Supplementary information for

A Comprehensive Public Health Evaluation of Lockdown as a Non-pharmaceutical Intervention on COVID-19 Spread in India: National Trends Masking State Level Variations

Maxwell Salvatore[1,2], MPH; Deepankar Basu[3], PhD; Debashree Ray[4,5], PhD; Mike Kleinsasser[1], MS; Soumik Purkayastha[1], MS; Rupam Bhattacharyya[1], MS; Bhramar Mukherjee[1,2]*, PhD.

[1]: Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA

[2]: Center for Precision Health Data Science, University of Michigan, Ann Arbor, MI 48109, USA

[3]: Department of Economics, University of Massachusetts, Amherst, MA 01002, USA

[4]: Department of Epidemiology, Johns Hopkins University, Baltimore, MD 21205, USA

[5]: Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205, USA

*: Corresponding author. Address - Department of Biostatistics, School of Public Health, University of Michigan, 1420 Washington Heights, Ann Arbor, MI 48109-2029, USA. Telephone – (734) 764-6544. Email – bhramar@umich.edu.

Table of Contents

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 31 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.

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 (CFR2, ratio of the total number of deaths and the sum of the same and the total number of recovered cases).

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.

Using Log-linear models to calculate doubling time: Assuming a constant growth rate of $r\%$ within time t, the doubling time T_d can be calculated using the formula $T_d = t\frac{\ln(2)}{\ln(1+r)}$. For the estimation of r and the subsequent computation of the doubling time, we use the 'fit' function from the R package 'incidence'[1***] which, given a vector of daily incidence of cases $\boldsymbol{y}$, fits a log-linear model of the form $\log(y_t) = rt + b$ where t is the time coded in days and b is the intercept at origin. Based on this fitted model and the corresponding estimate and 95% confidence interval (CI) for r, T_d (and a 95% CI for it) can then be calculated by using the transformation mentioned above. The estimate (95% CI) can directly be obtained by calling the object 'Modelname$info$doubling' ('Modelname$info$doubling.conf') in R, where 'Modelname' is the name of the fitted model object using the 'fit' function. We calculated the doubling time for the nation and the states of interest using incidence data during the dates from March 15 to May 31.

Time-Varying R Estimates

We estimate the effective reproduction number for COVID-19 in India using the EpiEstim package in R and data from COVID-19 India, a crowdsourced effort that relies on volunteer validation of state bulletins and official handle reports.[2,3] We refer to the effective reproduction number as "R" throughout, which is similar to the concept of R_0, however, R_0 assumes a fully susceptible population and is time-invariant. This instantaneous R is recommended for evaluating effective control measures.

We use the "parametric_SI" estimation method and a 5-day window ("estimate_R" function, which was used to describe the progression of the outbreak in Wuhan).[2,4] We also use a gamma distribution prior with a mean of 7 days and a standard deviation of 4.5 days, based on research by Wu and colleagues, for the generation time (a distribution of the onset of disease used to estimate R).[5]

We looked at the effective reproduction number for COVID-19 nationwide in India using data from March 1 to May 31. Because the estimation requires several days of data for reliable, consistent results, we only observe data from March 15 to May 31. 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 31. State-level data was first reported by COVID-19 India on March 15 and we begin the plots on March 24 to allow the estimates to stabilize.[3] 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 Figures 3 and 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 Positivity Rate

The test positivity 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 state-level testing data begins April 1).[2] While COVID-19 India 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

This metric is only relevant when the pandemic is in a control phase with steady decline in TPR and effective R for an extended period, say, 14 consecutive days. This metric is mostly for surveillance when the community prevalence is low. 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 positivity 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.

Extended SIR (eSIR) Model Predictions

Overview: The national and state-wide forecasts as available on the R Shiny dashboard at covind19.org are computed using an extension of the standard Susceptible-Infected-Removed (SIR) model, called the extended SIR (eSIR) model.[6] When using the eSIR model with time-varying disease transmission rate, it can depict a series of time-varying changes caused by either external variation like government-initiated macro isolation measures, community-level protective measures and environment changes, or internal variations like mutations and evolutions of the pathogen. To implement the eSIR model, a Bayesian hierarchical framework is assumed. Using the current time series data on the proportions of

infected and the removed people, a Markov chain Monte Carlo (MCMC) implementation of this Bayesian model provides not only posterior estimation of parameters and prevalence of all the three compartments in the SIR model, but also predicted proportions of the infected and the removed people at future time points. The R package for implementing this general model for understanding disease dynamics is publicly available at https://github.com/lilywang1988/eSIR. The next few subsections describe the parameter specifications used for the predictions. All the specifications are summarized in Parameter Table at the end of the Supplementary Methods.

Mathematical framework of the eSIR model: The eSIR model works by assuming that the true underlying probabilities of the three compartments follow a latent Markov transition process, and that we only observe the daily proportions of infected cases and removed. First, let us set up some notations. Assume that the observed proportions of infected and removed cases on day t are denoted by Y_t^I and Y_t^R, respectively. Further, denote the true underlying probabilities of the S, I, and R compartments on day t by θ_t^S, θ_t^I, and θ_t^R, respectively, and assume that for any t, $\theta_t^S + \theta_t^I + \theta_t^R = 1$. Assuming a usual SIR model on the true proportions (Supplementary Figure 6), we have the following set of differential equations:

$$\frac{d\theta_t^S}{dt} = -\beta\theta_t^S\theta_t^I,$$

$$\frac{d\theta_t^I}{dt} = \beta\theta_t^S\theta_t^I - \gamma\theta_t^I,$$

$$\frac{d\theta_t^R}{dt} = \gamma\theta_t^I$$

Here, $\beta > 0$ denotes the disease transmission rate, and $\gamma > 0$ denotes the removal rate. The basic reproduction number $R_0 \coloneqq \frac{\beta}{\gamma}$ indicates the expected number of cases generated by one infected case in the absence of any intervention and assuming that the whole population is susceptible. At this stage, for

the observed infected and removed proportions, we assume a Beta-Dirichlet state-space model, independent conditionally on the underlying process:

$$Y_t^I|\boldsymbol{\theta_t}, \boldsymbol{\tau} \sim Beta\left(\lambda^I\theta_t^I, \lambda^I(1-\theta_t^I)\right)$$

$$Y_t^R|\boldsymbol{\theta_t}, \boldsymbol{\tau} \sim Beta\left(\lambda^R\theta_t^R, \lambda^R(1-\theta_t^R)\right)$$

Further, the Markov process on the latent proportions is built as:

$$\boldsymbol{\theta_t}|\boldsymbol{\theta_{t-1}}, \boldsymbol{\tau} \sim Dirichlet\left(\kappa f(\boldsymbol{\theta_{t-1}}, \beta, \gamma)\right)$$

where $\boldsymbol{\theta_t}$ denotes the vector of the underlying population probabilities of the three compartments, whose mean is modeled as an unknown function of the probability vector from the previous time point, along with the transition parameters; $\boldsymbol{\tau} = (\beta, \gamma, \boldsymbol{\theta_0^T}, \lambda, \kappa)$ denotes the whole set of parameters where λ^I, λ^R and κ are parameters controlling variability of the observation and latent process, respectively. The function $f(.)$ is then solved as the mean transition probability determined by the SIR dynamical system, using a fourth order Runge-Kutta approximation.

Priors and the MCMC algorithm setup of the eSIR model: The prior on the initial vector of latent probabilities is set as $\boldsymbol{\theta_0} \sim Dirichlet\ (1 - Y_1^I - Y_1^R,\ Y_1^I,\ Y_1^R)$, $\theta_0^S = 1 - \theta_0^I - \theta_0^R$. The prior distribution of the basic reproduction number is $R_0 \sim LogNormal(0.582, 0.223)$ so that $E(R_0) = 2$ and $SD(R_0) = 1$, where E and SD denote the mean and standard deviation respectively. The prior distribution of the removal rate is $\gamma \sim LogNormal(-2.955, 0.910)$ so that $E(\gamma) = 0.082$ and $SD(\gamma) = 0.1$. The prior mean of the removal rate γ indicates an average infectious period of 12 days, which is originally set using the estimation from SARS outbreak in Hong Kong[7] due to the similarity between the two viruses; and this value also aligns well with a couple of recent studies on COVID-19 in China.[8–10] The prior mean of the basic reproduction number, 2.0, is approximately the average of the estimates from many other

COVID-19 studies on the Indian population.[11–15] Note that the prior mean of the distribution of the transmission rate β equals γR_0. For the variability parameters, the default choice is to set large variances in both observed and latent processes, which may be adjusted over the course of epidemic with more data becoming available.

$$\kappa,\ \lambda^I, \lambda^R \sim\ iid\ Gamma(2,\ 0.0001)$$

Denoting t_0 as the last date of data availability, and assuming that the forecast spans over the period $[t_0 + 1, T]$, our algorithm is as follows.

0. Take M draws from the posterior $[\boldsymbol{\theta}_{\mathbf{1}:\boldsymbol{t_0}}, \boldsymbol{\tau}|\boldsymbol{Y}_{\mathbf{1}:\boldsymbol{t_0}}]$.

1. For each solution path $m \in \{1, \dots, M\}$, iterate between the following two steps via MCMC.

i. Draw $\boldsymbol{\theta}_{\boldsymbol{t}}^{(\boldsymbol{m})}$ from $\left[\boldsymbol{\theta}_{\boldsymbol{t}} \middle| \boldsymbol{\theta}_{t-1}^{(m-1)}, \boldsymbol{\tau}^{(m)}\right], t \in \{t_0 + 1, \dots, T\}$.

ii. Draw $\boldsymbol{Y}_{\boldsymbol{t}}^{(\boldsymbol{m})}$ from $\left[\boldsymbol{Y}_{\boldsymbol{t}} \middle| \boldsymbol{\theta}_{t}^{(m)}, \boldsymbol{\tau}^{(m)}\right], t \in \{t_0 + 1, \dots, T\}$.

Modeling intervention: We model the effect of interventions by assuming that the intervention will result in a decrease in the transmission from the S compartment to the I compartment. We do so by decreasing the effective rate of transition (or, equivalently, the chance of interaction between members of S and I), by introducing a time-varying transmission rate modifier $\pi(t) \in [0,1]$. This updates the flow between the three compartments (Supplementary Figure 7) via a set of differential equations as follows:

$$\frac{d\theta_t^S}{dt} = -\beta\pi(t)\theta_t^S\theta_t^I,$$

$$\frac{d\theta_t^I}{dt} = \beta\pi(t)\theta_t^S\theta_t^I - \gamma\theta_t^I,$$

$$\frac{d\theta_t^R}{dt} = \gamma\theta_t^I.$$

The reproductivity is, thus, modified by the intervention over time as $R_0\pi(t)$. In effect, this $\pi(t)$ modifies the chance of a susceptible person meeting with an infected person which is termed as a transmission modifier.

Implementation of the eSIR model: We implemented the proposed algorithm in R package rjags and the differential equations were solved via the fourth-order Runge–Kutta approximation. To ensure the quality of the MCMC, we set the adaptation number to be 10^4, thinned the chain by keeping one draw from every 10 random draws to reduce autocorrelation, set a burn-in period of 10^5 draws to let the chain stabilize, and starting from 4 separate chains. Thus, in total, we have 2×10^5 effective draws with about 2×10^6 draws discarded. This implementation provides not only posterior estimation on parameters and prevalence of all the three compartments in the SIR model, but also predicted proportions of the infected and the removed people at future time point. To get predicted case-counts from the predicted prevalence, we used 1.34 billion as the population of India, thus treating the country as a homogeneous system for the outbreak.

Parameter Table.

eSIR parameter	**Value used for prediction**
Prior mean for $\boldsymbol{R_0}$	2
$\boldsymbol{\pi}$ values by scenario	
Social distancing and travel ban	0.75
Normal (pre-intervention) return	1
Moderate return	0.75
Cautious return	0.60
Lockdown	0.40
Lockdown date	
Start	25 March 2020
End	14 April 2020
$\boldsymbol{\pi}$ transition lengths	
Pre-lockdown to lockdown	7 days
Lockdown to post-lockdown	21 days
Proportion of death in removed compartment	0.2

REFERENCES

1. Jombart T, Zhian N. Kamvar, Cai J, et al. *Reconhub/Incidence: Incidence Version 1.6.0*. Zenodo; 2019. doi:10.5281/ZENODO.2584018

2. Cori A, Ferguson NM, Fraser C, Cauchemez S. A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. *Am J Epidemiol*. 2013;178(9):1505-1512. doi:10.1093/aje/kwt133

3. COVID-19 India. COVID-19 Tracker Updates for India for State-wise and District-wise data. COVID19 India. Published 2020. https://www.covid19india.org/

4. Pan A, Liu L, Wang C, et al. Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China. *JAMA*. 2020;323(19):1915. doi:10.1001/jama.2020.6130

5. Wu JT, Leung K, Bushman M, et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat Med*. 2020;26(4):506-510. doi:10.1038/s41591-020-0822-7

6. Wang L, Zhou Y, He J, et al. *An Epidemiological Forecast Model and Software Assessing Interventions on COVID-19 Epidemic in China*. Infectious Diseases (except HIV/AIDS); 2020. doi:10.1101/2020.02.29.20029421

7. Mkhatshwa T, Mummert A. Modeling Super-spreading Events for Infectious Diseases: Case Study SARS. *ArXiv10070908 Q-Bio*. Published online October 15, 2010. Accessed September 19, 2020. http://arxiv.org/abs/1007.0908

8. Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet*. 2020;395(10223):507-513. doi:10.1016/S0140-6736(20)30211-7

9. Li Q, Guan X, Wu P, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *N Engl J Med*. 2020;382(13):1199-1207. doi:10.1056/NEJMoa2001316

10. Ryu S, Chun BC. An interim review of the epidemiological characteristics of 2019 novel coronavirus. *Epidemiol Health*. 2020;42:e2020006. doi:10.4178/epih.e2020006

11. Das S. Prediction of COVID-19 Disease Progression in India : Under the Effect of National Lockdown. *ArXiv200403147 Cs Q-Bio*. Published online April 7, 2020. Accessed September 19, 2020. http://arxiv.org/abs/2004.03147

12. Deb S, Majumdar M. A time series method to analyze incidence pattern and estimate reproduction number of COVID-19. *ArXiv200310655 Q-Bio Stat*. Published online March 24, 2020. Accessed September 19, 2020. http://arxiv.org/abs/2003.10655

13. Ranjan R. *Predictions for COVID-19 Outbreak in India Using Epidemiological Models*. Epidemiology; 2020. doi:10.1101/2020.04.02.20051466

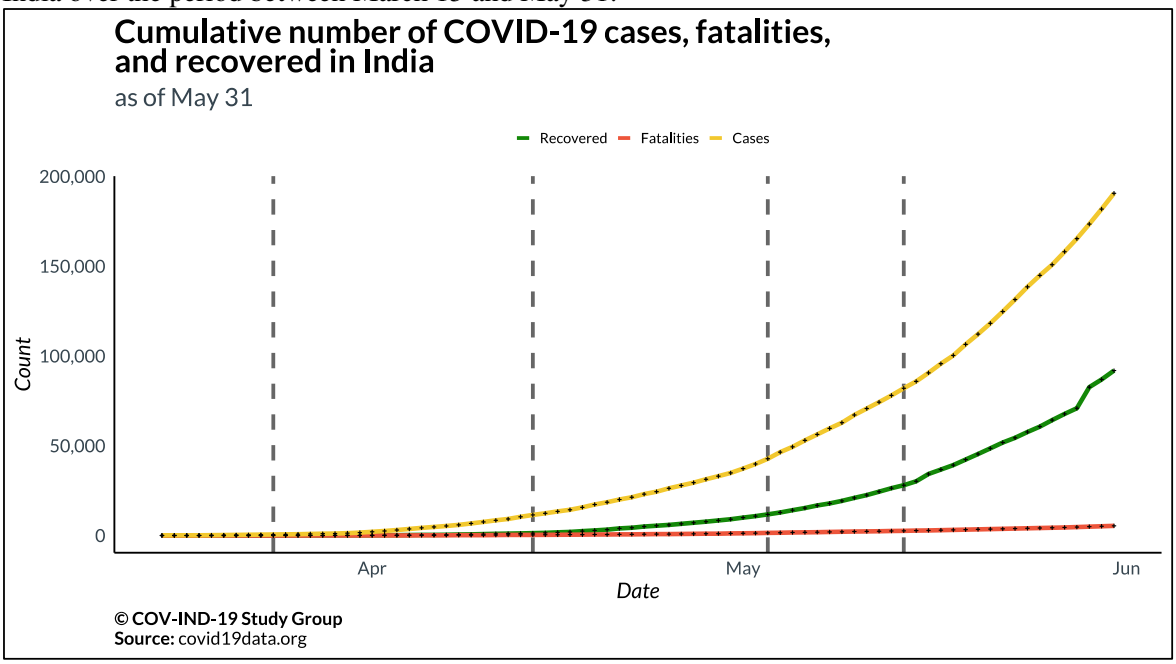
14. Sardar T, Nadim SS, Rana S, Chattopadhyay J. Assessment of lockdown effect in some states and overall India: A predictive mathematical study on COVID-19 outbreak. *Chaos Solitons Fractals*. 2020;139:110078. doi:10.1016/j.chaos.2020.110078

15. Singh R, Adhikari R. Age-structured impact of social distancing on the COVID-19 epidemic in India. *ArXiv200312055 Cond-Mat Q-Bio*. Published online March 26, 2020. Accessed September 19, 2020. http://arxiv.org/abs/2003.12055

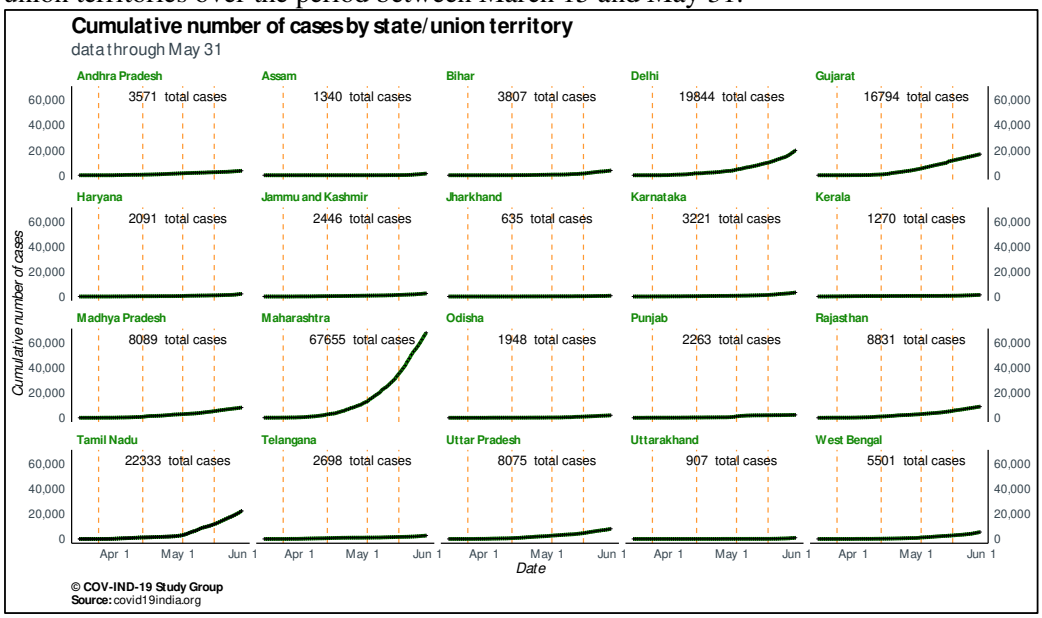
16. Ray D, Salvatore M, Bhattacharyya R, et al. Predictions, Role of Interventions and Effects of a Historic National Lockdown in India's Response to the the COVID-19 Pandemic: Data Science Call to Arms. *Harv Data Sci Rev*. Published online May 14, 2020. doi:10.1162/99608f92.60e08ed5

SUPPLEMENTARY FIGURES

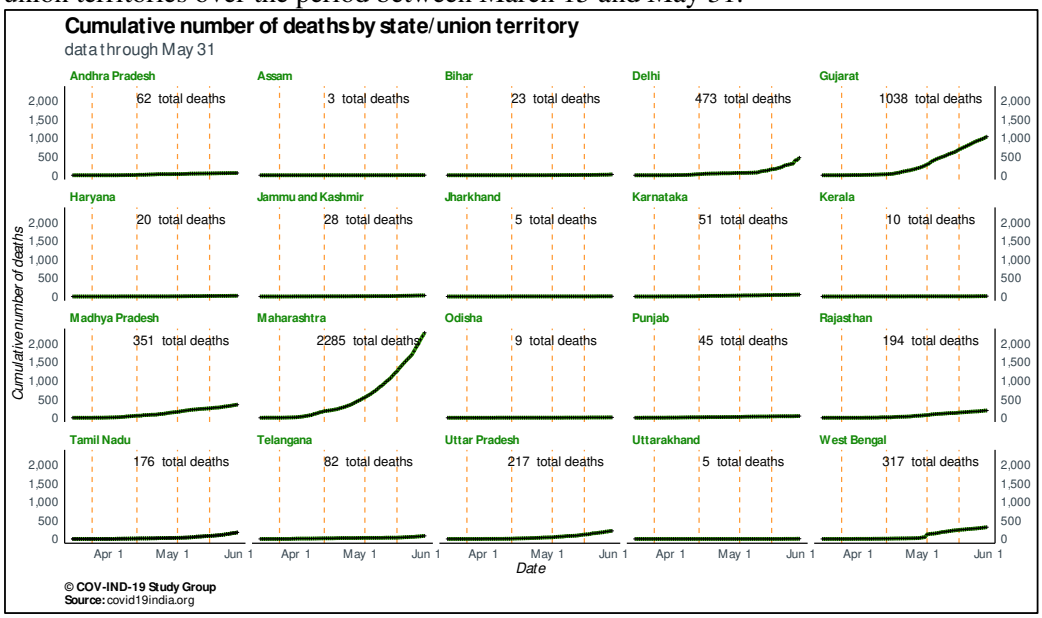
Supplementary Figure 1. Cumulative number of reported cases, fatalities, and recovered cases in India over the period between March 15 and May 31.



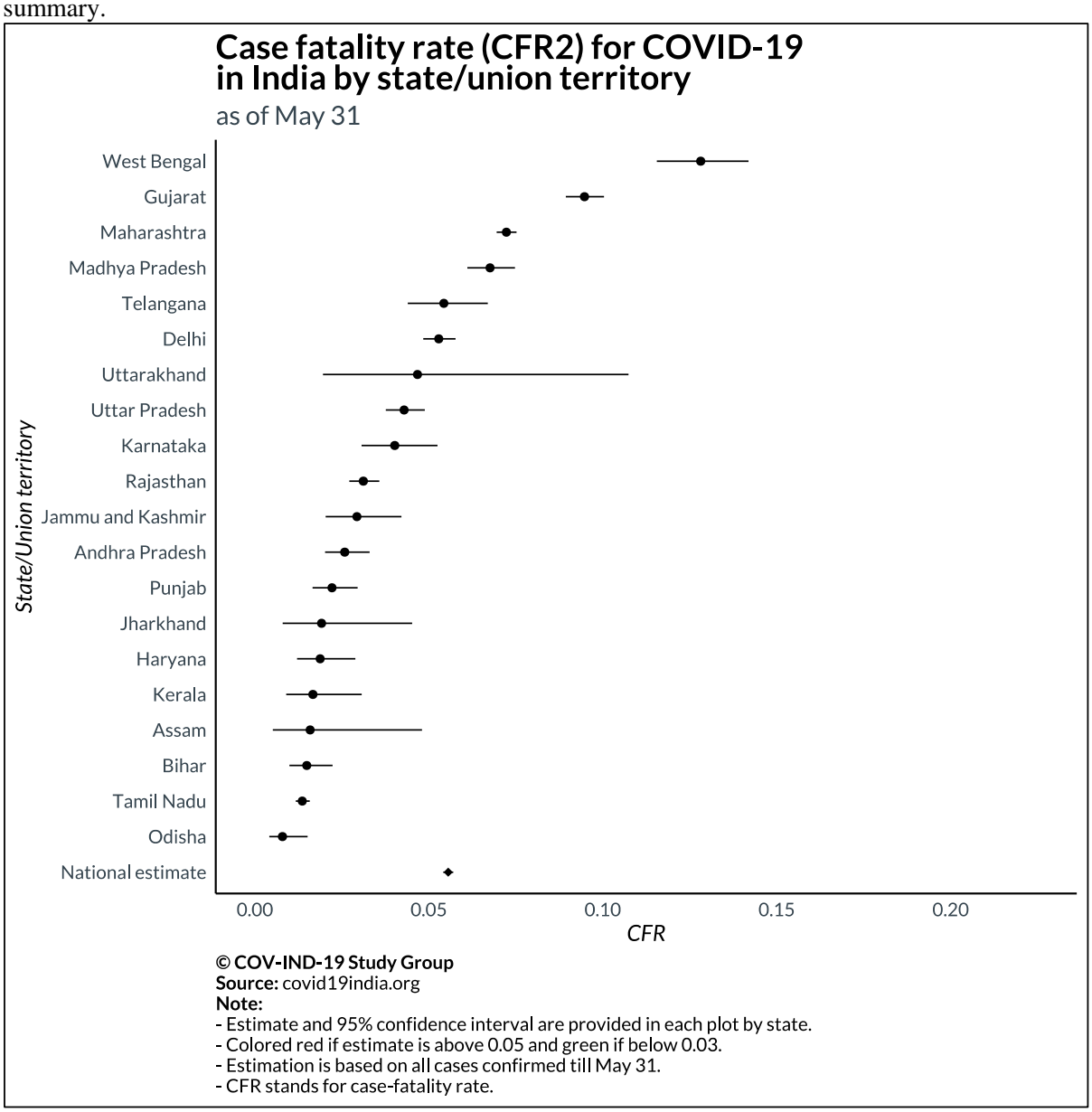
Supplementary Figure 2. Cumulative number of reported COVID-19 cases in 20 Indian states and union territories over the period between March 15 and May 31.



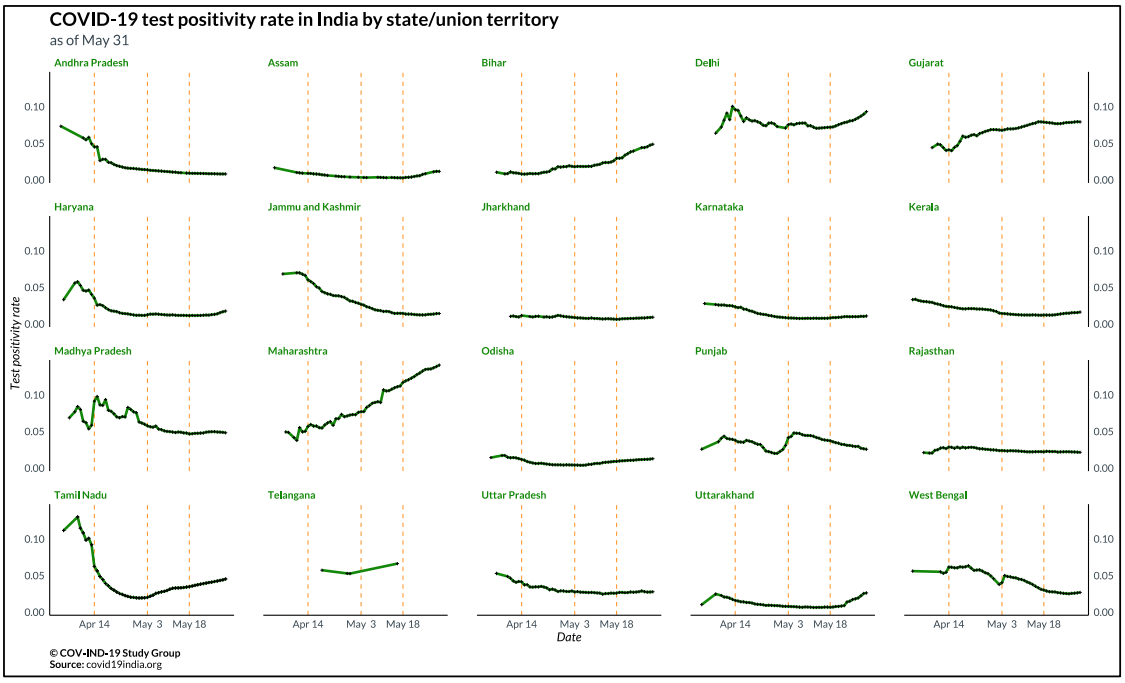
Supplementary Figure 3. Cumulative number of reported COVID-19 deaths in 20 Indian states and union territories over the period between March 15 and May 31.



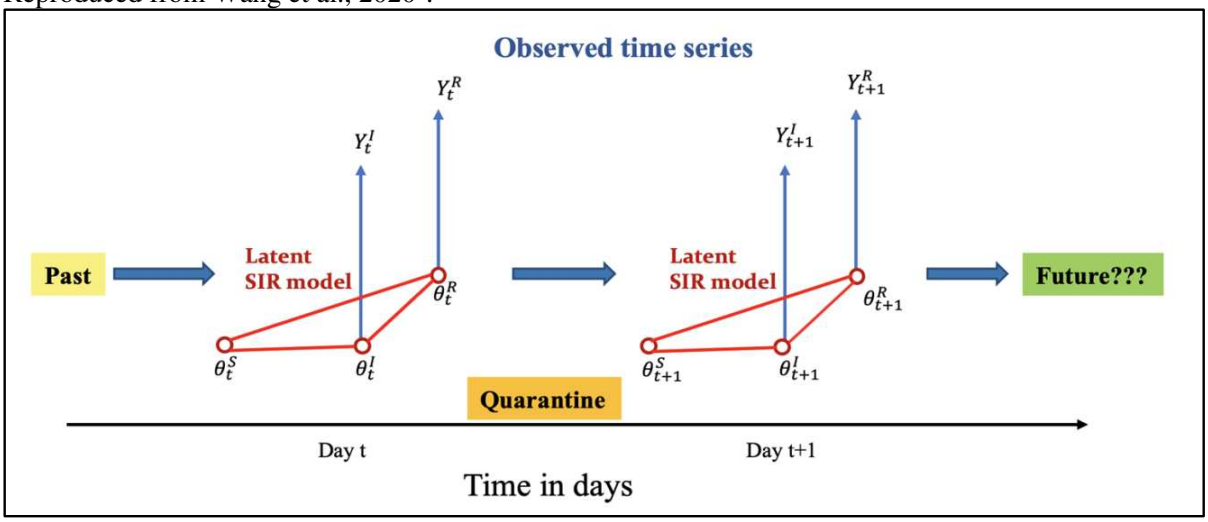
Supplementary Figure 4. Forest plot of estimated case-fatality rates based on closed cases only as of May 31, along with 95% confidence intervals, for 20 states and union territories of India, and a national summary.



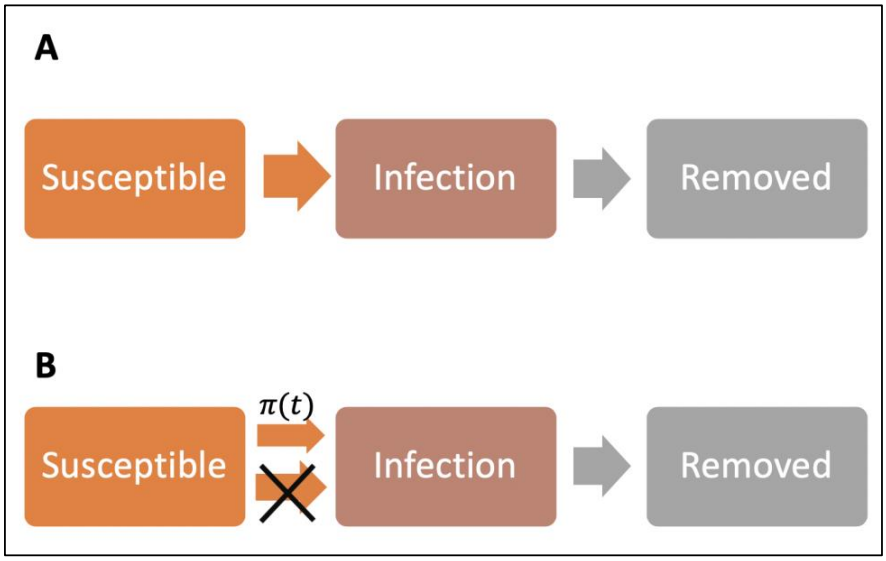
Supplementary Figure 5. Time series plots of test positivity rates for 20 Indian states and union territories.



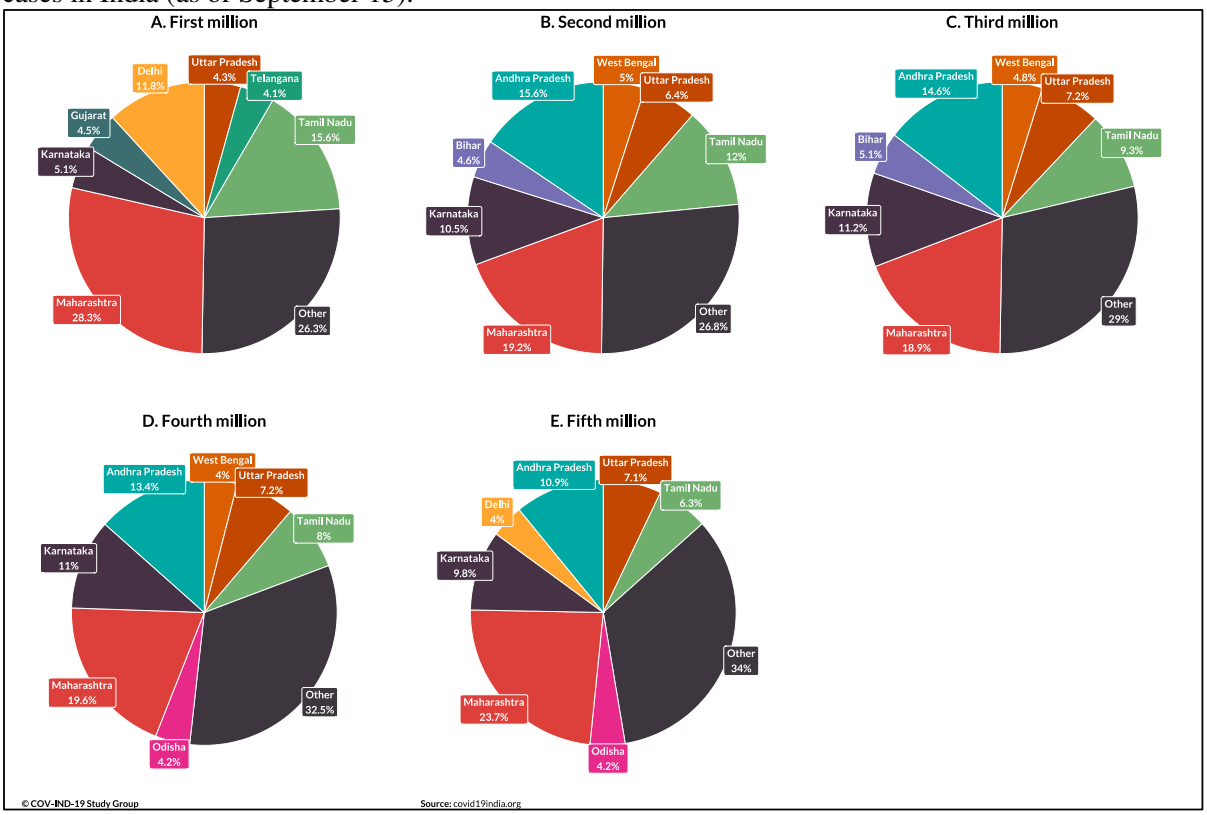
Supplementary Figure 6. The eSIR model with a latent SIR model on the unobserved proportions. Reproduced from Wang et al., 2020[6].



Supplementary Figure 7. The SIR model with (A) or without (B) considering human intervention by introducing a transmission rate modifier π(t). Reproduced from Ray et al., 2020[16].



Supplementary Figure 8. Case distribution by top 7 states for each of the first five million COVID-19 cases in India (as of September 15).



SUPPLEMENTARY TABLES

Supplementary Table 1. COVID-19 metrics table for India and the 20 states with the most cumulative case counts as of September 15, 2020.

Assessing COVID-19 in India

as of September 15

	METRICS							PREDICTED CASES (10/06)	
LOCATION	R	DOUBLING TIME (DAYS)	CFR	TEST POSITIVITY RATE	TOTAL TESTED	POPULATION	PPT (%)	CAUTIOUS RETURN	MODERATE RETURN
National estimate	1.12	33.3	0.016	0.086	57,239,428	1,332,830,000	4.29	5,908,852	6,296,267
Maharashtra	1.24	30.7	0.028	0.202	5,321,116	122,153,000	4.36	1,348,177	1,449,632
Andhra Pradesh	0.98	35.1	0.009	0.123	4,661,355	52,221,000	8.93	672,987	712,328
Karnataka	1.02	32.5	0.015	0.122	3,846,937	65,798,000	5.85	560,696	596,401
Tamil Nadu	0.95	57.3	0.017	0.085	5,968,209	75,695,000	7.88	549,313	565,817
Delhi	1.47	39.7	0.022	0.101	2,184,316	19,814,000	11.02	256,902	272,514
Uttar Pradesh	1.12	29.8	0.014	0.042	7,636,000	224,979,000	3.39	388,070	417,900
West Bengal	1.03	41.5	0.019	0.082	2,517,595	96,906,000	2.60	228,717	238,418
Telangana	0.93	41.9	0.006	0.073	2,169,339	37,220,000	5.83	184,727	194,693
Odisha	1.13	24.0	0.004	0.063	2,472,517	43,671,000	5.66	197,370	216,890
Assam	0.92	36.6	0.003	0.052	2,750,037	34,293,000	8.02	176,558	190,407
Haryana	1.32	24.0	0.010	0.063	1,514,575	28,672,000	5.28	123,585	136,458
Kerala	NA	24.1	0.004	0.050	2,152,585	35,125,000	6.13	141,842	155,999
Bihar	0.86	64.1	0.005	0.032	4,986,747	119,520,000	4.17	172,865	178,163
Madhya Pradesh	1.26	27.8	0.020	0.053	1,700,929	82,232,000	2.07	114,008	124,561
Punjab	1.32	22.7	0.030	0.058	1,410,759	29,859,000	4.72	108,181	119,840
Chhattisgarh	1.33	13.1	0.009	0.084	806,045	28,724,000	2.81	127,891	157,059
Rajasthan	1.08	40.2	0.012	0.039	2,672,224	77,264,000	3.46	119,250	125,613
Gujarat	1.03	55.6	0.028	0.034	3,360,318	67,936,000	4.95	127,466	132,406
Jharkhand	0.96	23.6	0.009	0.045	1,407,470	37,403,000	3.76	78,991	86,301
Jammu and Kashmir	1.56	23.4	0.016	0.044	1,248,495	13,203,000	9.46	73,976	81,409

© COV-IND-19 Study Group

Source data: covid19india.org

Notes: Only states/union territories with the highest cumulative case counts as of September 15 are shown. Predicted cases are for October 6 based on data through September 15. National Commission on Population 2019 projections used to calculate PPT.

Abbrev: CFR, Case-fatality rate; PPT, Proportion of population tested

Supplementary Table 2. Existing articles that incorporate migration in COVID-19 models

Sl.No.	Model type	Reference	Research question
1.	Modified SEIR model	Maji et al (2020)	Predict the temporal variation in confirmed and active cases of COVID-19 in selected states of India with high outflux of migrant workers.
2.	Network-based model	Kumar (2020)	Predict spread of COVID-19 at different geographical locations in India using reported COVID-19 cases, census migration data, and monthly airline data of passengers.
3.	A hybrid of SIR and spatial network model	Pujari and Shekatkar (2020)	Study the spread of COVID-19 in India using domestic transport networks, such as aviation and railways, and incorporating distance-dependent temporal delays in migration.
4.	Extended SEIR model	Gupta et al (2020)	Generate qualitative projections of COVID-19 spread in India, and investigate the effects of different public health interventions by incorporating heterogeneity at geographical and infrastructural levels and in local responses. P.S.: Authors mention that they use mobility patterns in normal times and do not take into account large-scale worker migration that took place at the start of (un-)lockdown.
5.	Spatial network-based extended SEIR model	Sharma et al (2020)	Study the spread dynamics of COVID-19 in different states of India accounting for time delay, spatial heterogeneity, and population migration networks; and examine the impact of the most significant lockdown measure in containing the pandemic spread.

Abbreviations: SIR, Susceptible-Infected-Removed; SEIR, Susceptible-Exposed-Infected-Removed

Supplementary Table 3. Effect of underreporting of cases and deaths on infection fatality rate

Place	Estimated Seroprevalence (%)	Observed data (as of 9/15)		CFR	URF_D	IFR			
						Based on seroprevalence	URF_C 10	URF_C 20	URF_C 30
Delhi	22.86[a]	225,796	4,806	0.021	1	0.0011	0.0021	0.0011	0.0007
					5	0.0055	0.0106	0.0053	0.0035
					10	0.0111	0.0213	0.0106	0.0071
Mumbai	40.5[b]	173,596	8,230	0.047	1	0.0011	0.0047	0.0024	0.0016
					5	0.0055	0.0237	0.0119	0.0079
					10	0.011	0.0474	0.0237	0.0158
Pune	51.5[c]	239,481	4,888	0.020	1	0.003	0.002	0.001	0.0007
					5	0.0152	0.0102	0.0051	0.0034
					10	0.0305	0.0204	0.0102	0.0068

[a] https://indianexpress.com/article/explained/delhi-serological-survey-shows-antibodies-in-23-participants-what-does-this-mean-6516512/
[b] https://indianexpress.com/article/cities/mumbai/higher-share-in-slums-exposed-to-virus-than-in-societies-mumbai-sero-survey-6527865/
[c] https://indianexpress.com/article/cities/pune/first-sero-survey-shows-extensive-spread-of-covid-19-from-36-1-to-65-4-in-selected-areas-sampled-in-pune-6558853/
Observed data collected from covid19india.org.
Abbreviations: *CFR*, case fatality rate; *IFR,* infection fatality rate; URF_C, underreporting factor for reported cases; URF_D, underreporting factor for deaths.