Lockdown Effect on COVID-19 Spread in India: National Data Masking State-Level Trends

Deepankar Basu[1], Maxwell Salvatore[2,3], Mike Kleinsasser[2], Soumik Purkayastha[2], Rupam Bhattacharyya[2], Bhramar Mukherjee[2,3]

[1]: Department of Economics, University of Massachusetts, Amherst
[2]: Department of Biostatistics, University of Michigan, Ann Arbor
[3]: Center for Precision Health Data Science, University of Michigan, Ann Arbor

Abstract: National data on the COVID-19 outbreak in India is of limited use for nuanced analyses or informed policy interventions. The immense variation of the pandemic across states is obscured by our focus on national-level data and hampers assessment of effective policy measures.

We illustrate the masking of state-level trends by national data with a sample of 20 states that reported the highest number of COVID-19 cases by mid-May. Taken together, these 20 states accounted for more than 99% of the cumulative case and death counts in India on May 17, 2020. We highlight variation across states by presenting evidence on many crucial aspects of the COVID-19 outbreak from a public health point of view: case and death counts with case fatality rates; doubling times of cumulative number of cases; effective basic reproduction numbers; and the scale of testing.

Some interesting patterns of change over lockdown periods emerge. For example, the estimated effective reproduction number for India stood at 3.36 (95% confidence interval (CI): [3.03, 3.71]) on March 24, the day on which the lockdown was announced, whereas the average for the same over the week of May 12 - May 18 stands at 1.24 (95% CI: [1.22, 1.26]). Similarly, the estimated doubling time at an all-India level was at 3.56 days on March 24, and the past 7-day average for the same on May 18 is 13.08 days. The case fatality rate for India is estimated to be 3.25% (95% CI: [3.04%, 3.14%]) based on all confirmed cases, as of May 18. The average daily number of tests have increased from 1717 (March 19-25) to 99,403 (May 12-18) with test positivity rate during March 12-18 around 2.0% (Range: [1.3%, 2.4%]) and May 12-18 at an average of 3.9% (Range: [3.7%, 4.1%]). However, these overall national metrics do not reflect the state level heterogeneity. We present a comprehensive summary that can capture recent trends in all of these public health metrics across the states to aid a policymaker to titrate resources, set priorities and evaluate lockdown effects.

By looking at several public health and policy relevant measures, our framework offers a holistic picture of the pandemic across Indian states and union territories. Our analysis has a simple take-home message for analysts and policy makers. The COVID-19 outbreak in India, as in any other large country, displays large state-level variations. Identifying these variations can help in both understanding the dynamics of the pandemic and formulating effective public health policies of containment and mitigation. The visualization tools are available in our interactive application at covind19.org.

1. Introduction:

Coronavirus disease of 2019 (COVID-19) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Mayo Foundation for Medical Education and Research, 2020). It was first identified in December 2019 in Wuhan, China, and has since spread globally, resulting in an ongoing pandemic (Hui, et al., 2020). As of May 21, 2020, more than 5 million cases have been reported across 188 countries and territories, resulting in more than 328,000 deaths. More than 1.91 million people have recovered (Johns Hopkins University, 2020). In order to attenuate the spread of this pandemic, India implemented a strict population-wide lockdown on March 25, 2020. According to latest government announcements, the lockdown is in effect till May 31 (Pradhan and Sen, 2020). As of May 21, the number of total confirmed cases in India has crossed 112,000, of whom 3,456 have died and 45,886 have recovered (COVID 19 INDIA, 2020).

In light of tremendous public interest, numerous data repositories, along with statistical models, are being developed with the aim of studying and analyzing COVID-19 epidemiology. The focus of modeling has shifted from producing precise quantitative predictions about the extent or duration of disease burdens to estimating and evaluating the effect of various interventions on the spread of the virus (Jewell, Lewnard, & Jewell, 2020). With databases becoming increasingly granular, models are not only focusing on national trends, but also state-level, and even district-level trends in India. As of May 21, 2020, 3,824 COVID-19 SARS-CoV-2 preprints have been uploaded to medRxiv and bioRxiv, of which at least 30 articles focus on analyzing the efficacy of the social distancing measures implemented by the Indian government. A few measures used to judge the performance of epidemic curtailment strategies (such as case-isolation and population-level lockdown) include basic and effective reproductive numbers, doubling time and growth rates.

(Ray, et al., 2020) studied the short- and long-term impact of the initial stages of the lockdown on the total number of cases in India using standard epidemiological forecasting models, making use of Bayesian estimation methods. The authors concluded that the lockdown, if implemented properly stood a good chance of reducing the total number of cases in India in the short term. This finding has since been supported by subsequent studies. Noting the size and diversity of India, (Ghosh, Ghosh, & Chakraborty, 2020) investigated the spread of the virus and subsequent impact of preventive measures on the same at a state-level in India. The authors considered a wide variety of statistical models to analyze data gathered from both official, as well as crowdsourced data repositories, and noted that the lockdown has had differential effects on daily infection rates for various states in India. While some states do not show a decreasing trend in daily infection rates, some report an 'almost-decreasing trend', while a few states exhibit a decreasing trend and decreasing growth in active infected cases. (Jakhar, Ahluwalia, & Kumar, 2020) modeled data released by the Ministry of Health and Family Welfare, Government of India using the classical SIR (Susceptible-Infected-Recovered) model. In addition to predicting the trajectories of infection and recovery curves, the authors calculated the basic reproduction number (R_0) for India as a whole, along with state-specific values of the same. Along similar lines, (Gupta, et al., 2020) focused on the estimation of key epidemiological parameters and evaluate the effect of control measures on the COVID-

19 epidemic in India and its states. Using crowdsourced data, the authors computed both basic and effective reproduction numbers at a national, as well as state, level in India after adjusting for imported cases and reporting lags. Through a dynamic compartment-based SEIR-QDPA modeling approach which accounted for asymptomatic cases, the authors assessed the impact of lockdown relaxation and increased testing. Both these papers noted that state-specific R_0 values exhibit high variability when compared to the computed national value of R_0. Using baseline R_0 values (calculated 15 days into the lockdown) and effective R_0 values (calculated 30 days into the lockdown), along with doubling times and growth rates, another article (Mitra, Pakhare, Roy, & Joshi, 2020) suggested that curtailment strategies employed by the Indian government seems to have been effective in controlling the spread of the pandemic in India. However, much is left to be done, as noted by (Venkatesan, 2020). While the reproduction number has come down significantly during the lockdown period, both at the national level and in most states, the computed R_0 values suggest that 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, as suggested by (Rajendrakumar, et al., 2020). Using crowdsourced time series data, the authors compute time dependent reproductive numbers for India and high burden states using a variety of statistical methods. They used a standard SIR model, modified to accommodate an extra compartment for deaths, and worked with sequential Bayesian methods to model the spread of the virus.

Although the statistical methods employed along with reported findings vary, almost all studies point to the underlying variability of state level reproduction numbers when compared to their national analogs. This lurking variability might have an adverse effect on formulation of a long-term strategy to ensure resumption of normal social mobility and economic growth in India. The need of the hour is to study and analyze infection, recovery and fatality trends on a more granular level so as to ensure the formulation of effective public health policies aimed at containment and mitigation.

Faced with the prospect of a raging epidemic, on March 24, the Indian government implemented one of the strictest lockdowns in the world (Mehrotra, 2020). Did the lockdown succeed in slowing down the spread of the pandemic in India? There is a lively debate on this issue in the Indian media, with answers varying with measures used to track spread (Rukmini S., 2020), (Narain, 2020).Moving forward, one can learn from the strategies that have been successful in various states. We look at an ensemble of metrics across states for a deeper and policy-relevant understanding of the situation after three contiguous periods of lockdown from March 25 to May 18 (Lockdown 1.0: March 25-April 14, Lockdown 2.0: April 14-May 3, Lockdown 3.0: May 3-May 18).

2. Methods:

In addition to case and death counts we present measures of spread and testing adequacy. The detailed definition of each reported metric and methods for computation of corresponding measures of uncertainty are presented in the supplementary methods section. We use publicly available data from (covid19India.org, 2020) and (Roser, Ritchie, Ortiz-Ospina, & Hasell, 2020) for all our analysis. All source code and interactive plots are available at (COV-IND-19 Study Group, 2020).

3. Results:

Total number of cases and deaths

India had reported its first case of COVID-19 on January 30. The first death from COVID-19 was reported about 6 weeks later, on March 12. Since then, the epidemic has spread rapidly in the country. The cumulative number of reported COVID-19 cases topped 100,000 on May 18, even as total reported COVID-19 deaths had crossed 3,000 a day earlier. In the second week of May, India recorded the highest growth in case counts among Asian countries (Pradhan, 2020). As the pandemic has spread across the world, India has gradually climbed the list of countries with high case counts. On May 18, only 10 countries in the world, all outside Asia, had recorded more cases than India (Johns Hopkins University, 2020).

Appendix Figure 1 highlights national trends of the COVID-19 outbreak in India by plotting the cumulative number of confirmed cases, fatalities, and recovered cases. It highlights the rising caseload and deaths at the national level. But these national trends obscure an important fact: the pandemic's spread has a pronounced geographic pattern. Figure 1 compares the daily profile of the pandemic at the national level with four states: Kerala, Punjab, Maharashtra and Delhi in terms of daily new cases, recovered cases and fatalities. It is clear that Punjab has been doing well, Kerala seems to have some new cases after the strong initial control, Maharashtra has an increasing trend while Delhi has high number of cases but stabilizing slowly but surely. Since Maharashtra contributes nearly 35-40% of India's total number of cases, the national pattern has more resemblance to Maharashtra,

To highlight this geographic pattern, let us offer snapshots at three dates, separated by a month each, using data from (COVID-19 India, 2020). On March 18, the top 10 states in terms of case counts were Maharashtra, Kerala, Uttar Pradesh, Haryana, Karnataka, Telangana, Delhi, Ladakh, and Jammu & Kashmir. India had reported a total of 171 cases by that day, and these top 10 states accounted for 93.5% of the total cases. A month later, on April 18, India had a total of 15,724 cases, and the top 10 states – Maharashtra, Delhi, Madhya Pradesh, Gujarat, Tamil Nadu, Rajasthan, Uttar Pradesh, Telangana, Andhra Pradesh, and Kerala – accounted for 87.9% of the total cases. On May 18, India's total case count was 100,304, and the top 10 states – Maharashtra, Gujarat, Tamil Nadu, Delhi, Rajasthan, Madhya Pradesh, Uttar Pradesh, West Bengal, Andhra Pradesh and Punjab – accounted for 90.9% of the cumulative case count.

Two crucial points emerge from considering these snapshots. First, the concentration of the caseload among the top 10 states has remained relatively stable, at around 90% of the national case count, over this two-month period. Second, and more importantly, the membership of the top-10 states has changed gradually – even as Maharashtra, Delhi and Uttar Pradesh have continued to figure in the list at all three points. Appendix Figures 2 and 3 plot cumulative case and death counts, respectively, across states and over time to highlight the geographical pattern of the pandemic. (Our state/union territory-level graphs focus on the 20 states/union territories with the highest cumulative case counts as of May 18). Policy makers must take this geographical pattern into account while shaping effective responses. A perpetual national lockdown may not be necessary or feasible.

Case-fatality rates

We created two measures of the 'true' case-fatality rate (CFR). The first measure, denoted as CFR1, uses information on all the reported cases. CFR1 is defined as the ratio of cumulative number of deaths and cumulative number of reported cases till the day of interest. The second measure, which we denote as CFR2, only considers the cumulative number of closed, i.e., recovered or deceased cases. It is defined as the ratio of cumulative number of deaths and the sum of the cumulative number of deaths and recoveries. Since the numerator in the two measures is identical while the denominator in the first is smaller than in the second, CFR2 is always larger than CFR1. Most of the time the government and the media report CFR1.

Figure 6 Panel a (CFR1) and Appendix Figure 7 (CFR2) present forest plots of the estimated values of the two measures of the case fatality rate, along with a 95% confidence interval (CI), for the 20 states and union territories in our sample. To highlight differences across states, we have classified them into three groups: doing well, adequate, needs improvement. The three groups are color-coded red, gray and green, respectively.

Using CFR1, the members of the group that needs improvement (with CFR1 larger than 3%) – with relatively high case fatality rates – are West Bengal, Gujarat, Madhya Pradesh and Maharashtra; Tamil Nadu, Kerala, Bihar and Odisha belong to the group that is doing well; all other states fall in the adequate category. Using CFR2, the picture is grimmer. The groups needing improvement now has 10 states/union territories – West Bengal, Maharashtra, Gujarat, Madhya Pradesh, Karnataka, Chandigarh, Rajasthan, Uttar Pradesh, Delhi, Telangana and Andhra Pradesh – and the group that is doing well has only one member – Kerala. The other 9 states/union territories belong to the adequate group.

Measures of Growth: How far has the pandemic slowed down?

We will address this important question with evidence on two related, but different, measures of the pandemic's spread and growth.

Doubling Time and Reproduction Number

The first measure we use is the doubling time for total reported cases, computed with a 7-day backward-looking window. This measure gives the number of days it would take for total cases to double if its trajectory remained as observed in the past week. An increase in the doubling time is evidence of the pandemic slowing down.

The second, and related, measure we use is the *effective* basic reproduction number, R, which measures the average number of persons infected by an infected individual. Whereas the more-commonly used R_0, which is the basic reproduction number, is a constant and is inherent to the pathogen, the *effective* basic reproduction number, R, is time-variant and can be impacted by various public health interventions, including physical distancing and lockdown. When R falls below 1, the epidemic starts slowing down.

Figure 2 plots the estimated doubling times and Figure 3 plots the estimated values of the effective basic reproduction number, R, for COVID-19 in India. Since consistent and reliable estimates of doubling time and R requires many days of data, Figures 2 and 3 both start on March 15. In both graphs, we report the estimate (along with the 95% confidence interval for R) on three landmark dates: March 24, April 14, and May 3, corresponding to the initial lockdown and subsequent extensions, respectively.

The time series patterns of doubling time and estimated R at the national level show that the lockdown did slow down the spread of the pandemic. It took about 2 weeks for the doubling time to start moving up in a sustained manner. Since early April, the doubling time has increased from about 5 days to over 10 days at the middle of May (Figure 2). Turning to Figure 3, we see that the estimated value of R fell over the period of the first lockdown from **3.36 [95% CI: 3.30, 3.71]** on March 24 to **1.71 [95% CI: 1.66, 1.76]** on April 14, with significant fluctuation in between. Since then, the estimated R has fallen at a slower pace. The trailing 7-day average value of R for the week ending on **May 18 is 1.27 [95% CI: 1.25, 1.29]**.

The national pattern hides substantial state-level variations, and to see this we need to turn to Appendix Figures 4 and 5. Appendix Figure 4 has time series plots of doubling times at the state-level; Appendix Figure 5 summarizes information about estimated R for the 20 states and union territories in our sample. The plots in Appendix Figure 4 show that across most states doubling times have increased. Significant exceptions to this general pattern can be seen in Delhi, Karnataka, Odisha, and Tamil Nadu. In Appendix Figure 5, we see that, starting from relatively high values, estimates of R have generally fallen across all states. But there are significant differences across states – some continue to have high values, and in some low estimates of R have reverted to relatively high value.

Figure 6 Panels b and c present forest plots of the average value of the doubling time (with 7-day range) and estimated R (with 95% CI) over the week before May 18, respectively, and help us understand the state-level variation in a convenient way. Using the information in this figure, we can divide all states into three groups. With respect to doubling time, the first group consists of 14 states with estimated doubling times greater than two weeks (considered to be performing well). The second group contains 5 states with doubling times between one and two weeks. Odisha is the only state in the third group with an estimated doubling time less than one week. With respect to R, the first group consists of 3 well-performing states (Haryana, Andhra Pradesh and Punjab), where the average estimate of R lies below 1. The second group consists of 4 poor-performing states (Odisha, Tripura, Telangana, and Bihar), where the average estimate of R lies above 2. All the other states in our sample have average values of the estimated R falling within 1 and 2 (for some states, the CIs include one or both the extreme values).

Markers of Testing

Testing Coverage and Test Positivity Rate

As India gradually eases out of the lockdown, the focus of public policy intervention should be reoriented towards testing, tracing and isolating. A rigorously implemented strategy of testing suspected patients, tracing contacts of patients, and isolating infected persons can effectively break the chain of transmission and slow down the pandemic. In a poor country like India, which can ill afford the severe economic disruption caused by a lockdown, this alternative approach has much to recommend itself (Cash & Patel, 2020).

Such an approach crucially rests on adequate testing. While looking at testing coverage, i.e., persons tested per million population, is a common way to gauge adequacy of testing rates, an alternative approach is to track the test positive rate, i.e., fraction of positives in the total number of persons tested (Stein, Wroth, & Hurt, 2020). High and/or rising test positive rates provide indication of inadequate testing. It means that only those with a high probability of having the novel coronavirus that causes COVID-19 are being tested. These would predominantly be patients with strong symptoms of the disease. Hence, many more patients with lower probability of being infected and many of those without any symptoms (yet) would be out of the ambit of testing. Thus, the estimated prevalence rate of the disease would be a gross underestimate of the ‘true’ prevalence rate.

Going by national-level data, India seems to be doing pretty well in terms of testing. Since mid-April, India’s test-positive rate has fluctuated around 0.04 (see Figure 4). This is lower than not only many European and North American countries, but also significantly lower than its neighbors, like Bangladesh and Pakistan. But this national-level trend about testing in India is misleading because of the wide variation across states.

Appendix Figure 6 plots the test-positive rate over time for our sample of 20 states and union territories. The variation of test positive rates across states is obvious and striking. If we use a benchmark of 2% test positive rate to assess adequacy – this being the rate that Kerala has achieved for many weeks now – then we see that many states are still to achieve adequate testing.

What is even more worrisome is that many states, starting from a value that is higher than the benchmark, have witnessed rising test-positive rates. The problem of rising test-positive rates gets compounded by the fact that most of these states are where the pandemic is geographically concentrated. Delhi, Gujarat, Maharashtra and Tamil Nadu are four important examples – these states have high case counts and high/rising test-positive rates. Bihar, Chandigarh and Punjab, which have so far seen relatively low case counts, are witnessing rising test positive rates.

We quantify the magnitude of shortfall in COVID-19 testing across states in a simple way. Using a benchmark test-positive rate of 0.02 (which comes from Kerala’s experience), we define the adequate number of total tests as the ratio of observed test-positive rates (averaged over the week of May 12-May 18) and the target test-positive rate (in this case, 0.02) multiplied by the cumulative number of tests performed. The shortfall in testing is, then, quantified as the difference between the adequate number of tests and the actual number of daily tests (averaged over the past week). If the actual number of tests is larger than the 'adequate' number, then we note the shortfall as zero.

Figure 5 presents a bar chart of the magnitude of testing shortfall across the 20 states and union territories in our sample. Eleven states show a shortfall with Maharashtra, Gujarat, Tamil Nadu and Delhi leading the slate. Maharashtra is the clear outlier in terms of testing shortfall. It requires roughly an additional 1.25 million tests to come close to achieving the adequacy benchmark at 2% test positive rate. Maharashtra has done 282, 437 tests as of May 18. For Gujarat, Tamil Nadu and Delhi, the testing shortfall figure is approximately between 230 and 430 thousand tests. While Madhya Pradesh, West Bengal, Punjab, Uttar Pradesh, Rajasthan, Bihar, and Chandigarh have lower testing shortfalls, the other states and union territories in our sample display adequate testing so far. This is definitely a positive sign.

Summary state-level dashboard: Comprehensive Display of Metrics

With a complete data tsunami, different metrics telling us different features of the pandemic, and a rapidly evolving landscape, we offer a summary dashboard (Figure 6) where we have created classification groups and highlighted states according to various metrics. This captures, at a glimpse, a snapshot of where things stand across states and the nation. Daily updates are available in our app hosted at (COV-IND-19 Study Group, 2020)

In an effort to make these data more digestible, our 4-panel state-level dashboard (Figure 6) follows a consistent color-coding system to indicate states/union territories that are doing well (colored green) and states/union territories that need improvement (colored red).

Panel a shows the case-fatality rate along with the 95% confidence interval (doing well: below 1%; needs improvement: above 3%). While the all-India CFR on May 18 was 3.14%, state-level CFRs ranged from 8.6% (West Bengal) to 0.5% (Odisha). Four states had high levels of CFR, i.e., higher than the threshold value of 3%: West Bengal (8.6%), Gujarat (5.9%), Madhya Pradesh (4.8%) and Maharashtra (3.6%). Four good-performing states had low CFR: Tamil Nadu (0.7%), Kerala (0.6%), Bihar (0.6%) and Odisha (0.5%). The other 12 states fall in between.

Panel b shows the 7-day average doubling time along with the range (doing well: above 14; needs improvement: below 7). This metric is related to the effective reproduction number. The quickest doubling time is in Odisha (5.5 days, range: [4.3, 6.6]), while the slowest doubling time is in Kerala (51.4 days, range: [25.1, 113.3]). Kerala, again, is an example of how R is more sensitive to a recent increase in daily cases, while the doubling time, being a function of cumulative cases, is relatively less sensitive to daily movements. This is why Kerala's recent R estimate is around 2, even as it has the longest doubling time. The national estimate is 13.1 days (range: [11.9, 13.9]), with most states having doubling time exceeding 14 days.

Panel c shows the 7-day average effective basic reproduction estimate (R) along with the 95% confidence interval (doing well: below 1; needs improvement: above 2). We see that 7-day average estimates range from 0.33 (95% CI: [0.29, 0.39]) in Punjab to 2.53 (95% CI: [2.26, 2.84]) in Odisha, with the national estimate being 1.24 (95% CI: [1.22, 1.26]). Most states/union territories are between 1 and 2, suggesting that smaller outbreaks are being seen across the board. It is worth noting, as we have pointed out above, that this metric is sensitive to the number of daily cases being reported. For example, Kerala has done well controlling the outbreak, but a small recent increase in observed cases results in a 7-day average estimate of 1.99 (95% CI: [1.53, 2.53]).

Panel d shows the 7-day average test-positive rate along with the range (doing well: below 3%; needs improvement: above 6%). The lowest 7-day average test-positive rate is seen in Jharkhand (0.69%, range: [0.64%, 0.72%]), with the highest being seen in Maharashtra (10.98%, range: [10.52%, 11.70%]). Generally, we see that states with larger cumulative case counts have higher test-positive rates. Considering the test shortfall (Figure 5), this is likely explained by relatively fewer number of tests being performed given the scale of the outbreak in these places. The national 7-day test-positive rate is 3.82% (range: [3.73%, 4.10%]).

It is important to consider these metrics together and keep their nuances in mind:

- CFR is an indicator of the fatality associated with the epidemic, but its value is sensitive to the number of tests being performed. A high CFR might very well arise from inadequate testing. Hence, the CFR is best used in conjunction with some measure of adequate testing.
- Effective reproduction number can indicate a recent outbreak but is sensitive to the level of daily cases being observed (i.e., a state/union territory with few cases can have a high R).
- Doubling time is a longer-term measure since it is a function of cumulative cases (i.e., this metric is more robust to increases in recent daily cases).
- Test-positive rate is both a function of the size of the outbreak in an area and the number of tests being performed. A higher test-positive rate can indicate insufficient levels of testing.

4. Discussion and conclusion:

While it is common for analysts and policymakers to predict a peak for the COVID-19 in India (Belfin, Brodka, Radhakrishnan, & Rejula, 2020), our analysis shows that the concept of a peak for the whole country is at best an ambiguous concept. Differences in estimates of R (Appendix Figure 5) and estimated doubling times (Appendix Figure 4) suggest that peaks will vary across states. Predictions from the eSIR model (Wang, et al., 2020) show that peaks might start as early as the end of July in some states and go all the way to September in many others.

These predictions are in line with basic intuition about the dynamics of the pandemic in India. Initial cases were imported, and the initial growth was limited to a few states which saw the arrival of international travelers. These initial cases seeded the epidemic and saw the initial explosion of cases. With the lockdown, mobility was limited at the macro level (inter-state, inter-city), which reduced transmission rates (Figures 2-3, Appendix Figures 4-5). Pre-lockdown infections and micro-mobility showed up in growth of cases within states – note that the top 10 states on April 18 and May 18 are largely same (other than West Bengal and Punjab). As we approach the partial or complete end of the lockdown, internal migration will start playing an increasingly important role in the spread of the pandemic.

India has a large migrant worker population. Estimates of out-of-state and out-of-district migrants ranged from 60 million to 80 million in 2011 (Economic Survey 2016-17, Chapter 12). Average work-related migration flows between states over the period 2011-16 was about 9 million per year (Economic Survey 2016-17, Chapter 12). Moreover, there is an interesting pattern of migration flows. States in the North and East (other than Delhi) see net migration outflows, because they are relatively poorer; states in the West and South, being relatively richer, witness net migration inflows (Economic Survey 2016-17, Chapter 12, Table 2). This is important for our analysis because the initial spurt in COVID-19 cases has largely been confined to the richer states. As the lockdown is eased, a large migrant population will travel back to their homes in the East and the North. Unless effective testing, tracing and isolating protocols are in place, India could very well see the next surge in cases in the poorer Eastern and Northern states.

Given this spatial and temporal pattern of the pandemic's spread, it is extremely important to prioritize policies. Resources must be mobilized to help one cluster of states and then move to the next cluster. It might be useful for the central government and Indian Council of Medical Research (ICMR) to classify states in terms of the phases of the epidemic. Even as the worst-hit states are being addressed, the next set could be put on high alert. It is this dynamic policy intervention that will be required to deal effectively with the cascading pattern of the pandemic across Indian states.

In implementing such a dynamic policy, it is important to facilitate replication of successful strategies across states. Kerala's rapid response in terms of testing, contact tracing and quarantining; Odisha and Kerala's use of local governance structures and community health networks for surveillance and dissemination of correct information; Punjab's use of data analytics and district level granular contact tracing, tracking and isolation – all these experiences will be of use in other states that are likely to see a surge in cases in the coming weeks. The success of some states gives us hope that there are strategies to beat the virus that have worked well. Resources can be mobilized and optimally deployed to address the acute situations in high density population areas like Maharashtra, Gujarat and Delhi.

5. Supplementary Methods:

Estimation and confidence interval (CI) for case fatality rates

Let us denote the cumulative number of confirmed cases and deaths for a region of interest (India or one of the states/union territories) at a given date (May 18 for our purpose) respectively by C and D. Assuming that the proportion of underreporting (due to impossibility of testing all cases and imperfection of the tests) in the fatal and non-fatal cases are same, $D|C \sim Bin(C, \pi)$ where π is the true underlying case fatality ratio. Therefore, assuming sufficiently large number of cases, via central limit theorem, we can write $\sqrt{C}(\hat{\pi}-\pi) \sim AN\big(0, \pi(1-\pi)\big)$, where $\hat{\pi} = \frac{D}{C}$. Using delta method on this, we get $\sqrt{C}\big(logit(\hat{\pi}) - logit(\pi)\big) \sim AN\left(0, \frac{1}{\pi(1-\pi)}\right)$, where $logit(x) = \log\left(\frac{x}{1-x}\right)$. Therefore, one estimator the standard deviation of $\hat{\pi}$ is given by $s = \sqrt{\frac{1}{C\hat{\pi}(1-\hat{\pi})}} = \sqrt{\frac{C}{D(C-D)}}$. Using this, we can get a 95% CI for $logit(\pi)$ as $(logit(\hat{\pi}) \pm z_{0.975}s)$. Inverting this by applying the function $expit(x) = \frac{e^x}{1+e^x}$, we get a 95% CI for π. Figure 6 Panel a summarizes these estimates and CIs across states/union territories.

It is important to note that this method inherently assumes that all events (deaths/recoveries) that could possibly happen from the set of observed confirmed cases has happened by the day on which the data is observed, which of course is not true in general. One standard alternative approach here is to look at the closed cases only. Assume that the cumulative number of recovered cases at the same date for the same region as before is denoted by R. Then, using $D + R$ in place of C in the above calculations throughout, we can get another estimate and CI for the true case fatality rate. Appendix Figure 7 summarizes these estimates and CIs across states/union territories.

Doubling time

We calculate doubling time, T_d, assuming a constant growth rate r% within time t using the formula

$$T_d = t\frac{\ln(2)}{\ln(1+r)},$$

where r is calculated as

$$r = \frac{T_{end} - T_{start}}{T_{start}}.$$

We calculated the doubling time using a trailing 7-day window, i.e., the doubling time for May 7 represents how long cases would take to double assuming a constant growth in cases from May 1 to May 7.

Time-varying R estimates

We estimate the effective basic reproduction number for COVID-19 in India using the EpiEstim package in R (Cori, Ferguson, Fraser, & Cauchemez, 2013) and data from (COVID-19 India, 2020), a crowdsourced effort that relies on volunteer validation of state bulletins and official handle reports. We refer to the effective basic reproduction number as "R" throughout, which is similar to the concept of R_0, however, R_0 is a constant that is inherent to the pathogen and is not time-variant or impacted by interventions (such as social distancing or lockdown).

We use the "parametric_SI" estimation method and a 5-day window, described by (Cori, Ferguson, Fraser, & Cauchemez, 2013) ("estimate_R" function, which was used to describe the progression of the outbreak in Wuhan (Pan, et al., 2020)). We also use a gamma distribution with a mean of 7 days and a standard deviation of 4.5 days, based on research by (Wu, et al., 2020), for the generation time (a distribution of the onset of disease used to estimate R).

We looked at the effective basic reproduction number for COVID-19 nationwide in India using data from March 1 to May 18. Because the estimation requires several days of data for reliable, consistent results, we only observe data from March 15 to May 18. We also estimated R over the time period for the 20 states/union territories with the greatest number of total reported cases as of May 18 according to (COVID-19 India, 2020). State-level data was first reported by (COVID-19 India, 2020) on March 15 and we begin the plots on March 24 to allow the estimates to stabilize. There are some states/union territories for which the first cases were not reported until after March 24 (e.g., Tripura), in which case we see the initial elevated R estimates because the estimates have not yet stabilized.

We see that the estimated R varies across states/union territories and, in some cases, does drop below one (indicated by the dashed horizontal lines in Figure 3 and Appendix Figure 5). It is worth noting that in several of these cases, it returns to above 1 after it drops below 1, highlighting that, despite time-varying estimates, no state/union territory is in the clear yet. The plots report the average R and 95% CIs *for the past 7 days* corresponding to the highlighted state/union territory.

Test positive rate

The test positive rate was calculated as the ratio of cumulative reported number of positive tests to the reported total number of tests on a given date ((COVID-19 India, 2020) state-level testing data begins April 1). While (COVID-19 India, 2020) also has national-level testing data, it is spotty, and in recent weeks, have not been reporting the number of positive tests. As such, for national test-positive rates, we sum the positive tests and total tests over all the 35 states and union territories for which data were reported for national counts and rates. It will be to acquire consistency across data sources on the testing data.

Testing shortfall

The testing shortfall is a metric used to estimate the increase in the number of tests that should be seen relative to a 2% benchmark test positive rate. First, we calculate the desired number of tests, T_D:

$$T_D = \frac{TPR_O}{TPR_D} T_O$$

Where TPR_0 is the 7-day average of the observed, cumulative test-positive rate, TPR_D is the target test-positive rate (in this case $TPR_D = 0.02$, and T_O is the observed number of cumulative tests.

With this value, we calculate the shortfall, or the number of additional total tests required to achieve the test-positive rate as:

$$shortfall = \max(T_D - T_O, 0)$$

When shortfall is equal to 0, the number of tests being performed is theoretically sufficient given the number of cases being observed. When shortfall is greater than 0, it represents the number of additional tests that should be performed given the number of cases being observed.

References:

Belfin, R. V., Brodka, P., Radhakrishnan, B. L., & Rejula, V. (2020). COVID-19 peak estimation and effect of nationwide lockdown in India. *medRxiv*. https://www.medrxiv.org/content/10.1101/2020.05.09.20095919v1

Cash, R., & Patel, V. (2020). Has COVID-19 subverted global health? *The Lancet*. https://doi.org/10.1016/S0140-6736(20)31089-8

Cori, A., Ferguson, N. M., Fraser, C., & Cauchemez, S. (2013, November). A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. *American Journal of Epidemiology, 178*(9), 1505-1512. https://doi.org/10.1093/aje/kwt133

COVID-19 India. (2020, May). Retrieved from Coronavirus Outbreak in India: https://www.covid19india.org/

COVIDIndia.org. (2020, May 18). Retrieved from COVID-19 Tracker Updates for India for State-wise and District-wise data: https://covidindia.org/

COV-IND-19 Study Group. (2020, May 19). Retrieved from COVID-19 Outbreak in India: https://umich-biostatistics.shinyapps.io/covid19/

Economic Survey 2016-17, Chapter 12. Government of India, Ministry of Finance. Retrieved from Union Budget. https://www.indiabudget.gov.in/budget2017-2018/es2016-17/echap12.pdf

Ghosh, P., Ghosh, R., & Chakraborty, B. (2020). *COVID-19 in India: State-wise Analysis and Prediction.* medRxiv. https://www.medrxiv.org/content/10.1101/2020.04.24.20077792v3

Gupta, M., Mohanta, S. S., Rao, A., Parameswaran, G., Agarwal, M., Arora, M., . . . Bhatnagar, S. (2020). *Transmission dynamics of the COVID-19 epidemic in India, and evaluating the impact of asymptomatic carriers and role of expanded testing in the lockdown exit strategy: a modelling approach.* medRxiv. https://www.medrxiv.org/content/10.1101/2020.05.13.20096826v2

Hui, D. S., Azhar, E. I., Madani, T. A., Ntoumi, F., Kock, R., Dar, O., . . . Petersen, E. (2020). The continuing 2019-nCoV epidemicthreatof novel coronaviruses to global health: the latest 2019 novel coronavirus outbreak in Wuhan, China. *International Journal of Infectious Diseases*, 264-266. https://doi.org/10.1016/j.ijid.2020.01.009

Jakhar, M., Ahluwalia, P. K., & Kumar, A. (2020). *COVID-19 Epidemic Forecast in Different States of India using SIR Model.* medRxiv. https://www.medrxiv.org/content/10.1101/2020.05.14.20101725v1

Jewell, N. P., Lewnard, J. A., & Jewell, B. L. (2020). Jewell, N. P., Lewnard, J. A., & Jewell, B. L. (2020). Predictive mathematical models of the COVID-19 pandemic: underlying principles and value of projections. *Journal of the American Medical Association, 323*(19), 1893-1894. https://doi.org/10.1001/jama.2020.6585

Johns Hopkins University. (2020). *COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)*. Retrieved from Coronavirus Resource Center: https://coronavirus.jhu.edu/map.html

Mayo Foundation for Medical Education and Research. (2020). *Coronavirus disease 2019 (COVID-19)—Symptoms and causes*. Retrieved May 2020, from Mayo Clinic COVID-19 Response: https://www.mayoclinic.org/diseases-conditions/coronavirus/symptoms-causes/syc-20479963

Mehrotra, K. (2020, May 8). *Explained: India enforced one of the strongest lockdowns, here's how it stacks up against other countries.* Retrieved from The Indian Express: https://indianexpress.com/article/explained/coronavirus-india-lockdown-vs-global-lockdown-covid-19-deaths-cases-cure-6399181/

Mitra, A., Pakhare, A. P., Roy, A., & Joshi, A. (2020). *Impact of COVID-19 epidemic curtailment strategies in selected Indian states: an analysis by reproduction number and doubling time with incidence modelling.* medRxiv. https://www.medrxiv.org/content/10.1101/2020.05.10.20094946v1

Narain, N. (2020, April 24). *Month After Junta Curfew, a Look at How Successful India Has Been Post 'World's Most Stringent' Covid-19 Lockdown.* Retrieved from News 18 India: https://www.news18.com/news/india/month-after-junta-curfew-a-look-at-how-successful-india-has-been-post-worlds-most-stringent-covid-19-lockdown-2589017.html

Pan, A., Liu, L., Wang, C., Guo, H., Hao, X., Wang, Q., . . . Wu, T. (2020). Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China. *JAMA, 323*(19), 1915-1923. http://doi.org/10.1001/jama.2020.6130

Pradhan, B. (2020, May 19). *India's Virus Cases Cross 100,000, Jump at Fastest Pace in Asia*. Retrieved from Bloomberg: https://www.bloomberg.com/news/articles/2020-05-19/india-s-virus-cases-cross-100-000-jump-at-fastest-pace-in-asia?srnd=coronavirus

Rajendrakumar, A. L., Nair, A. T., Nangia, C., Chourasia, P. K., Chourasia, M. K., Ghouse, S. M., . . . Koya, S. F. (2020). *Epidemic Landscape and Forecasting of SARS-CoV-2 in India.* medRxiv. https://www.medrxiv.org/content/10.1101/2020.04.14.20065151v1

Ray, D., Salvatore, M., Bhattacharyya, R., Wang, L., Mohammed, S., Purkayastha, S., . . .Mukherjee, B.. (2020, May 14). *Predictions, role of interventions and effects of a historic national lockdown in India's response to the COVID-19 pandemic: data science call to arms.* Retrieved from Harvard Data Science Review: https://hdsr.mitpress.mit.edu/pub/r1qq01kw

Roser, M., Ritchie, H., Ortiz-Ospina, E., & Hasell, J. (2020, May 18). *Coronavirus Pandemic (COVID-19).* Retrieved from Our World in Data: https://ourworldindata.org/coronavirus

Rukmini S. (2020, May 6). *Covid-19 lockdown, successful or not? There are 5 ways to answer this question.* Retrieved from The Print: https://theprint.in/opinion/covid-19-lockdown-successful-or-not-there-are-5-ways-to-answer-this-question/415049/

Stein, R., Wroth, C., & Hurt, A. (2020, May 7). *U.S. Coronavirus Testing Still Falls Short. How's Your State Doing?* Retrieved from npr: https://www.npr.org/sections/health-shots/2020/05/07/851610771/u-s-coronavirus-testing-still-falls-short-hows-your-state-doing

Venkatesan, P. (2020). *National and state wise estimate of time varying reproduction number for COVID-19 in India during the nationwide lockdown.* medRxiv. https://www.medrxiv.org/content/10.1101/2020.05.01.20087197v1

Wang, L., Zhou, Y., He, J., Zhu, B., Wang, F., Tang, L., . . . Song, P. X. (2020). An epidemiological forecast model and software assessing interventions on COVID-19 epidemic in China. *medRxiv*. https://www.medrxiv.org/content/10.1101/2020.02.29.20029421v1

Wu, J. T., Leung, K., Bushman, M., Kishore, N., Niehus, R., de Salazar, P. M., . . . Leung, G. (2020). Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nature Medicine, 26*, 506-510. https://doi.org/10.1038/s41591-020-0822-7

Figures:

Figure 1. Daily number of reported cases, fatalities, and recovered cases in India over the period between March 15 and May 18 with four states showing the variation. Kerala and Punjab are example states of "doing well" whereas case-counts in Delhi and Maharashtra are still increasing.

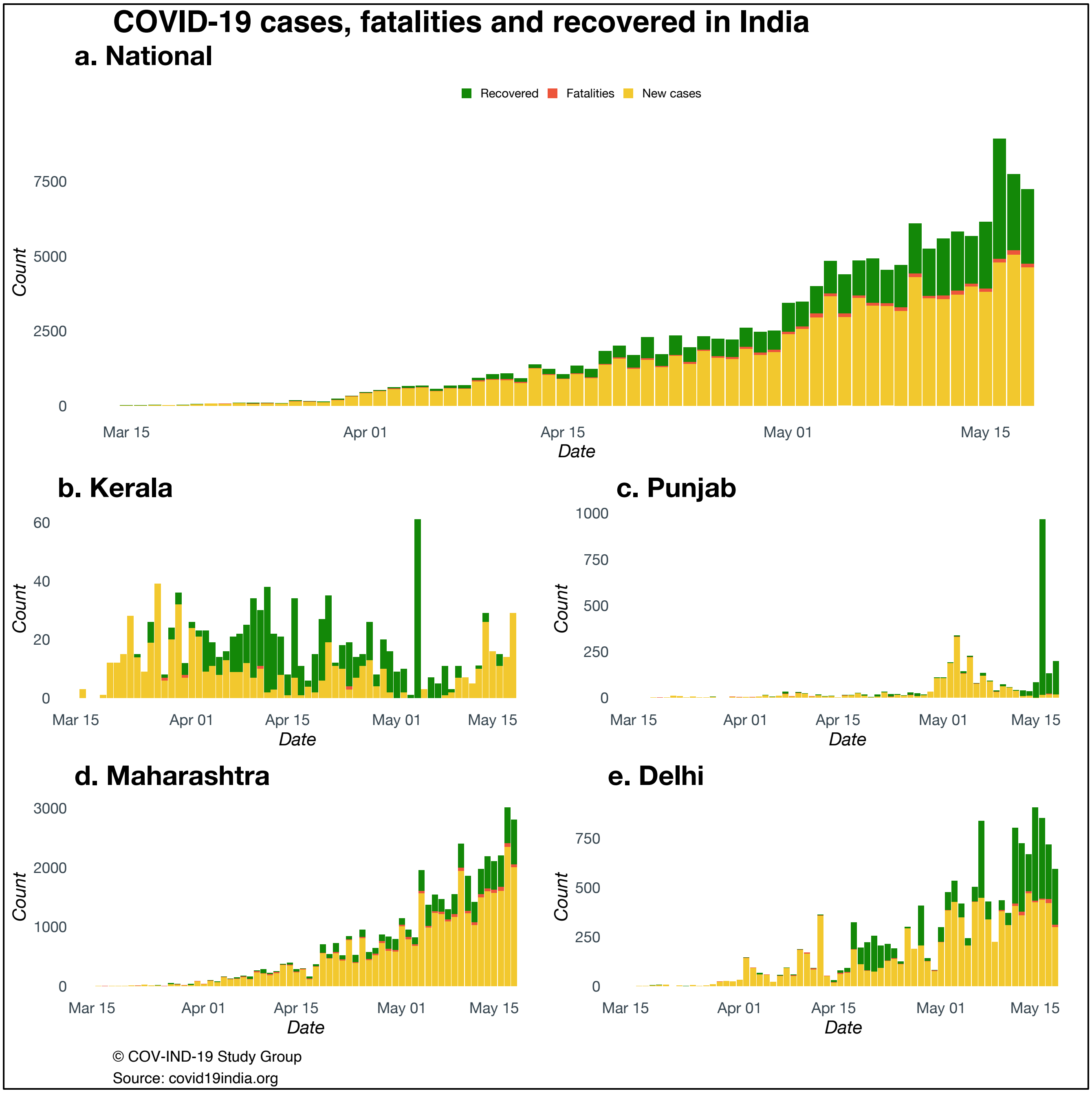


Figure 2. Estimated doubling times of total number COVID-19 cases in India, with averages for the pre- and post-lockdown periods and past 7-day average as of May 18.

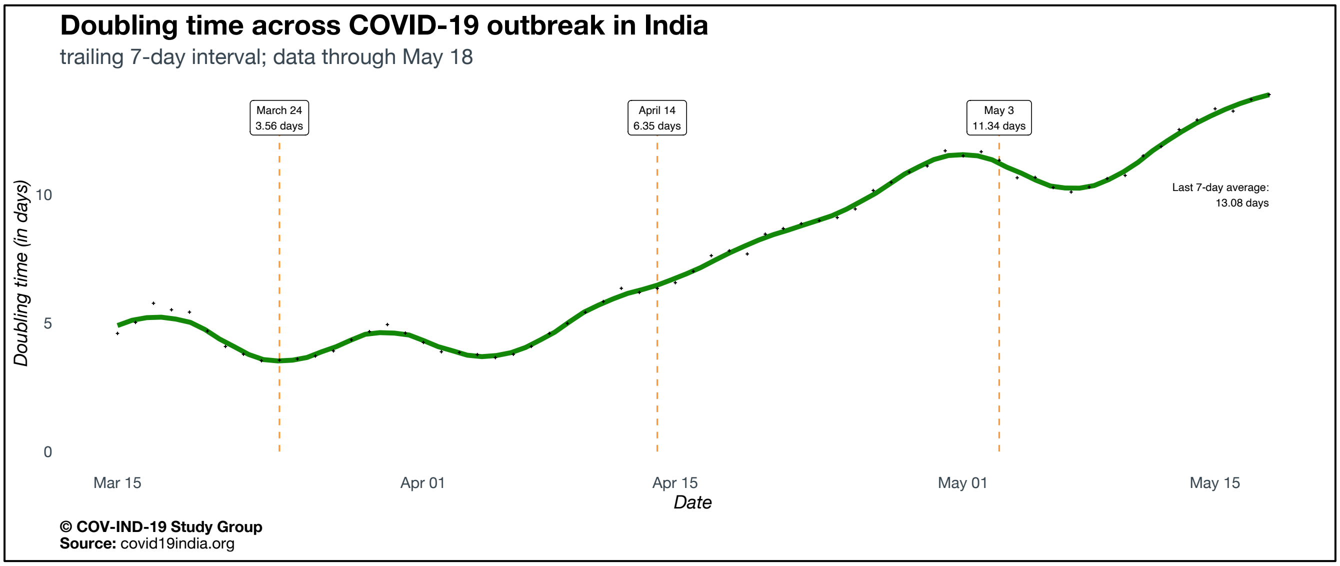


Figure 3. Estimated time-varying R (effective basic reproduction number) for COVID-19 in India with averages for the pre- and post-lockdown periods and past 7-day average as of May 18, along with 95% confidence intervals.

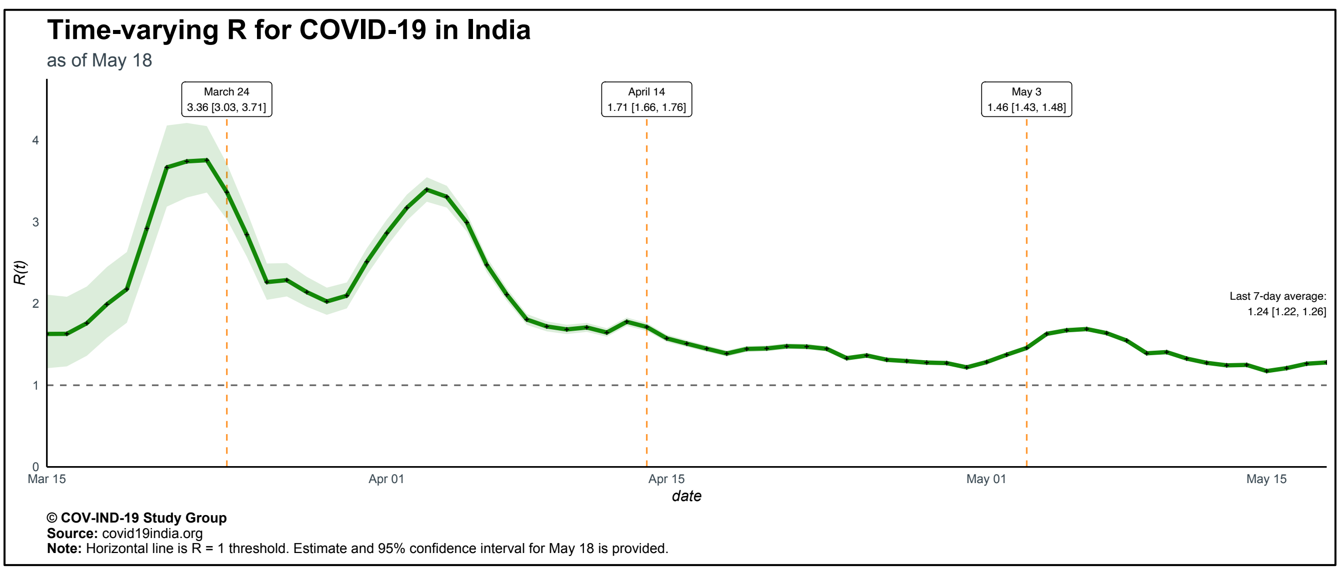


Figure 4. Time series plot of test positive rates for India over the period between April 1 and May 18.

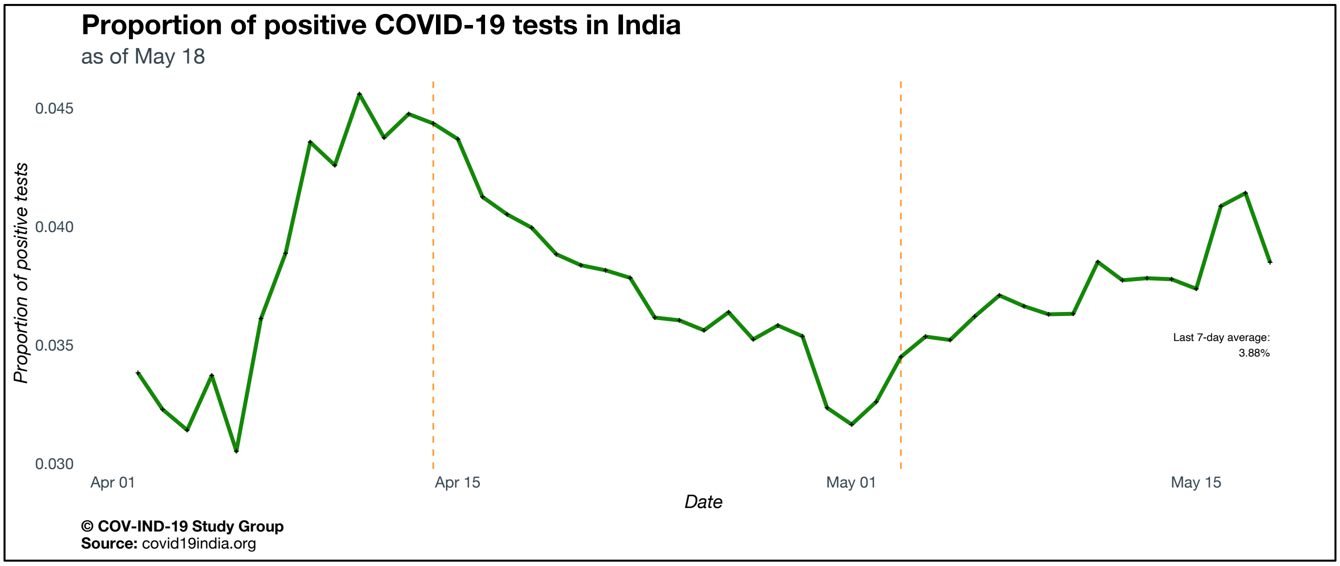


Figure 5. Shortfall of number of tests across 20 Indian states and union territories, relative to a benchmark test positive rate of 2%. Based on testing data up to May 18.

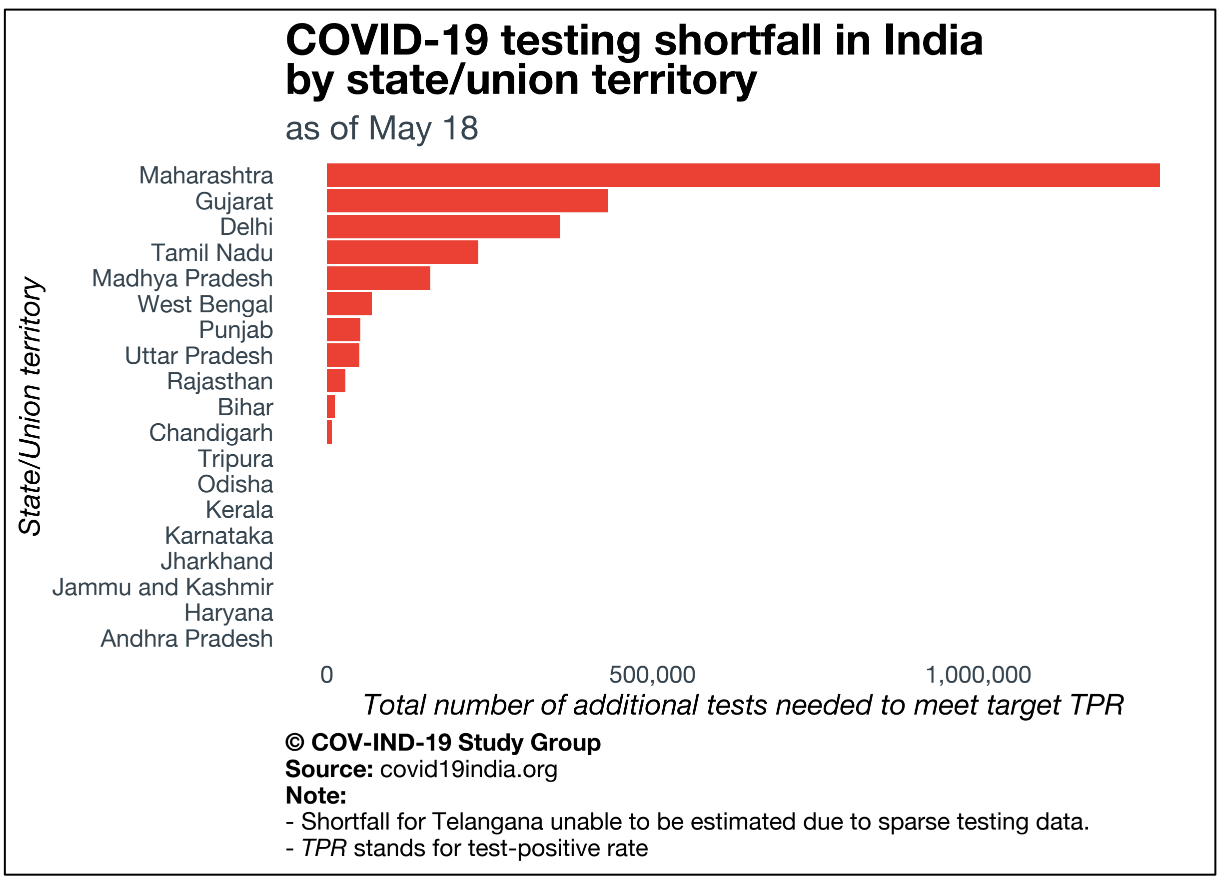


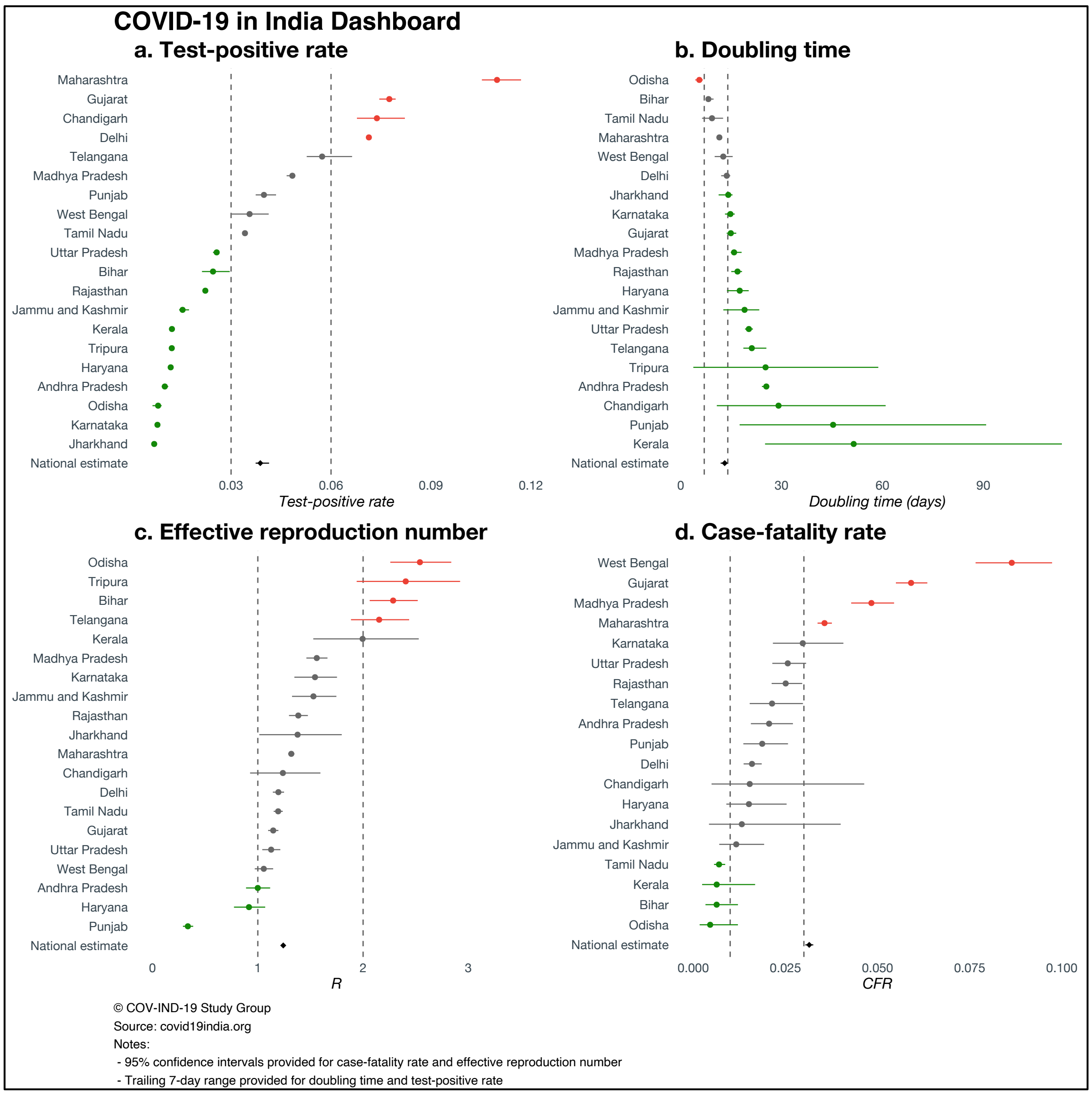
Figure 6. Forest Plot Dashboard

Panel a. Forest plot of estimated case fatality rates based on all confirmed cases as of May 18, along with 95% confidence intervals, for 20 states and union territories of India, and a national summary.

Panel b. Forest plot of estimated doubling times (in days) based on data from a 7-day past window from May 18, along with 95% confidence intervals, for 20 states and union territories of India, and a national summary.

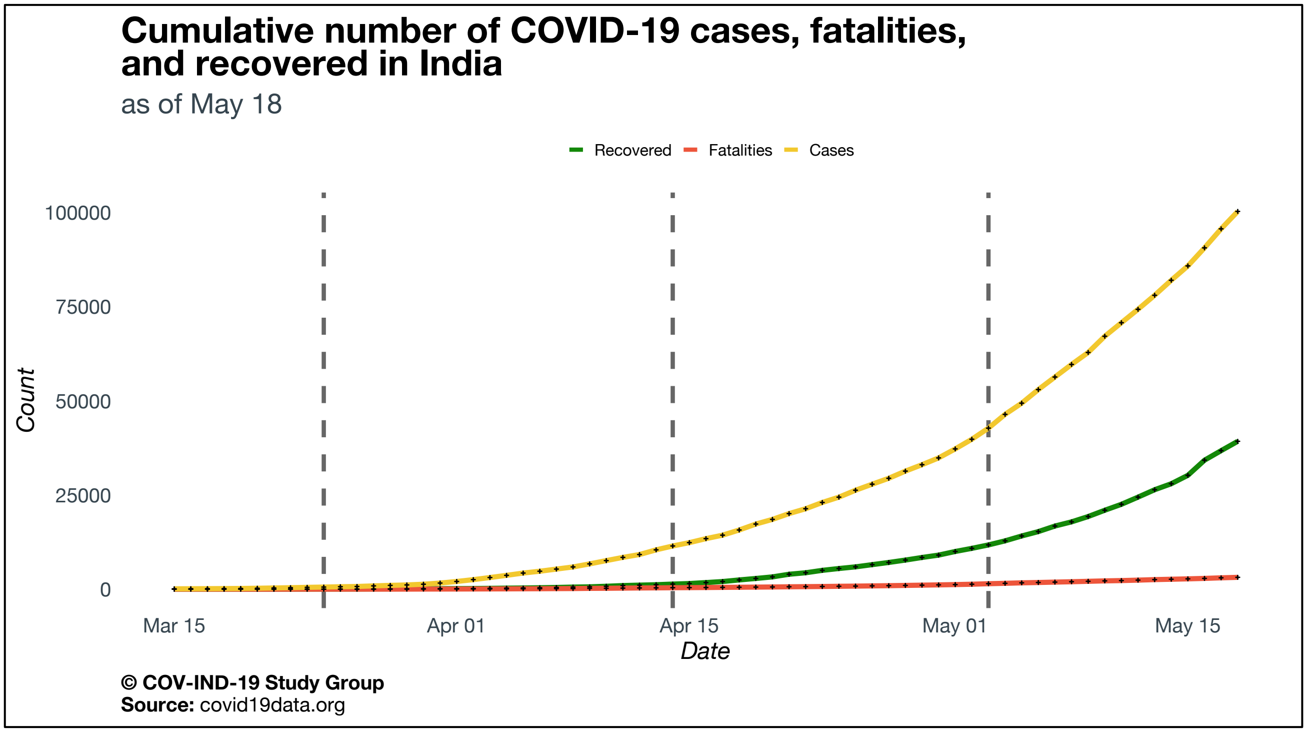
Panel c. Forest plot of estimated time-varying R (effective basic reproduction number) based on data from a 7-day past window from May 18, along with 95% confidence intervals, for 20 states and union territories of India, and a national summary.

Panel d. Forest plot of test positive rates (proportion scale) based on data as of May 18, for 20 states and union territories of India, along with a national summary.

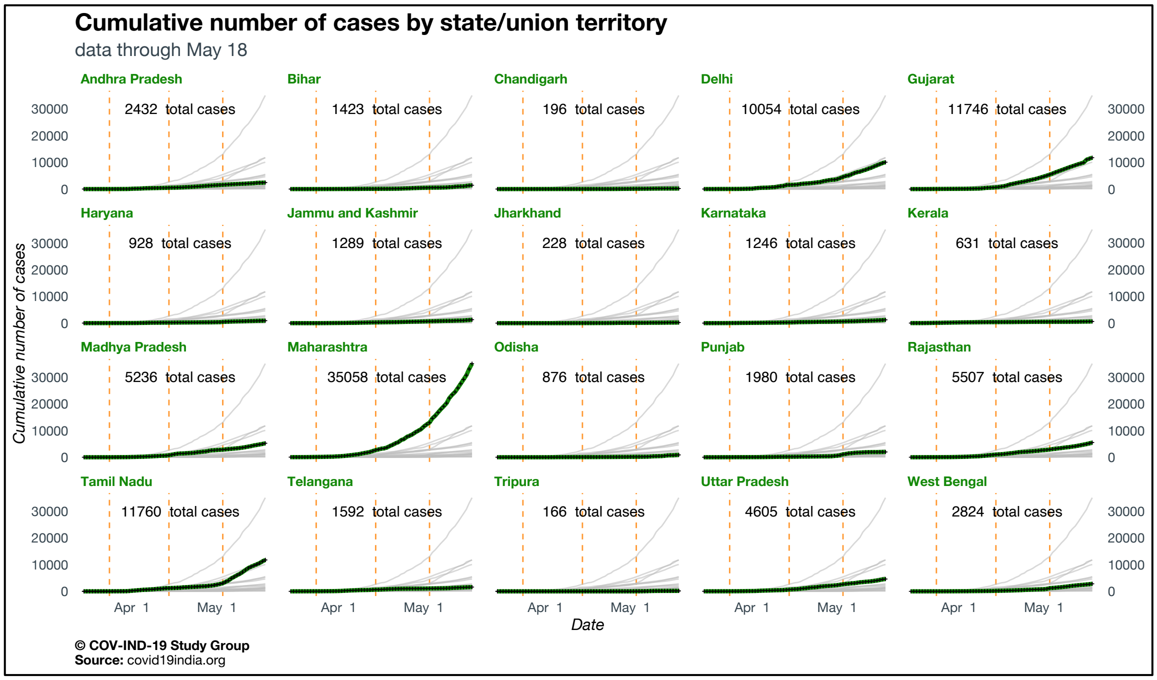


Appendix:

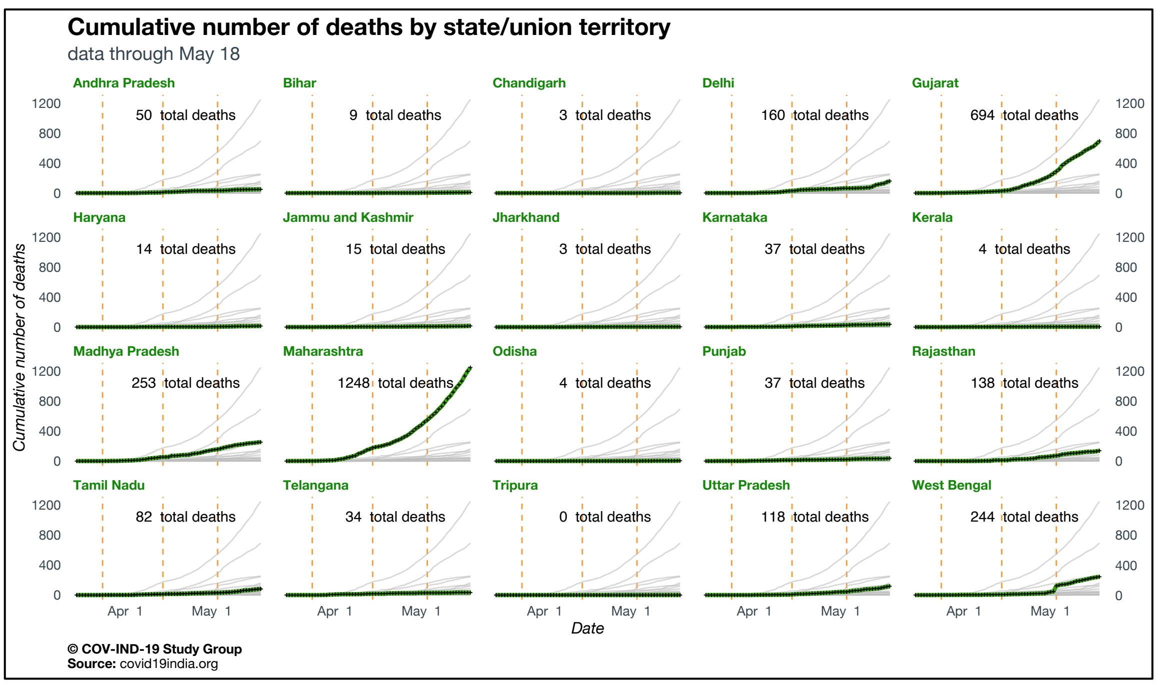
Appendix Figure 1. Cumulative number of reported cases, fatalities, and recovered cases in India over the period between March 15 and May 18.



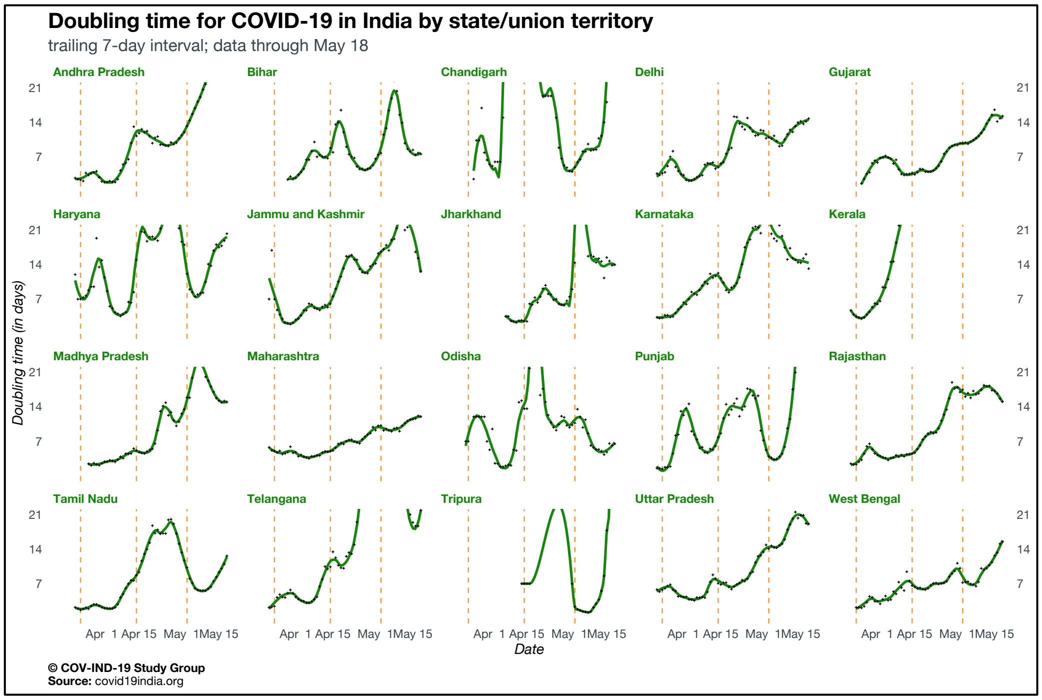
Appendix Figure 2. Cumulative number of reported COVID-19 cases in 20 Indian states and union territories over the period between March 15 and May 18.



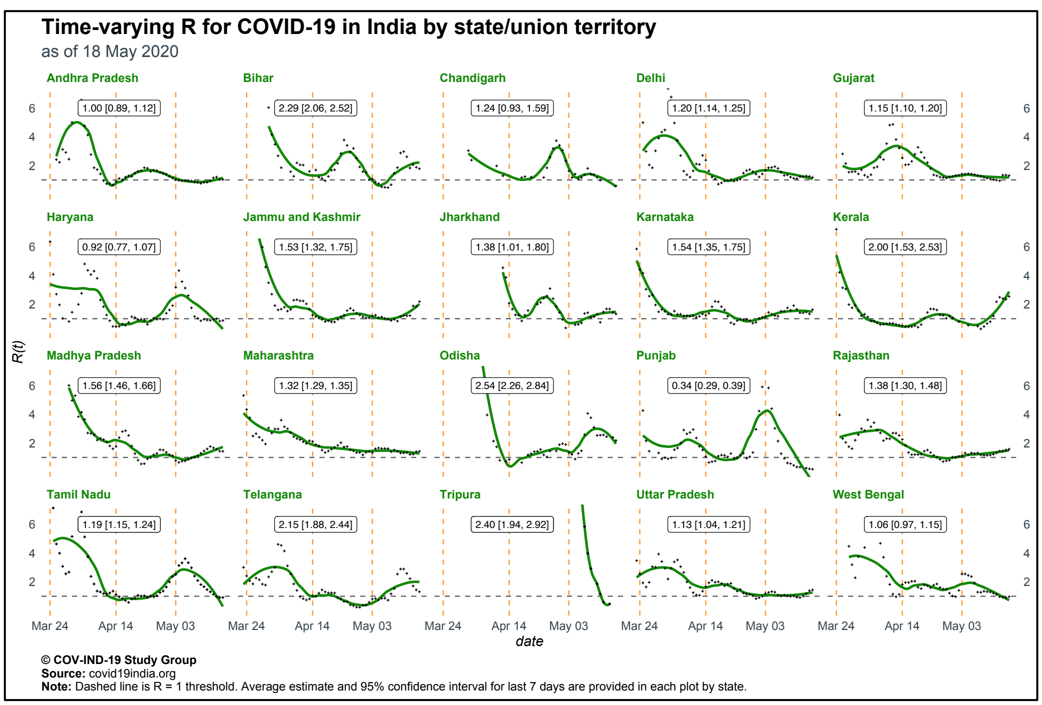
Appendix Figure 3. Cumulative number of reported COVID-19 deaths in 20 Indian states and union territories over the period between March 15 and May 18.



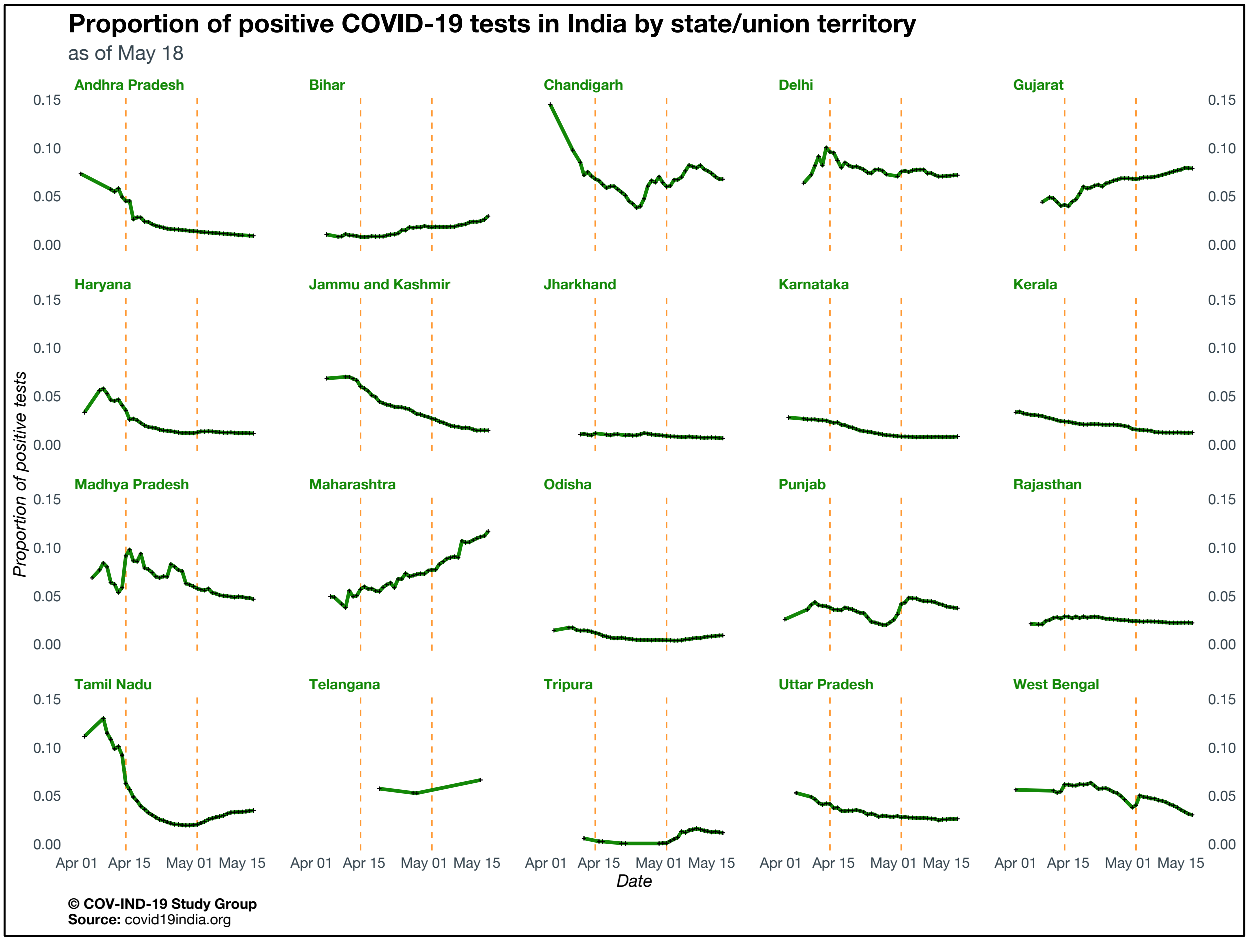
Appendix Figure 4. Estimated doubling times of total number COVID-19 cases in 20 Indian states and union territories.



Appendix Figure 5. Estimated time-varying R (effective basic reproduction number) for COVID-19 in 20 Indian states and union territories along with 95% confidence intervals.



Appendix Figure 6. Time series plots of test positive rates for 20 Indian states and union territories.



Appendix Figure 7. Forest plot of estimated case fatality rates based on closed cases only as of May 18, along with 95% confidence intervals, for 20 states and union territories of India, and a national summary.