	7.6	3.7
Karnataka	1.15	4.96	0.92	12.81	1.48	11.4	6.6
Kerala	0.40	8.73	0.15	15.63	1.65	17.2	3.5
Madhya Pradesh	1.10	5.91	0.61	25.32	1.53	9.2	8.2
Maharashtra	1.56	16.19	0.60	26.59	1.10	10.4	12.2
Manipur	1.27	5.01	1.85	5.58	1.93	5.1	0.31
Meghalaya	1.03	3.43	0.91	8.17	1.77	5.4	0.322
Mizoram	0.24	1.79	0.00	7.41	1.77	13.9	0.119
Nagaland	0.75	8.94	0.00	8.55	1.31	6.9	0.215
Odisha	0.54	3.87	0.09	12.29	1.80	11.5	4.4
Puducherry	1.48	6.55	0.88	15.85	1.40	11.2	0.15
Punjab	2.62	4.59	1.80	14.59	1.15	8.4	3
Rajasthan	0.75	5.47	0.44	29.78	1.63	14.5	7.7
Sikkim	2.03	8.01	0.00	–	1.74	25.2	0.066
Tamil Nadu	1.31	4.71	0.40	9.80	1.40	6.3	7.6
Telangana	0.52	3.01	0.35	4.82	1.48	8.5	3.7
Tripura	1.15	5.11	0.00	2.30	1.51	22.8	0.399
Uttar Pradesh	1.14	2.29	0.59	14.07	1.75	4.8	22.5
Uttarakhand	1.50	3.75	1.11	7.12	1.59	14.8	1.1
West Bengal	1.59	6.79	0.45	20.01	1.58	9.3	9.7

^aNumbers in bold indicate a greater probability of virus spread. ^bAt least one dose has been administered.

Table 2. Characteristics of the epidemic spread in six most affected countries as on 19 April 2021

Region	Cumulative		Daily		R_t	Vac. (%)	Population (in crores)
	CFR (%)	TPR (%)	CFR (%)	TPR (%)			
USA	1.79	7.54	0.94	3.18	1.01	63.3	33.2
India	1.18	5.72	0.68	16.15	1.37	9.5	139.1
Brazil	2.68	–	4.64	–	1.02	15.9	21.4
France	1.91	7.29	6.68	1.59	0.93	25.1	6.5
Russia	2.25	3.74	4.03	1.43	1.00	11.1	14.6
UK	2.90	3.05	0.13	0.17	0.90	63.5	6.8

procured and administered, which could further accelerate the vaccination effort.

To further characterize the second wave, we employ the following well-known ratios

$$\text{Test positivity rate (TPR)} = \frac{\text{Total infections}}{\text{Total tests}}, \quad (1)$$

$$\text{Case fatality rate (CFR)} = \frac{\text{Total deaths}}{\text{Total infections}}. \quad (2)$$

These ratios can be calculated based on the cumulative or daily data. While the estimates based on cumulative data are more smooth, the daily ratios reflect sudden changes more prominently. Therefore, we employed both

definitions to illustrate different aspects, although the daily ratios were obtained after a seven-day moving average to remove large fluctuations due to reporting delays and other uncertainties.

TPR typically indicates whether the number of tests is sufficient to contain the spread of the virus by isolating and quarantining the positive cases. Figure 5 *a* illustrates the temporal variations of these parameters based on cumulative data for both the first and second waves. India shows an increase in cumulative TPR during the acceleration phase of the first wave and it starts declining from August 2020. Recently, the cumulative TPR curve shows an upward trend commensurate with the spurt of cases, with a TPR value of 5.72% as on 19 April 2021. This increase in TPR in the second wave is better reflected in the daily estimates as shown in Figure 5 *b*. The daily TPR curve shows a sharp spike in its value since late March 2021. As on 19 April 2021, the daily TPR value for India is about 16%. WHO recommends that this TPR value should be less than 5% for at least two weeks so that the transmission can be brought under control. A sudden increase in daily TPR to a very high value indicates an alarming situation and necessitates ramping up of daily testing capacity.

Further, as shown in Table 1, there are large variations among different regions in India. Ten states have alarming daily TPR of more than 20%. Maharashtra, which shows a slight decline in R_t value recently, has a very high TPR of 26.6% indicating very high transmission. Therefore, the actual number of infections is likely to be higher than that being reported due to limited diagnostics capacity. States which have a high R_t value, as well as high daily TPR ($>10\%$), are also at significant risk.

Next, we report CFR for India in Figures 5 *a* and *b* based on cumulative and daily data respectively. Both the CFR curves show downward trends with time. The cumulative CFR curve decreases from 3.5% in mid-April 2020 to 1.2% in mid-April 2021, with minor fluctuations.

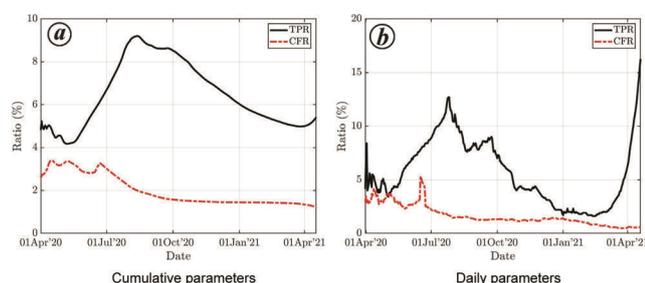


Figure 5. Temporal variations of two characterizing parameters – test positivity rate (TPR) and case fatality rate (CFR). These ratios were computed using (a) cumulative and (b) daily data. There is a sudden increase in TPR in the second wave, which presents a worrying situation. CFR in the second wave is lower than that in the first wave. However, (b) shows signs of its increase very recently. This may be due to reduced healthcare access to the patients due to a drastic increase in the number of infections.

Interestingly, although the second wave shows the virus to be more infectious, the decline in the CFR curve suggests a silver lining of a relatively less fatal mutant. However, considering an exponential increase of cases at a very high rate, it is expected that soon the healthcare facilities will be fully throttled resulting in the unavailability of hospital beds and ventilators to those in critical need. This may result in an increase in CFR. Even otherwise, in terms of absolute numbers, the daily deaths are already higher than the level of peak values in the first wave (Figure 1 *b*). Table 1 lists state-wise data of cumulative and daily CFRs, with Punjab having the highest value and Kerala the least among the severely affected states.

Finally, we describe the vaccination data in states as listed in Table 1. Among the most impacted states with a sizeable population, Kerala and Chhattisgarh have highest vaccination per 100 people. Uttar Pradesh and Bihar, which have very high R_t , also have the lowest level of vaccination per capita. This further suggests the need for strong interventions in these states while vaccination capacity is increased simultaneously with prioritized allocations based on social contacts²⁴.

After the characterization of the second wave, we employed mathematical and epidemiological models to understand the dynamics and provide actionable insights. As discussed earlier (see Figure 2), the daily infections data in the second wave can be classified into relatively initial slow and subsequent fast growth phases. Although the reason for the two different growth curves is not exactly known, it may be related to daily testing capacity and other regional factors. Based on trends, we used two exponential curves to fit the incidence (daily cases) data. Further, we compared this fit with that of the first wave, starting from a date when the number of infections was similar to that in the second wave. Figure 6 exhibits these exponential fits, while Table 3 lists the coefficients as well as statistical fitness data. The first exponential fit of the second wave has a lower exponent than the first wave (compare b_1 and b_2 in Table 3), although it has a relatively low value of the coefficient of determination (adjusted $R^2 = 0.62$). However, the exponent of the second fit on more recent data in the second wave is more than twice that of the first wave, which explains the rapid growth of the pandemic. Fits for both the first wave and rapid second wave are statistically significant with adjusted R^2 greater than 0.95.

We then employed the dynamical SIR model for an estimate of the progression of the pandemic²⁵. This model takes both the incidence and recovery data to estimate the peak as well as eventual decline of the pandemic. It is a compartmental model, where the total population is divided into three segments – susceptible, infected and recovered (sometimes also termed as ‘removed’), as described further in Appendix 1. Using the parameters infection (β) and recovery (γ) rates, the model plots the

Table 3. Exponential regression model for initial growth in the first and second waves: $y = a \exp(bt)$

Wave	Coefficients	Estimate	95% CI	SE	<i>t</i> -statistic	<i>P</i> -value	Adjusted R^2
First	a_1	1.22×10^4	1.12×10^4	1.33×10^4	531	23.0	9.79×10^{-32}
	b_1	0.0278	0.0261	0.0296	0.00089	31.0	4.36×10^{-39}
Second (1)	a_2	1.11×10^4	0.99×10^4	1.23×10^4	578	19.2	1.72×10^{-16}
	b_2	0.0194	0.0131	0.0256	0.00305	6.35	1.18×10^{-06}
Second (2)	a_3	2.65×10^3	2.23×10^3	3.07×10^3	208	12.75	1.13×10^{-15}
	b_3	0.0689	0.0662	0.0715	0.00129	53.3	8.77×10^{-39}

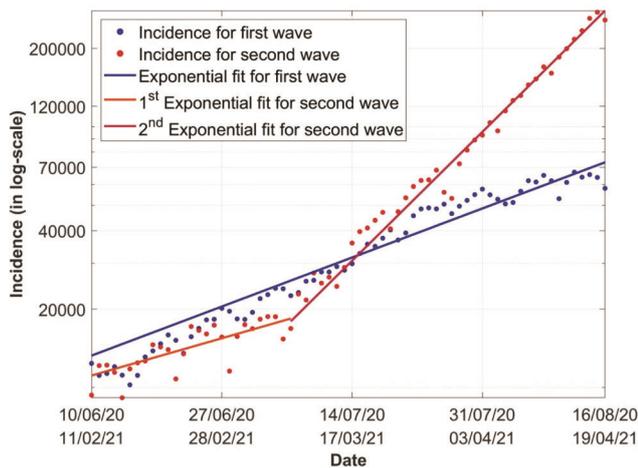


Figure 6. Comparison between first and second waves in equal intervals of time starting from the dates when the number of infections was almost similar. In the first wave, data between 10 June 2020 and 16 August 2020 were considered, which shows an exponential fit. In the second wave, two different exponential fits were used corresponding to green and orange regions shown in Figure 2. The exponential growth factor in the second wave fit on the recent data is more than double that of the first wave fit.

progression of disease with time in terms of these three segments using simple governing equations. This model has been widely used for the prediction of COVID-19 in several countries²⁶. More sophisticated variations of this model that further subdivide the population into ‘exposed’ (SEIR²⁷), ‘asymptomatic’ (SAIR²⁸) and ‘exposed–quarantined–deceased–insusceptible’ (SEIQRD²⁹) are also used for predictions of the first wave in India. A review of the predictive models can be found in Kotwal *et al.*³⁰. Although SIR model is the simplest epidemiological model, several studies suggest that it adequately represents COVID-19 dynamics for different geographical regions and can be used for predictions^{31,32}. Further, the uncertainty range for the SIR model is lower than more sophisticated models, as the two governing parameters (β and γ) for the former can be more easily determined using the limited publicly available infection and recovery data, while assumptions are used in the latter for several other parameters (such as the number of asymptomatic/exposed patients, intervention compliance, regional variations) in the absence of reliable data.

Note that forecasting the peak (both the time as well as number) with any model is difficult when the pandemic is

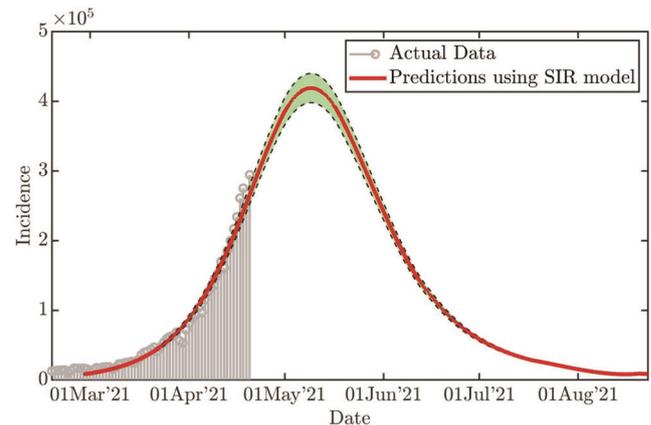


Figure 7. Predictive estimates of the COVID-19 epidemic in India using classical susceptible–infected–recovered (SIR model). Upper and lower bounds with 5% uncertainty are also shown. Data between 27 February 2021 and 17 April 2021 are taken for fitting of parameters. The predictions suggest the second wave to peak around mid-May with infections exceeding 0.42 million. Note that uncertainty in infection data due to high positivity rate and regional variations in the spread of the virus as well as interventions are not taken into account for predictions. Consideration of these factors could alter the estimates.

in the exponential growth phase. This is because there is large uncertainty in the estimation of underlying parameters for these models using the infection data only from the exponential growth phase. Pandemics usually follow the powerlaw beyond the exponential growth during which estimation of the peak is more accurate^{33,34}. Nonetheless, the projections provide ballpark estimates and therefore help policymakers to prepare for the future and plan necessary actions. Figure 7 shows the predictions using the SIR model with parameters obtained from the data between 27 February and 17 April 2021. As shown, the peak of the second wave is estimated in the middle of May, with the peak number of daily cases of more than 0.4 million. Furthermore, even with a low daily CFR of 0.68% as of 19 April 2021, India will have about 3000 deaths per day at its peak. This number is expected to increase as access to healthcare becomes more and more limited with the surge in active infections.

An important point to note here is that as mentioned above, the high TPR of India suggests that the actual number of infections could be higher than that reported. Since the model parameters are estimated using the reported number of infections, the projection of the peak

case count is still conservative. Therefore, the results presented above should be taken as an optimistic scenario, provided no new interventions are introduced. The high TPR also suggests that India may see a near plateau (of about 0.38–0.42 million daily cases until TPR goes below 5%) in May instead of a sharp peak, as testing capacity limits the number of daily cases. Thus, while the growth rate may be small in this period, active cases will continue to increase, thus throttling the healthcare infrastructure. Note that the proposed model does not include several factors such as asymptomatic cases, partial or full lockdowns, vaccination data, regional heterogeneities, etc. These factors may be considered for more realistic modelling of the second wave once more data are available. We plan to employ universal curve³⁵ and refined epidemiological models^{26,36} for improved prediction of the second wave in India.

Conclusion

The second COVID-19 wave in India, which began in the second week of February 2021, presents a grim situation with the number of daily cases reaching around 0.3 million on 19 April 2021. The data suggest that at present the virus is much more infectious than in the first wave, but the number of daily deaths per infection is lower. However, with an inordinate increase in the number of cases and over-stretched healthcare system, the daily death count may increase substantially. The effective reproduction number (R_t) has been estimated for India as well as for the Indian states. At present, almost every state shows $R_t > 1$, suggesting that the second wave has spread everywhere, including rural areas which were largely untouched during the first wave. This includes populous states like Uttar Pradesh, Bihar and West Bengal: each has R_t value greater than 1.58. With rural areas being impacted, it may be necessary to take aggressive lockdown measures to arrest further spread while sufficient vaccines become available. The SIR model based on available data suggests the peak of the epidemic to occur in the middle of May 2021, with approximate daily infections exceeding 0.4 million.

In summary, using the available infection data, we have analysed the second COVID-19 wave in India. We observed that the epidemic is creating unprecedented havoc in the population. We hope that appropriate administrative intervention, aggressive vaccination drive and people’s participation will help flatten the curve earlier than the grim forecast of the epidemic.

Appendix 1. Description and implementation of the SIR model.

In the SIR model, the distributions for the number of susceptible S , the number of infectious I , and the number of

recovered or deceased (or immune) individuals R is given as¹⁹

$$\dot{S} = -\beta SI, \tag{A1}$$

$$\dot{I} = \beta SI - \gamma I, \tag{A2}$$

$$\dot{R} = \gamma I, \tag{A3}$$

where β is the transmission rate and γ is the average recovery rate. From eqs (A1)–(A3), $\dot{S} + \dot{I} + \dot{R} = 0$. Therefore $S + I + R = N$ is the population that remains constant throughout the pandemic and could be considered as the susceptible population before outbreak ($t < t_0$), since $I = R = 0$ for ($t < t_0$). Parameters β and γ govern the course of the pandemic and for an ongoing pandemic, they are determined by available infection and recovery data as follows. Combining eqs (A1) and (A2) yields

$$\frac{dI}{dS} = \frac{\nu}{S} - 1, \quad \nu = \frac{\gamma}{\beta}. \tag{A4}$$

If we assume that at the beginning of pandemic $I(t_0) = 1$, $R(t_0) = 0$ and therefore $S(t_0) = N - 1$, the above equation can be integrated to yield

$$I = \nu \ln S - S + N - \nu \ln(N - 1). \tag{A5}$$

Similarly, eqs (A2) and (A3) can be combined to obtain the equation for total number of cases, $C = I + R$, as

$$\dot{C} = \beta SI. \tag{A6}$$

Using eqs (A5) and (A6), the governing equation for C using the SIR model will be

$$\dot{C} = \beta(N - C)[\nu \ln S - S + N - \nu \ln(N - 1)]. \tag{A7}$$

Parameters β and γ are thus obtained by minimizing the difference between the reported and predicted number of cases for a given time interval. MATLAB function ‘fminsearch’ was used to obtain the minimum of the least squares problem:

$$\min \|C_a - C(\beta, \gamma)\|_2. \tag{A8}$$

Here C_a is the reported number of daily cases, while C is the predicted number of cases using eq. (A7).

1. Davies, N. G. *et al.*, Estimated transmissibility and impact of SARS-CoV-2 lineage b. 1.1. 7 in England. *Science*, 2021, **372**, 6538.
2. Volz, E. *et al.*, Assessing transmissibility of SARS-CoV-2 lineage b. 1.1. 7 in England. *Nature*, 2021, **593**, 266–269.

3. Chen, J., Wang, R., Wang, M. and Wei, G.-W., Mutations strengthened SARS-CoV-2 infectivity. *J. Mol. Biol.*, 2020, **432**(19), 5212–5226.
4. Leung, K., Shum, M. H., Leung, G. M., Lam, T. T. and Wu, J. T., Early transmissibility assessment of the n501y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. *Eurosurveillance*, 2021, **26**(1), 2002106.
5. Borghesi, A., Golemi, S., Carapella, N., Zigliani, A., Farina, D. and Maroldi, R., Lombardy, northern Italy: COVID-19 second wave less severe and deadly than the first? A preliminary investigation. *Inf. Dis.*, 2021, **53**, 370–375.
6. Xu, S. and Li, Y., Beware of the second wave of COVID-19. *The Lancet*, 2020, **395**(10233), 1321–1322.
7. Bagcchi, S., The world's largest COVID-19 vaccination campaign. *The Lancet. Inf. Dis.*, 2021, **21**(3), 323.
8. OWID; <https://ourworldindata.org/> (accessed on 19 April 2021).
9. COVID19INDIA; <https://www.covid19india.org/> (accessed on 19 April 2021).
10. Fraser, C., Riley, S., Anderson, R. M. and Ferguson, N. M., Factors that make an infectious disease outbreak controllable. *Proc. Natl. Acad. Sci.*, 2004, **101**(16), 6146–6151.
11. Cori, A., Ferguson, N. M., Fraser, C. and Cauchemez, S., A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am. J. Epidemiol.*, 2013, **178**(9), 1505–1512.
12. Wallinga, J. and Teunis, P., Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am. J. Epidemiol.*, 2004, **160**(6), 509–516.
13. Thompson, R. *et al.*, Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics*, 2019, **29**, 100356.
14. Batista, M., estimate_r, Matlab central file exchange, 2021 (accessed on 19 March 2021).
15. Du, Z., Xu, X., Wu, Y., Wang, L., Cowling, B. J. and Meyers, L. A., Serial interval of COVID-19 among publicly reported confirmed cases. *Emerging Inf. Dis.*, 2020, **26**(6), 1341.
16. Knight, J. and Mishra, S., Estimating effective reproduction number using generation time versus serial interval, with application to COVID-19 in the greater Toronto area, Canada. *Inf. Dis. Modelling*, 2020, **5**, 889–896.
17. Rai, B., Shukla, A. and Dwivedi, L. K., Estimates of serial interval for COVID-19: A systematic review and meta-analysis. *Clin. Epidemiol. Global Health*, 2020, **9**, 157–161.
18. Nishiura, H., Linton, N. M. and Akhmetzhanov, A. R., Serial interval of novel coronavirus (COVID-19) infections. *Int. J. Inf. Dis.*, 2020, **93**, 284–286.
19. Hethcote, H. W., The mathematics of infectious diseases. *SIAM Rev.*, 2000, **42**(4), 599–653.
20. Batista, M., Estimation of the final size of the COVID-19 epidemic. *medRxiv*, doi, 2020, 10, 2020.02, 16–20023606.
21. Ranjan, R., Predictions for COVID-19 outbreak in India using epidemiological models. *medRxiv*, 2020.
22. Grubaugh, N. D., Hanage, W. P. and Rasmussen, A. L., Making sense of mutation: what D614G means for the COVID-19 pandemic remains unclear. *Cell*, 2020, **182**(4), 794–795.
23. Volz, E. *et al.*, Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. *Cell*, 2021, **184**(1), 64–75.
24. Chen, J. *et al.*, Prioritizing allocation of COVID-19 vaccines based on social contacts increases vaccination effectiveness. *medRxiv*, 2021.
25. Anderson, R. M. and May, R. M., *Infectious Diseases of Humans: Dynamics and Control*, Oxford University Press, 1992.
26. Ranjan, R., Temporal dynamics of COVID-19 outbreak and future projections: A data-driven approach. *Trans. Indian Natl. Acad. Eng.*, 2020, **5**, 109–115.
27. Gopal, R., Chandrasekar, V. and Lakshmanan, M., Dynamical modelling and analysis of COVID-19 in India. *Curr. Sci.*, 2021, **120**(8), 1342–1349.
28. Agrawal, M., Kanitkar, M. and Vidyasagar, M., Modelling the spread of SARS-CoV-2 pandemic – impact of lockdowns and interventions. *Indian J. Med. Res.*, 2021, **153**(1), 175–181.
29. Ranjan, R., COVID-19 spread in India: Dynamics, modeling, and future projections. *J. Indian Statist. Assoc.*, 2020, **58**(2), 47–65.
30. Kotwal, A., Yadav, A. K., Yadav, J., Kotwal, J. and Khune, S., Predictive models of COVID-19 in India: a rapid review. *Med. J. Armed Forces India*, 2020, **76**(4), 377–386.
31. Postnikov, E. B., Estimation of COVID-19 dynamics ‘on a back-of-envelope’: does the simplest SIR model provide quantitative parameters and predictions? *Chaos, Solitons & Fractals*, 2020, **135**, 109841.
32. Cooper, I., Mondal, A. and Antonopoulos, C. G., A SIR model assumption for the spread of COVID-19 in different communities. *Chaos, Solitons & Fractals*, 2020, **139**, 110057.
33. Asad, A., Srivastava, S. and Verma, M. K., Evolution of COVID-19 pandemic in India. *Trans. Indian Natl. Acad. Eng.*, 2020, **5**(4), 711–718.
34. Verma, M. K., Asad, A. and Chatterjee, S., COVID-19 pandemic: Power law spread and flattening of the curve. *Trans. Indian Natl. Acad. Eng.*, 2020, **5**, 103–108.
35. Sharma, A., Sapkal, S. and Verma, M. K., Universal epidemic curve for COVID-19 and its usage for forecasting, *Trans. Indian Natl. Acad. Eng.*, 2021, **6**, 405–413.
36. Shayak, B., Sharma, M. M., Rand, R. H., Singh, A. K. and Misra, A., Transmission dynamics of COVID-19 and impact on public health policy. *medRxiv*, 2020.

ACKNOWLEDGEMENT. This study is based partly on work supported by a SERB MATRICS project (SERB/F/847/2020-2021).

Received 28 April 2021; revised accepted 6 May 2021

doi: 10.18520/cs/v121/i1/85-93