Statistics and Applications {ISSN 2452-7395(online)}
Volume 18, No. 1, 2020 (New Series), pp 333–350

Epidemiological Dynamics of the COVID-19 Pandemic in India: An Interim Assessment

Sitabhra Sinha
The Institute of Mathematical Sciences
CIT Campus, Taramani, Chennai 600113, India

Received: 20 July 2020; Revised: 29 July 2020; Accepted: 30 July 2020

Abstract

We have analyzed the time-series for the number of active cases of COVID-19 pandemic in India, as well as, in other countries around the world using a variety of statistical fitting procedures. We obtain robust estimates of the exponential growth rate for the number of active cases, which is then used for calculating the reproduction number of the epidemic. We estimate the basic reproduction number of COVID-19 epidemic in India to be $R_0 \sim 1.82 \pm 0.02$, a value that lies at the lower end of the spectrum of values of different regions around the world where there have been major outbreaks of the disease. We have also investigated the change in the effective reproduction number over time, particularly following the introduction of unprecedented non-pharmaceutical interventions such as the stay-at-home order (lockdown) imposed over the entire country from 24 March 2020, and continued at varying levels of strictness, and with regional variations, up to the present (July). We observe that the reproduction number showed a large reduction within a couple of weeks of the imposition of lockdown, suggesting that this measure played a role (along with others such as compliance with physical distancing rules in public and use of masks) in reducing the rate of spreading of the contagion, although it was unable to break the chain of infection. We also note that there is considerable regional variation across India in the dynamics of the epidemic, with different regions registering rise and fall in the growth rate of the disease at different times.

Key words: COVID-19; Corona virus; Reproduction number; Epidemiological dynamics; pandemic.

AMS Subject Classifications: 92D30, 62M10

1. Introduction

The rapid spread of Coronavirus disease 2019 (COVID-19), that results from infection with the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pathogen, has in the few months following its initial identification in December 2019 (Wang, *et al.*, 2020) not only brought to mind recent pandemics such as the 2009 swine flu pandemic that is believed

Corresponding Author: Sitabhra Sinha
Email: sitabhra@imsc.res.in

to have affected around $11\% - 18\%$ of the world population (see Kelly, *et al.*, 2011), but also raised the specter of it eventually growing to rival the 1918-20 "Spanish Flu" pandemic. Believed to be one of the deadliest pandemics in recent history, the 1918-20 pandemic caused between 15 million (Spreeuwenberg, *et al.*, 2018) and 50 million deaths (Johnson and Mueller, 2002), with a third of these occurring in British-ruled India (Reyes, *et al.*, 2018). Indeed, in scale, COVID-19 has already surpassed the other two coronavirus epidemics of recent times, *viz.*, the 2002-04 SARS outbreak and the 2012-14 MERS outbreak. Compared to the 8439 cases that were reported for SARS worldwide (WHO, 2003), and the even smaller 2500+ cases of MERS from 2012 till date (WHO, 2020a), there have already been, as of 15 July 2020, over 13 million confirmed cases of COVID-19 and 574,466 deaths (WHO, 2020b) spread across more than 200 countries and territories.

In the absence of any possibility that a vaccine for COVID-19 will be available in the immediate future, the epidemic is also proving to be a testing ground for extreme non-pharmaceutical interventions that different countries have implemented in order to contain the spread of the disease. The principal among these are the stay-at-home orders (colloquially referred to as lockdowns) imposed on the populace of several countries around the world by their governments [see Wikipedia (2020) for a list] after its apparent success in containing the initial outbreak of the disease in Hubei province of China [Lai, *et al.* (2020)]. Such unprecedented measures have provoked controversy, not least in India where an initial 21-day lockdown was initiated on 24 March 2020 with the stated purpose of bringing the epidemic to an end by breaking the chain of infection through enforced distancing [see India Today (2020)]. While the physical isolation of individuals through such measures is very likely to restrict the likelihood of transmission of the pathogen from infected to susceptible individuals, there are high social and economic costs accompanying such a measure. Lockdowns also affect different sections of the population asymmetrically, and can result in aggravating existing inequalities in society - making it unsuitable for prolonged use. Thus, following the initial period of nationwide lockdown in India, it is now primarily being applied in a more restrictive manner at specific locations where the number of active cases is increasing in a particularly alarming rate, to ensure that the medical infrastructure is not overwhelmed by the rising number of infected individuals who need to be hospitalized.

To gauge the efficacy of such non-pharmaceutical interventions, it is imperative to understand the epidemiological dynamics of this novel infectious disease - particularly, as it manifests in diverse manners at different locations. In this paper, the transmissibility of COVID-19 has been investigated with special focus on India, but also considering many other locations around the world where there have been major outbreaks. For this purpose, we have estimated at each location the reproduction number of the disease (the basic reproduction number R_0 and effective reproduction number R are defined in the next section), which measures how rapidly the number of active cases of the disease changes over time (*active cases* refers to the individuals who are infected with the disease at a given time and who can potentially infect others by passing the pathogen to non-infected individuals, *e.g.*, via contact). As can be easily explained using the mathematical theory of epidemics, the reproduction number has to be greater than 1 for an epidemic to occur, and the larger the number, the faster the disease will spread. Although several studies have already appeared

that compute the basic/effective reproduction numbers for different locations at different times, as the numerical value of the reproduction number typically depends to an extent on the exact model used to estimate it from data, most of these numbers that appear in different studies cannot be compared to each other. Thus, in order to compare the epidemiological dynamics of COVID-19 across different geographical regions and temporal phases of the epidemic, the corresponding values of R_0 and R need to be calculated in a consistent manner across both space and time. With this aim in view, in this paper the reproduction numbers for the epidemic have been calculated not only for India but also for several countries where there have been major outbreaks, as well as, for different regions within India. This allows us to obtain an understanding of the spatio-temporal diversity in the spreading dynamics of the disease within India, apart from that between India and other parts of the world. The paper is organized as follows. In Section 2, we briefly describe the sources of the data used for the analysis and the method of calculating the key epidemiological parameters R_0 and R. In Section 3, the international situation is discussed with analysis of the epidemiological dynamics for the world as a whole and that of selected countries. In Section 4, we focus on India with data for the entire country, as well as for individual states and districts. We conclude with a discussion on the limitations of the study and its implications in Section 5.

2. Data Sources and Methods

Data aggregated for the world as a whole and at the level of each country affected by the epidemic was obtained from CSSE (2020), an online data repository on GitHub of COVID-19 cases worldwide that is operated by the Johns Hopkins University Center for Systems Science and Engineering. Information about the cumulative number of confirmed cases, deaths and recovered cases are updated daily, beginning from 22 January 2020. The data is collated from a large number of sources, such as various national government health departments, as well as, from the World Health Organization (WHO), the US and European CDCs and aggregating sites such as *WorldoMeters*. Disaggregated data for India was obtained from COVID19-India (2020), a crowdsourced database of COVID-19 cases. Its volunteers collate information from health bulletins issued periodically by various governmental organizations, as well as other sources, and compile the obtained numbers to create district-level and state-level daily time-series for confirmed cases, recovered cases, deaths, active cases and number of individuals tested for the disease. At the state-level, data is available from 14 March 2020, while for the bulk of the districts the time-series information is obtainable from 21 April 2020 onwards.

We consider the time-evolution of the number of active cases, *i.e.*, the number of individuals who remain infected with the virus on a particular date, which is obtained by subtracting the cumulative number of deaths and recoveries announced up to that date from the cumulative number of confirmed cases till then. During the initial phase of an epidemic, the number of infected individuals is expected to increase exponentially with time, as is the case for any multiplicative process (such as, a chain reaction) where the value adopted by a variable at each instant is obtained by multiplying the value at the immediately preceding instant by a constant factor. This is easy to see from the mathematical models of epidemiological dynamics that stem from the pioneering work of Kermack and McKendrick (1927) [for an example of how such theoretical modeling can accurately describe the empirical data

from influenza epidemics see Spicer and Lawrence (1984)]. In the most basic setting, one can divide the entire population comprising N individuals into three compartments, corresponding to those who are *susceptible* to contracting the disease ($\mathcal{S}$), those who are at present *infected* ($\mathcal{I}$), and those who have *recovered* or are *removed* by death ($\mathcal{R}$). Neglecting any demographic changes during the period that one is considering, the time-evolution of the number of individuals in each compartment can be described by the system of differential equations:

$$\frac{d\mathcal{S}}{dt} = -\beta\mathcal{S}\mathcal{I}, \frac{d\mathcal{I}}{dt} = \beta\mathcal{S}\mathcal{I} - \gamma\mathcal{I}, \frac{d\mathcal{R}}{dt} = \gamma\mathcal{I}, \tag{1}$$

where the parameters β is the rate of infection transmission through contact between an infected and a susceptible individual, while γ is the recovery rate ($= \tau^{-1}$, *i.e.*, the reciprocal of the time duration during which a person is free to pass on the infection to others). It is easy to see that as the total population is conserved ($\mathcal{S} + \mathcal{I} + \mathcal{R} = N$), only two of the equations are independent. Furthermore, at the earliest stage of the epidemic, we can assume the susceptible population size to be effectively equal to the total population N, and thus a constant. Thus, we are left with a single differential equation that describes the evolution of $\mathcal{I}$. Normalizing the variables by the total population size N and solving the equation, we see that the fraction of infected individuals in the population $i(= \mathcal{I}/N)$ will evolve from its initial value i_0 as $i(t) = i_0 \exp(\{N\beta - \gamma\}t) = i_0 \exp(\{[R_0 - 1]/\tau\}t)$. Here the parameter $R_0 = N\beta/\gamma$ is the key epidemiological parameter *basic reproduction number*, which is defined as the average number of secondary infections that results from a primary infection at the earliest stage of the epidemic, i.e., before a significant fraction of the population has been exposed to the disease. Note that the expression of R_0 remains unchanged even if we augment this basic model with an additional compartment $\mathcal{E}$ for the subpopulation of *exposed* individuals who have been infected but are not infectious, taking into account the latent or pre-infectious period after an infection.

Thus, R_0 can be estimated from the empirical time-series of the number of infected individuals (*i.e.*, the active cases) by accurately fitting it to an exponential growth curve, *viz.*, $i(t) \sim \exp(\lambda t)$, and obtaining the most reliable estimate for λ. Using the equivalence $\lambda = (R_0 - 1)/\tau$, and equating τ with the *generation time*, i.e., the mean interval between a person getting infected by another individual (the "infector") and the time at which the "infector" was infected, R_0 can be calculated from the data. We have used a generation time of 5.2 days that was estimated by Ganyani, *et al.* (2020) from the Singapore cluster of cases. The fitting procedure is carried out using a nonlinear least squares approach implemented by the function *fit* in *MATLAB* R2009b software (Mathworks (2009)). We have also obtained the 95% confidence bounds using the function *confint* which does the calculation through QR decomposition of the Jacobian. To assess the quality of fitting, we calculate the correlation coefficient r between the logarithm of the number of active cases and time, as well as the p-value indicating the measure of significance. We only use those estimates of λ for calculating R_0 for which $r > 0.99$ and $p \leq 0.002$. We have earlier shown in Jesan, *et al.* (2011) that using the above fitting procedure yields values of R_0 that are consistent with those calculated using alternative methods, such as bootstrapping. To aid fitting when the data exhibits large

fluctuations we have performed smoothing using a 3-day moving average.

Once the epidemic has had time to penetrate substantially into the population and/or containment measures put in place have had a discernible effect, it is no longer possible to view the process as a contagion freely infecting every contact of an infected person. At this stage, we speak of the growth rate in terms of the time-dependent *effective reproduction number*, R. As in the case of R_0, R is also defined as the mean number of infections arising from a single infected individual (with the difference that now we can no longer assume the population to be almost entirely susceptible to the disease). Thus, it can be estimated using the above technique provided that the susceptible population does not decrease perceptibly over the period in which the estimation is being done. Also, similar methods can be applied to calculate the reproduction number at different spatial scales. Obviously the smaller the area being considered the smaller is the total population, so that it becomes more likely that there will be discernible changes in the susceptible population as the epidemic progresses and consequently one has to be more careful in using the above procedure.

3. COVID-19: International Scenario

The outbreak of a novel disease (to be named COVID-19 eventually) came to the attention of public health authorities towards the end of December 2019 with the occurrence of a large number of pneumonia cases of unknown causes in Wuhan, the capital of Hubei province in China. Huang, *et. al.* (2020) have traced the earliest human infected case to 01 Dec 2019, although it is likely that the virus had been circulating in the population even earlier. Subsequently, the extremely rapid rise of the number of infections made the authorities impose unprecedented city-wide stay-at-home orders (lockdown) in Wuhan and other cities in the province on 23 January 2020. Less restrictive measures for ensuring physical distancing were introduced in several other locations in China. As a consequence, there was discernible decrease in the rate of growth in infections and from 17 February 2020, the number of active cases began to decline. However, as is evident from the time-series shown in Figure 1 (left), the active case count for the entire world started to increase again from 5 March 2020. This resulted from the focal point of the epidemic shifting outside China (where it continued to decline) to countries such as Italy and the United States of America. Indeed, by 8 March 2020, the total number of active cases of the disease outside China exceeded that from China for the first time, and while there have been subsequently resurgences of COVID-19 cases in China, these have been fairly limited in size. With the rapid spread of the epidemic across different countries in Europe and the Americas, apart from Asia, on 11 March 2020, the World Health Organization (WHO) declared the COVID-19 outbreak to be a global pandemic.

Figure 1 (right) shows the reproduction number R estimated from the time-series of the total number of active cases across all countries by using a moving window having different starting dates (t) and interval lengths (Δt). The choices of t and Δt for which the correlation coefficient r between time and logarithm of the number of active cases (that measures how closely the curve describing the number of active cases fits an exponential function) is greater than 0.998 are indicated within the black dotted contour lines (the regions within the blue dotted lines have $r > 0.995$). The corresponding measure of significance is $p < 0.001$. As

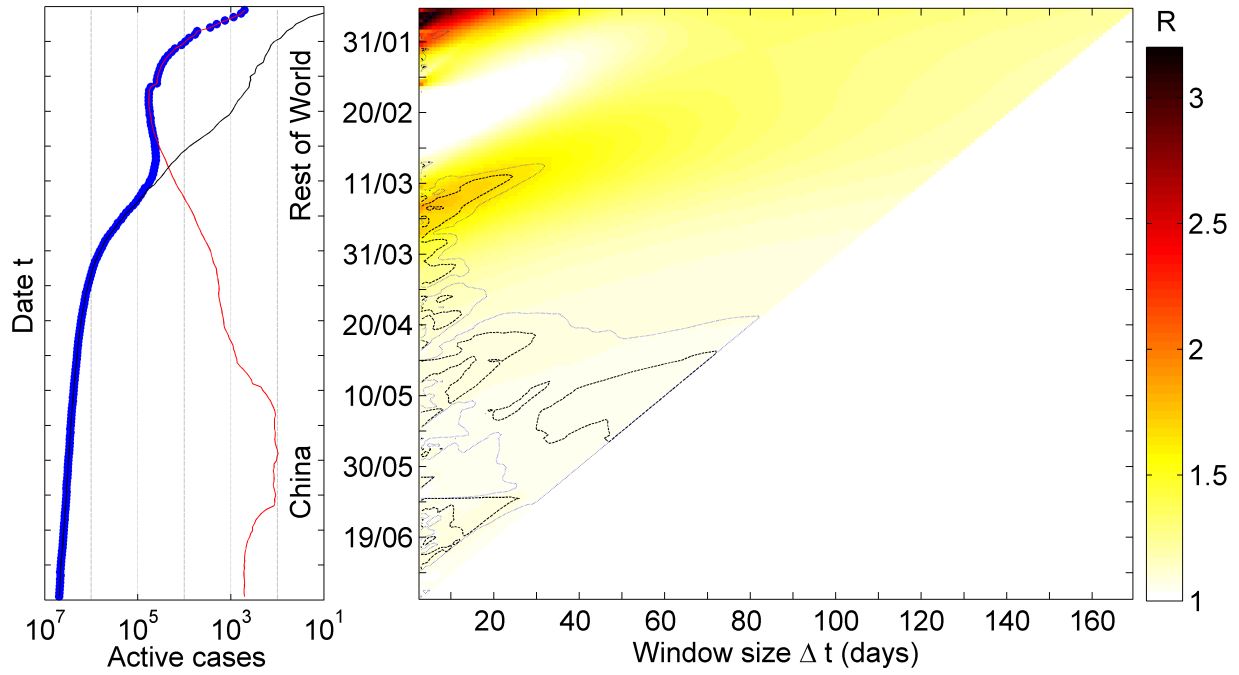

Figure 1: (Left) The time-series for the number of active cases of COVID-19 for the entire world (the data for China and the rest of the World are also shown separately) and (right) the estimated reproduction number R over time windows with different starting dates (t) and temporal intervals (Δt).

the data source we have used does not have information for the period prior to 22 January 2020, we cannot estimate the basic reproduction number. We note that R was 2.02 ± 0.16 between 23-27 January 2020, but then decreased to 1.85 ± 0.13 (02-05 February 2020) and subsequently to 1.31 ± 0.06 (07-10 February 2020). Throughout the second half of February R appeared to be equal or less than 1 so that, even though there were countries such as Italy and United Kingdom that had large outbreaks during the period, it still seemed possible that the disease can be contained and prevented from becoming a global pandemic. However, it started rising again in March, increasing from 1.18 ± 0.04 (06-09 March 2020) to 1.66 ± 0.02 (09 March-01 April 2020) - possibly resulting from the large number of outbreaks that occurred across countries in Europe and Latin America, as well as, USA and South Africa, at this time. From the month of April onward, however, we have seen a steady decrease in the global R, from 1.25 ± 0.01 (03-13 April 2020) to 1.16 ± 0.01 (12-21 April 2020) and 1.12 ± 0.01 (20-28 April 2020), notwithstanding the fact that new territories have been affected by the disease. Over the last couple of months, R has stood at 1.065 ± 0.001 (28 April-08 July 2020) which probably reflects the success of European countries in containing the epidemic and the fact that in USA, despite the large absolute number of cases, the growth rate has decreased substantially.

As the United States of America has the highest number of confirmed cases and is therefore contributing to the value of R for the entire world more than any other country, in Figure 2 we specifically look into how the situation has evolved there, beginning from

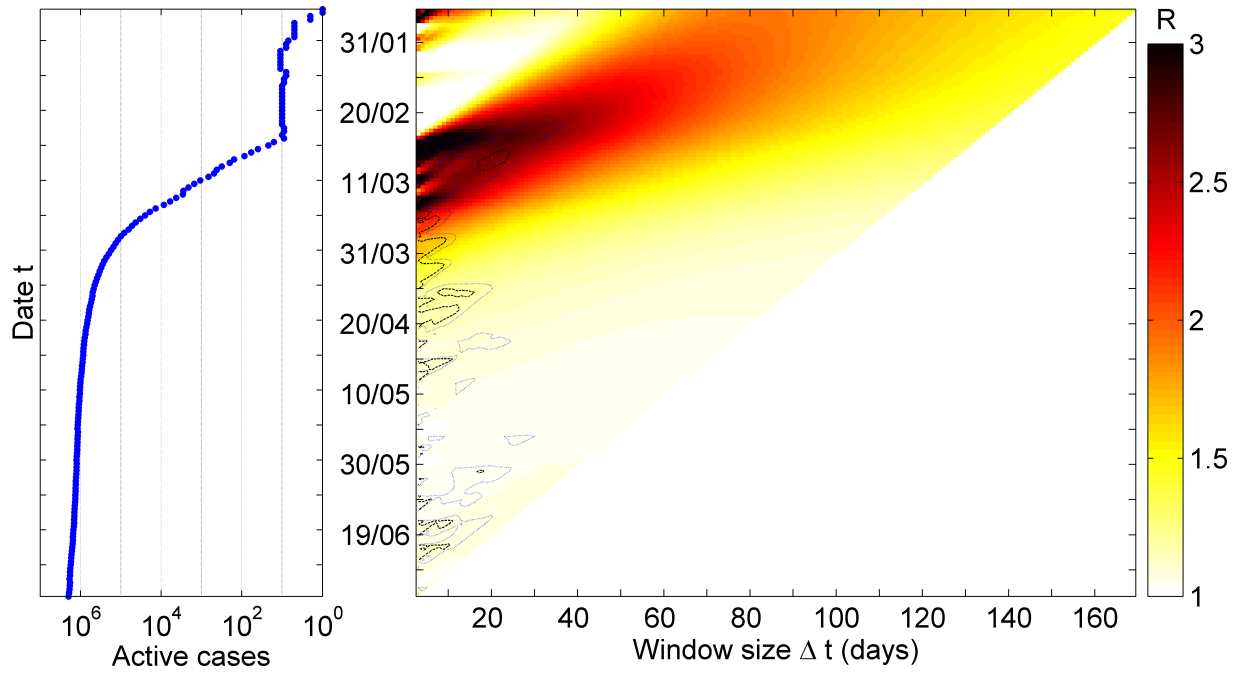

Figure 2: (Left) The time-series for the number of active cases of COVID-19 for United States of America and (right) the estimated reproduction number R over time windows with different starting dates (t) and temporal intervals (Δt).

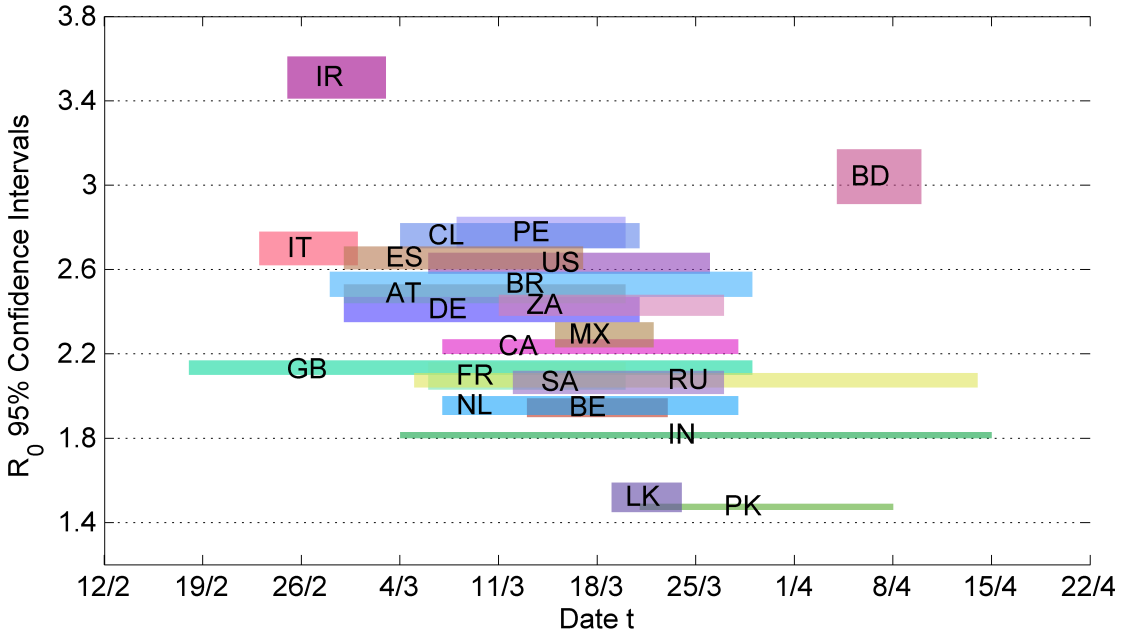

Figure 3: The 95% confidence intervals (represented by the vertical extent of the colored bars representing different nations) for the basic reproduction number R_0 estimated for COVID-19 outbreak in 22 countries across the world with the horizontal extent of the colored bars indicating the time period used for the estimation.

late January. The first case of COVID-19 in USA was reported on 19 January 2020 from Snohomish County in Washington state (Holshue, *et al.*, 2020). Following this, as seen from Figure 2 (left), the number of active cases rose only slightly to reach double digits and remained steady at fairly low numbers (~ 10) for the entire month of February. However, between 03-26 March 2020, the number of cases rose rapidly marking the advent of the epidemic with a basic reproduction number $R_0 = 2.63 \pm 0.05$. The case fatality ratio (CFR) was also high at this early period, peaking at 0.114. It is to be noted, however, that there is a high degree of heterogeneity in the disease incidence across the country, with the state of New York (followed by California, Florida and Texas) accounting for a large fraction of the cases.

As the epidemic unfolded, the value of the effective reproduction number has undergone several changes as can be seen from Figure 2 (right). Between 27 March-4 April 2020, the number reduced from the value of R_0 (mentioned above) to 1.68 ± 0.03. It further reduced to 1.38 ± 0.02 during 4-11 April 2020, 1.18 ± 0.01 during 11-26 April 2020, 1.10 ± 0.01 during 30 April-10 May 2020, and reached its lowest value so far $R = 1.05 \pm 0.01$ in the first half of June (1-18 June 2020). In more recent periods, it has marginally increased back to 1.10 ± 0.01 (22 June-1 July 2020). The trajectory of the epidemic in USA as described by these reproduction numbers is qualitatively similar to that seen for India (described below), with an initial period of extremely rapid spread lasting for about a month followed by gradual reduction in the transmission, with R eventually settling to a value just higher than 1 about four months after the outbreak established itself in the local population.

To see how much variation there is across geographical regions in the rate at which the epidemic has spread, in Figure 3 we graphically represent the 95% confidence intervals of the basic reproduction number for several countries where there have been major outbreaks of COVID-19, along with the period corresponding to the initial phase of growth of the epidemic (over which R_0 has been estimated). China is not included because as mentioned earlier, the data sources being used do not include information on the initial phase of the outbreak in China. The two letter symbols associated with each colored bar indicate the different countries (see Table 1, which provides the numerical values of R_0 for these and several additional countries). Note that all the countries which are currently in the top 15 in terms of confirmed cases have been included.

Even a cursory glance at Figure 3 is sufficient to establish a few exceptional features underlining the diversity in COVID-19 epidemiological dynamics in different locations. While the bulk of the countries investigated have had their R_0 values lying between 2 and 2.8, there have been exceptions such as Iran, which had an unusually high R_0. In contrast, R_0 for countries in South Asia such as Sri Lanka, Pakistan and Nepal (see Table 1) have been very low, with the notable exception of Bangladesh which had a R_0 of around 3. The R_0 for India, while higher than its southern and western neighbors, is still at the lower end of the range of values for the basic reproduction number that we have estimated for different countries.

4. COVID-19: The situation in India

Having discussed the international situation, we now focus on how the epidemic has

Table 1: Basic reproduction numbers (R_0) estimated for COVID-19 outbreaks in different countries

Region	Country	R_0	95% CI	r	p	Period	Peak CFR
North America	USA (US)	2.63	[2.58, 2.68]	0.998	< 0.001	03-26 Mar 2020	0.114
	Mexico (MX)	2.29	[2.23, 2.35]	0.999	< 0.001	15-22 Mar 2020	0.124
	Canada (CA)	2.23	[2.20, 2.27]	0.998	< 0.001	07-28 Mar 2020	0.082
South America	Brazil (BR)	2.53	[2.47, 2.59]	0.995	< 0.001	26 Feb-29 Mar 2020	0.070
	Peru (PE)	2.77	[2.70, 2.85]	0.998	< 0.001	08-20 Mar 2020	0.042
	Chile (CL)	2.76	[2.70, 2.82]	0.998	< 0.001	04-21 Mar 2020	0.022
Europe	Italy (IT)	2.70	[2.62, 2.78]	0.999	< 0.001	23 Feb-01 Mar 2020	0.145
	Spain (ES)	2.66	[2.60, 2.71]	0.998	0.001	29 Feb-Mar 17 2020	0.122
	UK (GB)	2.13	[2.10, 2.17]	0.995	< 0.001	18 Feb-29 Mar 2020	0.155
	France (FR)	2.10	[2.03, 2.16]	0.995	< 0.001	06-20 Mar 2020	0.159
	Belgium (BE)	1.94	[1.90, 1.99]	0.998	< 0.001	13-23 Mar 2020	0.165
	Netherlands (NL)	1.96	[1.91, 2.00]	0.995	< 0.001	07-28 Mar 2020	0.129
	Germany (DE)	2.41	[2.35, 2.47]	0.995	< 0.001	29 Feb-21 Mar 2020	0.047
	Austria (AT)	2.49	[2.44, 2.53]	0.998	< 0.001	29 Feb-20 Mar 2020	0.040
	Russia (RU)	2.08	[2.04, 2.11]	0.995	< 0.001	05 Mar-14 Apr 2020	0.015
Africa	South Africa (ZA)	2.43	[2.38, 2.48]	0.998	< 0.001	11-27 Mar 2020	0.022
Middle East	Iran (IR)	3.51	[3.41, 3.61]	0.999	< 0.001	25 Feb-03 Mar 2020	0.079
	Turkey (TR)	4.75	[4.52, 4.97]	0.998	< 0.001	14-21 Mar 2020	0.028
	Saudi Arabia (SA)	2.06	[2.01, 2.12]	0.996	< 0.001	12-27 Mar 2020	0.015
South Asia	India (IN)	1.82	[1.80, 1.83]	0.998	< 0.001	4 Mar-15 Apr 2020	0.036
	Pakistan (PK)	1.48	[1.46, 1.49]	0.998	< 0.001	21 Mar-08 Apr 2020	0.024
	Bangladesh (BD)	3.04	[2.91, 3.17]	0.998	< 0.001	04-10 Apr 2020	0.128
	Sri Lanka (LK)	1.52	[1.45, 1.59]	0.995	< 0.001	19-24 Mar 2020	0.037
	Nepal (NP)	1.48	[1.47, 1.49]	0.995	< 0.001	25 Mar-1 Jul 2020	0.007

developed in India. The first confirmed case of COVID-19 in India was recorded when an Indian student at Wuhan returned to Kerala on 30 January 2020. Subsequently two more Indian students returning from Wuhan were also tested to be positive for the disease in early February 2020. All of them subsequently recovered without having passed the infection to anybody else in India. However, the situation altered beginning from 02 March 2020, when an Indian citizen who had traveled to Austria tested positive in East Delhi, who was soon found to have infected six of his family members in Agra (all of whom tested positive on March 4). In a parallel development, after a member of an Italian tourist group tested positive on 03 March 2020, 16 other members of the group were found to have contracted the infection on the next day. Thus, with 22 new infected individuals being detected on 04 March 2020, the number of COVID-19 cases in India crossed single digits for the first time (see Figure 4). Following this the number of new cases steadily rose [see Figure 5 (left)], alarming the public health authorities into considering implementing extreme non-pharmaceutical intervention measures such as those successfully employed to control the epidemic in China. A "Janata curfew" (People's or self-imposed curfew) for 14 hours on 22 March 2020, a Sunday, tested the feasibility of imposing a nation-wide lockdown, and it was followed up by a stay-at-home order implemented from 24 March 2020. Figure 4 shows also the composition of the new cases reported each day between those returning from abroad and those who contracted it

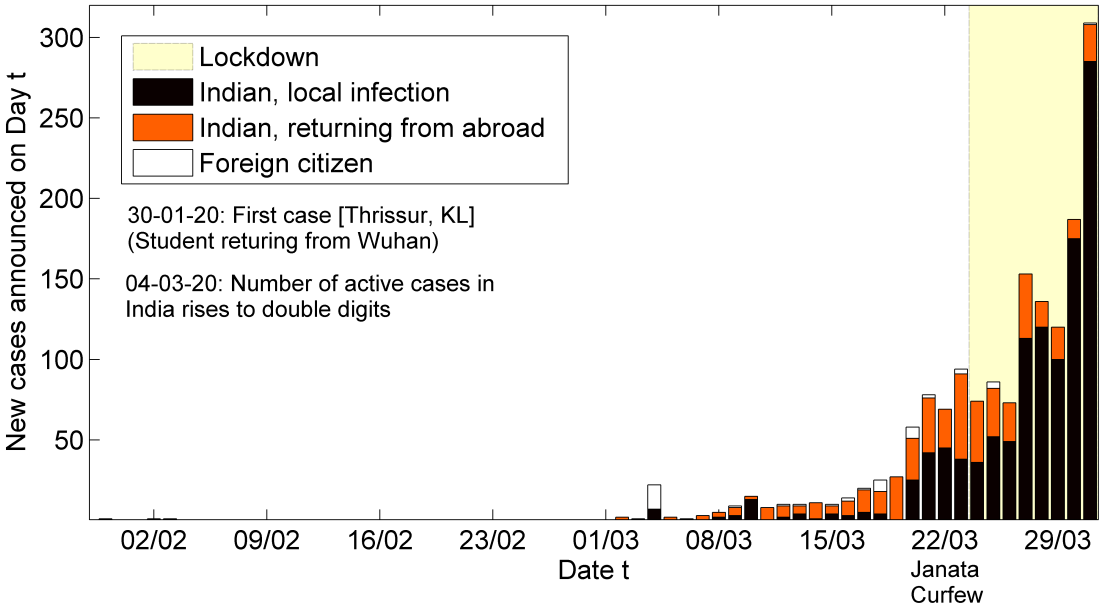

Figure 4: Initial phase of the COVID-19 epidemic in India, showing the first 1635 **confirmed cases and distinguishing between infections that were imported from abroad, either by returning Indians or visiting foreign citizens, and those that occurred among the local population.**

locally, and we note that, after 24 March 2020, the bulk of the cases were of the latter kind. It suggests that by this time, the disease had established itself in the local population.

Figure 5 shows how the reproduction number for the disease has evolved in India in response to the various measures that were successively put in place. Specifically, we indicate the various stages of the lockdown (differing in terms of the severity of the measures imposed to ensure social distancing) that was imposed from 23 March 2020, *viz.*, Phase 1 (23 March-14 April 2020), Phase 2 (15 April-3 May 2020), Phase 3 (4-17 May 2020), and Phase 4 (18-31 May 2020), followed by Unlock 1 (1-30 June 2020) and Unlock 2 (1-31 July 2020) which is still underway. In the initial stage between 04 March-15 April 2020 over which the R_0 value is estimated, the number of active cases rose from 25 to 10485. After this period, the rate of spreading lessened to a large extent and the effective reproduction number between 14 April-16 May 2020 was estimated to be 1.28 ± 0.01, which is a reduction of 30% from the value of $R_0 (\simeq 1.82)$. To understand the significance of this change in R, we note that had the epidemic continued with its initial growth rate for much longer, then we would have been very likely to have crossed 1 million active cases before the middle of May (see Figure 6). This would have undoubtedly put enormous stress on the medical infrastructure of the country.

The reduction in R may be attributed at least partially to the imposition of the lockdown and other related measures (such as, asking people to wear masks in public, *etc.*), especially as the deviation from the initial trend can be observed from the data by 06 April 2020 (see Figure 6), *i.e.*, after approximately two weeks following the imposition of Phase

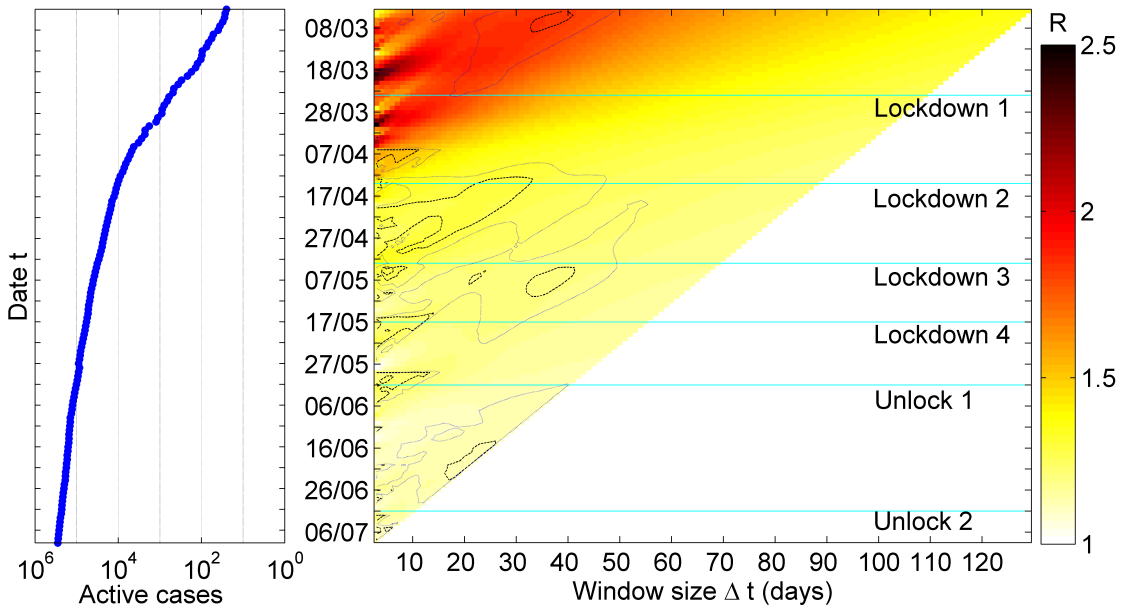

Figure 5: (Left) The time-series for the number of active cases of COVID-19 for India and (right) the estimated reproduction number R over time windows with different starting dates (t) and temporal intervals (Δt). The periods of different stay-at-home orders (referred to as Lockdown and Unlock) are indicated.

1 of the lockdown. The duration of this lag between an intervention and its manifestation in terms of changes in the number of cases is because a person infected with SARS-CoV-2 can take up to 14 days to manifest symptoms, upon which time they can be tested and then quarantined. Until this time, such individuals may be freely circulating in the population and aiding in the transmission of the pathogen. Thus, the bulk of the confirmed cases that were reported in the days immediately following the lockdown imposition would have resulted from infections that took place in the period prior to it.

Between 16-28 May 2020, the value of R slipped further to 1.22 ± 0.01, followed by a marginal decrease to $R = 1.21 \pm 0.01$ during 29 May-11 June 2020. Subsequently, during 12 June-11 July 2020, R reached its lowest value (up till the time of writing) of 1.13 ± 0.01. However, this continually decreasing trend in R was then broken and the most recent value was estimated to be 1.19 ± 0.01 between 11-20 July 2020. As already hinted in the previous section, this trajectory of the spreading dynamics, decrease in R for four months followed by a slight upward turn, resembles that of USA. Taking into account the approximately two-week delay between an event involving a population and its effect manifesting in the epidemiological data, the present rise in R can be possibly related to the relaxation of lockdown norms in the second half of June.

To obtain a better understanding of the temporal variation in the epidemiological dynamics, we should consider more spatially detailed information. As seen from the pie chart in Figure 7, the disease has not affected all regions of the country uniformly. Maharashtra

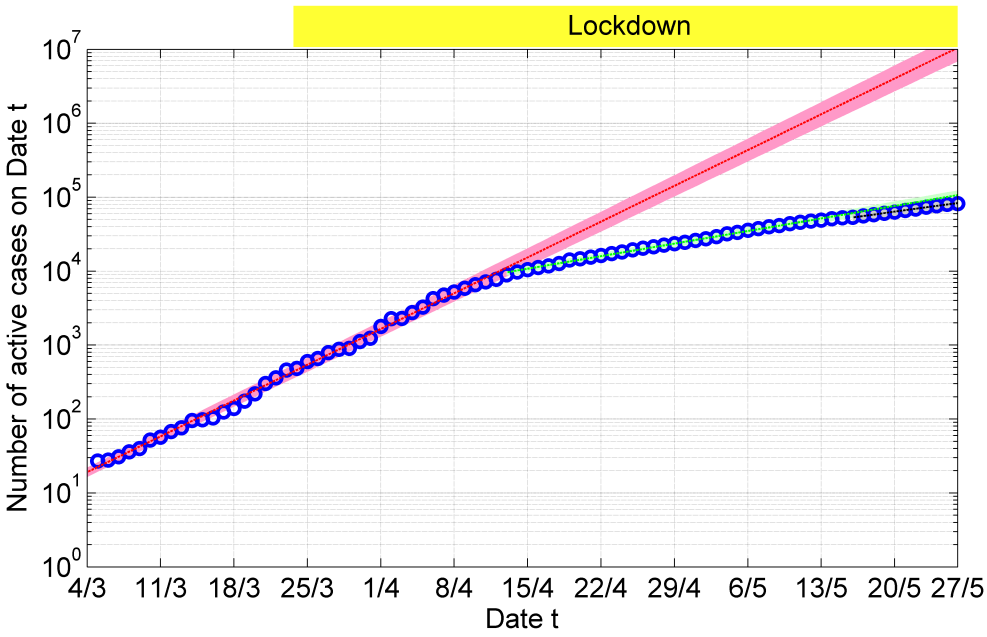

Figure 6: The progress of COVID-19 epidemic in India between 04 March-27 May 2020 showing the daily number of active cases (circles) in logarithmic scale. Log scale is used to visualize the quality of fit of the data to an exponential curve, that manifests as a straight line in such a scale. The red dotted curve indicates the projected increase in active cases in April and May had the epidemic continued to progress according to the rate given by the basic reproduction number of 1.82. The 95% confidence intervals are indicated by the shaded regions, and the period under lockdown by the colored horizontal bar on top.

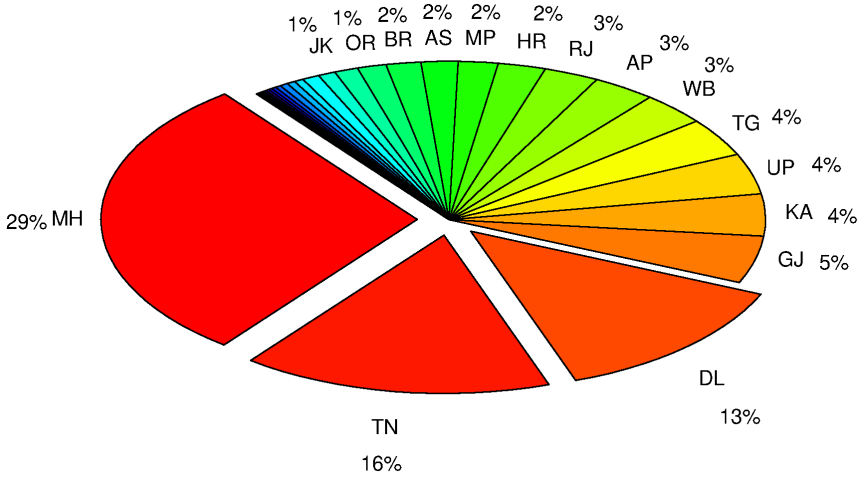

Figure 7: Pie chart showing the percentage contribution of the different states of India to the total number of confirmed cases of COVID-19 till 11 July 2020.

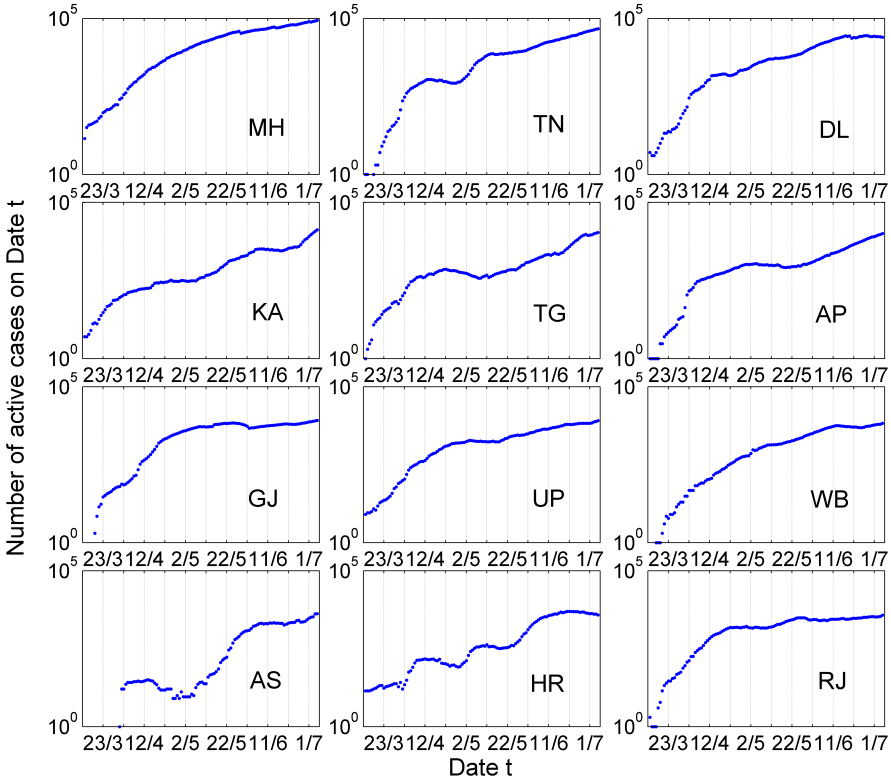

Figure 8: Progress of the epidemic in Indian states having the highest number of active cases on 05 July 2020 (in decreasing order, left to right & top to bottom).

has contributed to almost a third of the entire burden of the country, and along with Tamil Nadu and Delhi, are the three states which account for about 60% of all confirmed cases till date. On the other hand, both Kerala and Punjab, which were some of the first states to be affected, surprisingly had less than 1% of the total number of confirmed cases till July. Figure 8 also suggests a substantial amount of spatio-temporal heterogeneity in the manner in which the epidemic has evolved across the country, with the disease peaking early in west (*e.g.,* Maharashtra) and north (*e.g.,* Delhi), followed by the south (*e.g.,* Karnataka) and much later by the east (*e.g.,* West Bengal and Assam). While many of the states show a trajectory similar to that for the country, *viz.* a rapid growth phase initially, followed by a slowing of the spreading while continuing to be an epidemic (i.e., $R > 1$), certain states like Tamil Nadu and Haryana exhibit multiple rounds of growth and decay of the epidemic. Thus, between 12-30 April 2020, Tamil Nadu had a value of R around or less than 1, so that the number of active cases were decreasing over time indicating that more recoveries were happening than new infections. [1] However, the appearance of a cluster of cases originating in the Koyambedu wholesale market in Chennai towards the end of this period resulted in the

[1]We would like to note here that $R \sim 1$ does not necessarily imply that the number of active cases has remained constant, for instance, because the number of new infections reported each day is exactly balanced by the number of infected individuals who are recovering daily. Rather, the growth is slower than exponential (sub-exponential), *e.g.,* following a trend that fits a polynomial trend.

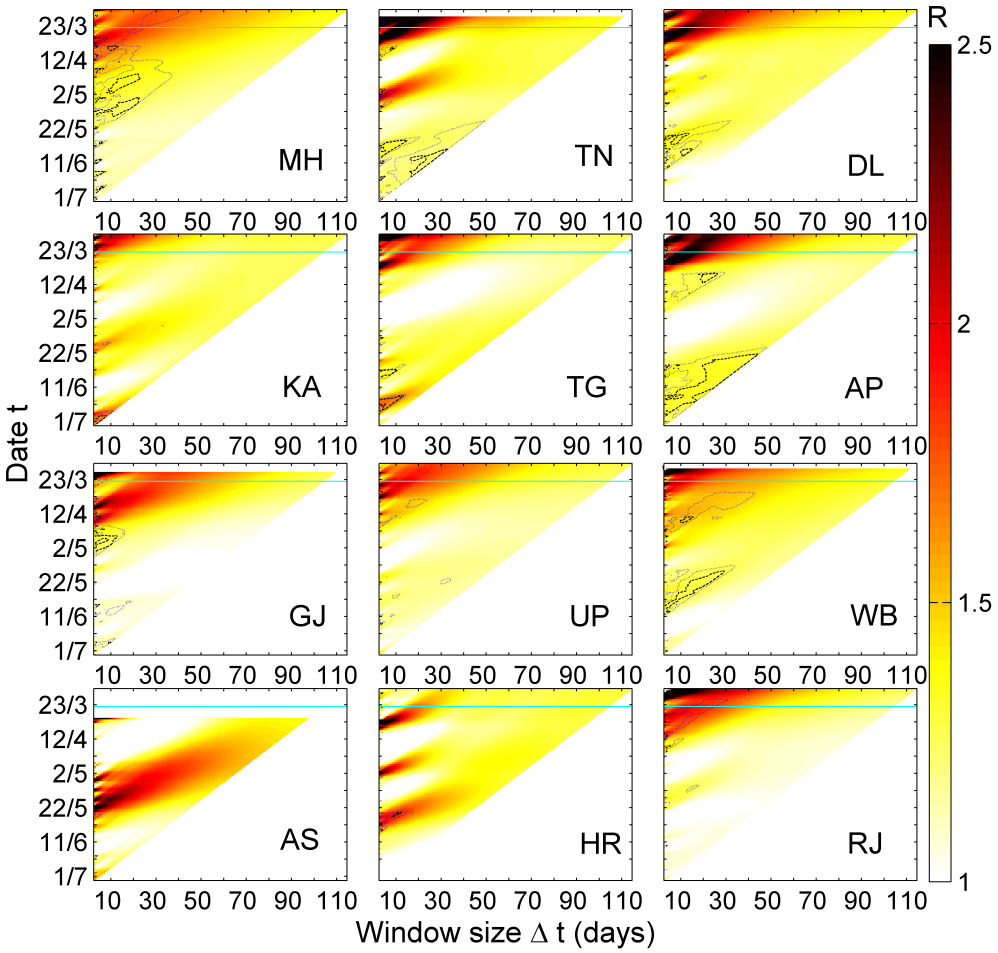

Figure 9: The basic/effective reproduction numbers for the states of India having the highest number of active cases as on 05 July 2020, estimated from the time-series of the number of active cases by using a moving window having different starting dates and interval lengths (Δt).

number of infections across the state to increase rapidly and R rose to 2.01 ± 0.10 between 30 April-07 May 2020. Subsequently, it decreased to 1.56 ± 0.06 between 06-13 May 2020 and then further to 1.31 ± 0.03 during 30 May-04 June 2020.

A perusal of the evolution of the reproduction number for the individual states (see Figure 8 where the horizontal line in each panel indicates the date of imposition of the national lockdown) shows the diversity of outcomes as COVID-19 has spread through India. Not surprisingly, the R value for the state of Maharashtra has largely driven that for the entire country, as it accounts for the largest share of COVID-19 cases among the states, even though it may not have had the highest value of R among them. To see why this is the case, consider a hypothetical situation where a country has an epidemic raging in two of its states with two very different values of R, *e.g.*, 2 in state A and 4 in state B. Consider also that at a given time, state A has 1000 active cases, while state B has 50 cases. After a time period

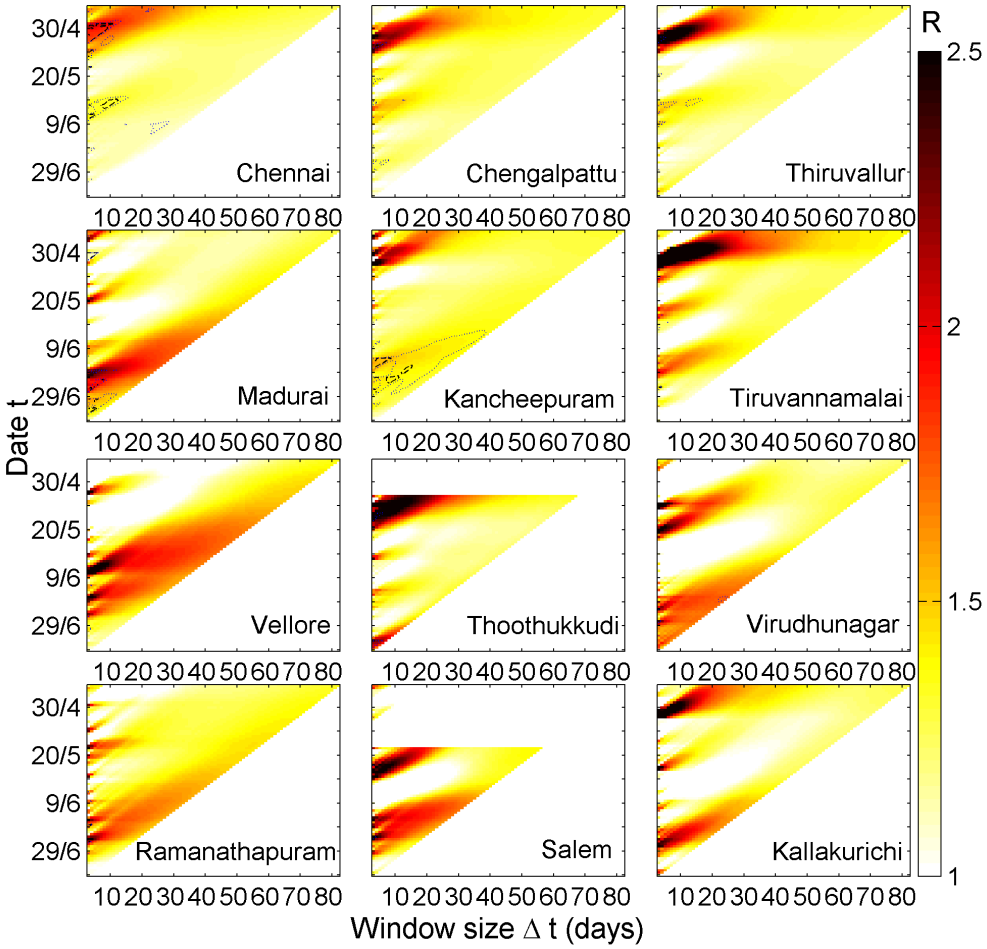

Figure 10: The basic/effective reproduction numbers for the districts of Tamil Nadu having the most confirmed cases as on 15 July 2020, estimated from the time-series of the number of active cases by using a moving window having different starting dates and interval lengths (Δt).

corresponding to one generation interval, state A will have 2000 active cases, while state B will have 200 cases (as per the definition of R). Thus, the effective reproduction number for the country as a whole will be 2.2, a value that is quite close to the R for state A which has the bulk of the active cases even though the epidemic is spreading much slower there than in state B. Note that if the growth rate for the two states remain unchanged, state B will soon surpass state A in terms of active cases and from that point onward will have its R value dominating the national R.

Similar heterogeneity is observed at an even finer spatial scale when we consider the evolution of the epidemic in each district of a state. Figure 10 shows the estimated reproduction numbers for twelve districts in Tamil Nadu that have had the highest number of confirmed cases. It can be easily observed that there has been much more temporal variability in the epidemiological dynamics at this more spatially resolved scale. Most districts show

multiple flare-ups of the epidemic growth rate because of local spreading events, followed by periods during which R has substantially decreased. We also note that the growth is far from being synchronized across a state. For example, while Madurai district had R of 1.41 ± 0.04 during 04-09 July 2020, at around the same time Chennai had the epidemic under effective control so that R was less than 1 (we estimate R as 0.78 ± 0.02 in the slightly later period of 10-14 July 2020). While data for spatial resolution higher than this is not publicly available, it seems reasonable to conclude that the spatio-temporal heterogeneity will be even more pronounced at that scale, not the least because of the relatively stronger daily fluctuations in the number of cases. In contrast, at the level of a state, and more so for a country as large as India, a self-averaging process occurring through rises in some regions being balanced by dips in others, decreases the fluctuations and aids the statistical analysis.

5. Discussion

While the reproduction number is not the only possible metric one can associate with an epidemic, nor does it necessarily contain all relevant information about the epidemic, it is nevertheless extremely informative about the dynamical process by which the disease is spreading (Heesterbeek, 2002). Its value is determined by multiple factors associated with demography and social structure of the population in which the disease is spreading, as well as, the biology of the pathogen, *viz.,* (i) the generation time (which can be considered as the period over which an infected individual passes the pathogen to others), (ii) the mean number of contacts between susceptible and infectious individuals, (iii) the probability of an infection resulting from such a contact, and (iv) the size of the susceptible population (Sinha, 2020). Typically, not all of these factors may be known for an epidemic. Thus, estimating R from empirical data can provide us with a means of making inferences about such factors. The basic reproduction number is also of practical importance from a public health perspective, as using the estimated value of R_0 we can estimate the overall burden of the disease, as well as, in the event of availability of a vaccine for the disease, the fraction of population who will need to be vaccinated to achieve herd immunity. It is for these purposes that it is imperative to accurately estimate R_0. We note in passing that the basic reproduction number for the pandemic is in the same range as the infamous Spanish Flu pandemic of 1918-19 [as estimated by Mills, *et al.* (2004)].

To conclude, it should be stressed that the estimated value of R_0 is at the lower end of the values reported for different regions in which the outbreak of COVID-19 has resulted in a large number of infected cases. Indeed, this seems to be true for most countries from the South Asian region (barring the notable exception of Bangladesh). Preliminary analysis of physical and climatic factors done by us appears to rule out the direct role of these in making R_0 for India low. While reliability of the available epidemiological data may be an issue, it is unlikely that this alone can be the explanation, because under-reporting, as long as it is done consistently at the same level over time, will not significantly alter the estimated value of R. The possibility that genetic or physiological features of the South Asian population may be responsible is a hypothesis that needs further investigation. One of the intriguing questions that arise from the analysis is the fact that the growth rate of the disease has continued to be low despite a large degree of relaxation that has happened in the lockdown norms. As India is still very far from achieving herd immunity, and there appears to be no

evidence that the pathogen has shown any change in its ability to infect, it is possible that the voluntary adherence to public hygiene has been responsible. If true, this may suggest that, under certain conditions, citizens can be mobilized to engage in a mass effort to achieve desirable public health outcomes.

Acknowledgments

I thank Izhar M. Ashraf for technical assistance with acquiring data and Soumya Easwaran for assistance with data analysis in the early phase of the project. The work was supported in part by the Center of Excellence in Complex Systems and Data Science, The Institute of Mathematical Sciences, funded by the Department of Atomic Energy, Government of India.

References

COVID19-India (2020). *A volunteer-driven, crowd-sourced database for COVID-19 stats and patient tracing in India.* https://www.covid19india.org/.

CSSE (2020). *COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University.* GitHub https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series.

Ganyani, T., Kremer, C., Chen, D., Torneri, A., Faes, C., Wallinga, J. and Hens, N. (2020). Estimating the generation interval for COVID-19 based on symptom onset data. *medRxiv* https://doi.org/10.1101/2020.03.05.20031815

Heesterbeek, J. A. P. (2002). A brief history of R_0 and a recipe for its calculation. *Acta Biotheoretica*, **50(3)**, 189-204.

Holshue, M. L., DeBolt, C., Lindquist, S., Lofy, K. H., Wiesman, J., Bruce, H., Spitters, C., Ericson, K., Wilkerson, S., Tural, A. and Diaz, G. (2020). First case of 2019 novel coronavirus in the United States. *New England Journal of Medicine*, **382(10)**, 929–936.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X. and Cheng, Z. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*, **395(10223)**, 497–506.

India Today (2020). *21-day lockdown in entire India to fight coronavirus, announces PM Narendra Modi.* Last updated March 25, 2020 00:05 IST, Retrieved on 17 July 2020 from https://www.indiatoday.in/india/story/india-lockdown-pm-narendra-modi-speech-coronavirus-1659266-2020-03-24.

MathWorks (2009). *MATLAB R2009b.* The MathWorks Inc., Natick, Massachusetts. https://in.mathworks.com/help/matlab/release-notes-R2009b.html.

Kermack, W.O. and McKendrick, A.G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London,* **A115 (772)**, 700–721.

Lai, S., Ruktanonchai, N. W., Zhou, L., Prosper, O., Luo, W., Floyd, J. R., Wesolowski, A., Santillana, M., Zhang, C., Du, X. and Yu, H. (2020). Effect of non-pharmaceutical interventions for containing the COVID-19 outbreak in China. *medRxiv* https://doi.org/10.1101/2020.03.03.20029843.

Jesan, T., Menon, G. I. and Sinha, S. (2011). Epidemiological dynamics of the 2009 influenza A (H1N1) v outbreak in India. *Current Science,* **100(7)**, 1051–1054.

Johnson, N. P. and Mueller, J. (2002). Updating the accounts: global mortality of the 1918-1920 "Spanish" influenza pandemic. *Bulletin of the History of Medicine,* **76(1)**, 105–115.

Mills, C. E., Robins, J. M. and Lipsitch, M. (2004). Transmissibility of 1918 pandemic influenza. *Nature*, **432(7019)**, 904–906.

Kelly, H., Peck, H. A., Laurie, K. L., Wu, P., Nishiura, H. and Cowling, B. J. (2011). The age-specific cumulative incidence of infection with pandemic influenza H1N1 2009 was similar in various countries prior to vaccination. *PLoS ONE* **6(8)**, e21828.

Reyes, O., Lee, E. C., Sah, P., Viboud, C., Chandra, S. and Bansal, S. (2018). Spatiotemporal patterns and diffusion of the 1918 influenza pandemic in British India. *American Journal of Epidemiology*, **187(12)**, 2550–2560.

Sinha, S. (2020). Why should we be mindful of nonlinear dynamics in the midst of a global pandemic. *Pramana: Journal of Physics*. To appear.

Spicer, C. C. and Lawrence, C. J. (1984). Epidemic influenza in Greater London. *The Journal of Hygiene,* **93(1)**, 105–112.

Spreeuwenberg, P., Kroneman, M. and Paget, J. (2018) Reassessing the global mortality burden of the 1918 influenza pandemic. *American Journal of Epidemiology,* **187(12)**, 2561–2567.

Wang, C., Horby, P. W., Hayden, F. G. and Gao, G. F. (2020). A novel coronavirus outbreak of global health concern. *The Lancet*, **395(10223)**, 470–473.

WHO (2020a). *MERS situation update, January 2020.* Retrieved on 17 July 2020 from http://www.emro.who.int/pandemic-epidemic-diseases/mers-cov/mers-situation-update-january-2020.html.

WHO (2020b). *WHO Coronavirus Disease (COVID-19) Dashboard.* Retrieved on 17 July 2020 from https://covid19.who.int/

WHO (2003). *Cumulative Number of Reported Probable Cases of SARS.* Retrieved on 17 July 2020 from https://www.who.int/csr/sars/country/2003_07_04/en/.

WikipediA: The Free Encyclopedia. National responses to the COVID-19 pandemic: Lockdowns. Retrieved on 17 July 2020 from https://en.wikipedia.org/wiki/National_responses_to_the_COVID-19_pandemic#Lockdowns.