HHS Public Access

Author manuscript

Harv Data Sci Rev. Author manuscript; available in PMC 2020 June 30.

Published in final edited form as:

Harv Data Sci Rev. 2020 ; 2020(Suppl 1): . doi:10.1162/99608f92.60e08ed5.

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

Debashree Ray[1,2,*], **Maxwell Salvatore**[3,4,*], **Rupam Bhattacharyya**[3], **Lili Wang**[3], **Jiacong Du**[3,4], **Shariq Mohammed**[3,5], **Soumik Purkayastha**[3], **Aritra Halder**[6], **Alexander Rix**[3,4], **Daniel Barker**[3], **Michael Kleinsasser**[3], **Yiwang Zhou**[3], **Debraj Bose**[3], **Peter Song**[3], **Mousumi Banerjee**[3,8], **Veerabhadran Baladandayuthapani**[3], **Parikshit Ghosh**[7], **Bhramar Mukherjee**[3,4,§]

[1]Department of Epidemiology, Johns Hopkins University

[2]Department of Biostatistics, Johns Hopkins University

[3]Department of Biostatistics, University of Michigan

[4]Center for Precision Health Data Science, University of Michigan

[5]Department of Computational Medicine and Bioinformatics, University of Michigan

[6]Department of Statistics, University of Connecticut

[7]Department of Economics, Delhi School of Economics

[8]Institute for Healthcare Policy and Innovation, University of Michigan

Abstract

With only 536 cases and 11 fatalities, India took the historic decision of a 21-day national lockdown on March 25. The lockdown was first extended to May 3 soon after the analysis of this paper was completed, and then to May 18 while this paper was being revised. In this paper, we use a Bayesian extension of the Susceptible-Infected-Removed (eSIR) model designed for intervention forecasting to study the short- and long-term impact of an initial 21-day lockdown on the total number of COVID-19 infections in India compared to other less severe non-pharmaceutical interventions. We compare effects of hypothetical durations of lockdown on reducing the number of active and new infections. We find that the lockdown, if implemented correctly, can reduce the total number of cases in the short term, and buy India invaluable time to prepare its healthcare and disease-monitoring system. Our analysis shows we need to have some measures of suppression in place after the lockdown for increased benefit (as measured by reduction in the number of cases). A longer lockdown between 42–56 days is preferable to substantially "flatten the curve" when

§ Corresponding author: bhramar@umich.edu.

*Co-first authors

Author contributions: P.G. and B.M. conceptualized this study. D.R., M.S., S.M., A.H. and D.B. curated the data. M.S., R.B., L.W., J.D., S.M. and S.P. analyzed the data. B.M. participated in funding acquisition, investigation and provided resources. L.W., P.S. and B.M. developed the methodology. M.S. and B.M. administered the project. M.S., A.R., D.B. and M.K. created the software/app. M.B., V.B., P.G. and B.M. supervised the study. D.R., M.S., R.B., L.W., A.R., M.K. and B.M. helped in the visualization. D.R., M.S., R.B., L.W., P.G. and B.M. wrote the original draft. All authors reviewed and edited the final manuscript.

compared to 21–28 days of lockdown. Our models focus solely on projecting the number of COVID-19 infections and, thus, inform policymakers about one aspect of this multi-faceted decision-making problem. We conclude with a discussion on the pivotal role of increased testing, reliable and transparent data, proper uncertainty quantification, accurate interpretation of forecasting models, reproducible data science methods and tools that can enable data-driven policymaking during a pandemic. Our software products are available at covind19.org.

Keywords

Basic reproduction number; Coronavirus; Credible interval; India; Intervention forecasting; SIR model

1. Introduction

Four months since the first case of COVID-19 in Wuhan, China, the SARS-CoV-2 virus has engulfed the world and has been declared a global pandemic ("WHO Director-General's Opening Remarks at the Media Briefing on COVID-19 – 11 March 2020," 2020). The number of confirmed cases worldwide stands at a staggering 1,930,780 (as of 9:20 AM EST April 14, 2020, Microsoft bing coronavirus tracker (*Microsoft bing COVID-19 tracker*)). Of these, 10,815 confirmed cases are from India (Figure 1), the world's largest democracy with a population of 1.34 billion (compare China at 1.39 billion and USA at 325.7 million) (World Bank). India has been vigilant and early in instituting strong public health interventions including sealing the borders with travel ban/canceling almost all visas, closing schools and colleges and diligently following up with community inspection of suspected/exposed cases with respect to adherence of quarantine recommendations (Table 1). On March 24, India took the historic decision of a 21-day national lockdown starting March 25, when it had reported only 536 COVID-19 cases and 11 fatalities. In the subsequent days, we have seen a steady growth in the number of new cases and fatalities, with growth rates slower than other affected countries but in 21 days, the curve has not yet "turned the corner" or showed a steady decline in the number of newly diagnosed cases (Figure 2). All forecasting models in this paper use data up to April 14 with the premise of a 21-day lockdown in place. The lockdown was further extended to May 3, soon after the analysis of this paper was completed. The lockdown was again extended to May 18 while this paper was being revised, thus entailing almost an 8-week lockdown period.

While India seems to have done relatively well in controlling the number of confirmed cases compared to other countries in the early phase of the pandemic (Figure 2), there is a critical missing or unknown component in this assessment: "The number of truly affected cases," which depends on the extent of testing, the accuracy of the test results and, in particular, the frequency and scale of testing of asymptomatic cases who may have been exposed. The frequency of testing has been low in India. According to the Indian Council of Medical Research (ICMR), only 229,426 subjects have been tested as of April 14 (<0.03% of the population) (Indian Council of Medical Research, 2020). When there is no approved vaccine or drug for treating COVID-19, entering phase 2 or phase 3 of escalation will have devastating consequences on both the already overstretched healthcare system of India, and

India's large at-risk sub-populations (Appendix Table 1). As seen for other countries like the US or Italy, COVID-19 enters gradually and then explodes suddenly.

We provide a table listing other highly affected countries along with their first reported case, initial interventions, crude fatality rates, and active case counts in Appendix Table 2 for reference. The estimated capacity of hospital beds in India is 70 per 100,000 (Hospital Beds (per 1,000 People), 2020), which is an upper bound on treatment capacity. Given an average occupancy rate of 75%, only a quarter of these are available (Sindhu et al. 2019). Moreover, critically ill COVID-19 patients (about 5–10% of those infected) will require ICU beds and ventilator support. India has only 35–58 thousand ICU beds with very high occupancy rates and at most 1 ventilator per 2 ICU beds (Times of India). In order to roll out interventions and plan for healthcare infrastructure, robust projection models for outcomes of interest are necessary. There are many outcomes that are of potential interest to policymakers; for example: how many infected cases will be hospitalized? How many will be admitted to the ICU? How many patients will need ventilators? And, finally, what will be the mortality due to COVID-19 infections? We focus on the number of active cases as our target of prediction due to the limited data from India on the other outcomes. From other nations we know that roughly 20% of infections will probably need hospitalization (Root, 2020) and 5% will need ICU admission (Guan et al., 2020), and case-fatality rates vary between 1–5% (Oke & Henegan, 2020). This may provide crude estimates of other outcomes from case-count predictions.

At the time of writing this manuscript, there exist several models that have been used to analyze the COVID-19 case-count data from India. The approaches for modeling the disease transmission and then forecasting the number of cases at a future time can be broadly categorized into two genres: exponential/Poisson type models, and compartmental epidemiologic models. For instance, Ranjan (2020) and Gupta and Shankar (2020) used the classical exponential model, S. Das (2020) used a Poisson regression model while Deb and Majumdar (2020) used an auto-regressive moving average model to analyze incidence pattern over time. The compartmental epidemiologic models include variations of the susceptible-infected-removed (SIR) model, which is guided by a set of differential equations relating the number of susceptible people, the number of infected people (cases) and the number of people who have been removed (either recovered or dead) at any given time. This simple SIR model has been used by Ranjan (2020) and Dhanwant and Ramanathan (2020). Singh and Adhikari (2020) used an age-structured SIR and social contact model, where an SIR model is assumed in each age category. Another extension of the SIR model is the SEIR model that incorporates an additional compartment of truly exposed people that is a latent variable. Mandal et al. (2020), Chatterjee et al. (2020), Sardar et al. (2020) and Senapati et al. (2020) used one or the other variation of the SEIR model. For example, Sardar et al. (2020) used an extra compartment for lockdown to capture in-home isolation and study the effects of lockdown on future case-counts. Appendix Table 3 compares and summarizes these existing models for India.

In this article, we apply a Bayesian extension of the SIR model (eSIR) to explore two primary forecasting objectives: (a) Forecasting future case-counts (short term and long term) with different forms of suppression measures in place (post lockdown) (b) Studying the

relative impact of length/duration of a lockdown on our predictions of cumulative COVID-19 infections. We carry out extensive sensitivity analysis to assess the robustness of our forecasting models. We conclude with a discussion regarding the need for reliable case-count data, increased testing, the importance of uncertainty quantification of the projected case-counts, and transparent data science methods that can inform and influence policymaking during a pandemic. Our data science products include three articles on Medium studying pre (Ray et al., 2020), during (Salvatore, Wang, et al., 2020) and post (Salvatore, Ray, et al., 2020) lockdown effects, providing critical information for policymakers and having an extensive reach (Reuters (Ghosal, 2020), Times of India (P. Das, 2020), The Guardian (Ellis-Petersen, 2020), The Economic Times (Noronha, 2020)), an interactive and dynamic *RShiny* app that daily updates forecasts as new case counts are reported, and publicly available codes for reproducible research.

The rest of the paper is organized as follows. In Section 2 we describe the structure of the eSIR model, our parameter choices and the Bayesian computational algorithm. In Section 3 we present results from analyzing the data from India that include a sensitivity analysis. We assess how our forecasting model updates itself with more accrual of data over time. In Section 4 we provide an itemized discussion of some of the salient data and data science issues related to intervention forecasting and case-count projections. Section 5 presents a brief conclusion.

2. Methods and Notation

2.1. Study design and data source

We used the current daily data on number of COVID-19 infected cases, recoveries and deaths in India to predict the number of infected and removed cases at any given time (L. Wang et al., 2020). We obtained the data (up to April 14) from the 2019 Novel Coronavirus Visual Dashboard operated by the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE) and from covid19india.org (covid19india, 2020; Dong et al., 2020; Johns Hopkins University Center for Systems Science and Engineering). Some of our testing data came from https://ourworldindata.org/coronavirus.

2.2. Our statistical model for predictions

Overview—The standard SIR model was recently extended to incorporate time-varying transmission rates or time-varying quarantine protocols and is known as the extended SIR (eSIR) model (L. Wang et al., 2020). 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.

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 observe only the daily proportions of infected cases and removed cases. First, let us establish some notation. 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 (Appendix Figure 1A), 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 $R_0 := \frac{\beta}{\gamma}$ 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:

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

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

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

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; $\boldsymbol{\tau} = \left(\beta, \gamma, \boldsymbol{\theta_0^T}, \lambda, \kappa\right)$ denotes the whole set of parameters where λ^I, λ^R and $\boldsymbol{\kappa}$ 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 $\boldsymbol{\theta_0} \sim Dirichlet\left(1 - Y_1^I - Y_1^R, Y_1^I, Y_1^R\right)$, $\theta_0^S = 1 - \theta_0^I - \theta_0^R$. 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 estimates from the SARS outbreak in Hong Kong (Mkhatshwa & Mummert, 2010) due to the similarity between the two viruses; this value also aligns well with several recent studies on COVID-19 in China (Chen et al. 2020; Li et al., 2020; Ryu & Chun, 2020). The prior mean of the basic reproduction number, 2.0, is approximately the average of estimates from many other COVID-19 studies on the Indian population (S. Das, 2020; Deb & Majumdar, 2020; Ranjan, 2020; Sardar et al., 2020; Singh & Adhikari, 2020). We have conducted sensitivity analysis to evaluate how robust the model is towards the prior settings using Indian population COVID-19 data. The sensitivity issue can be minimized with more observed data of longer exponentially increasing period and stronger intensities by focusing on cities or states that are highly exposed. 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 the 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 $\left[\boldsymbol{\theta}_{\mathbf{1}:\boldsymbol{t_0}}, \boldsymbol{\tau} | \boldsymbol{Y}_{\mathbf{1}:\boldsymbol{t_0}}\right]$.

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

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

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 (Appendix Figure 1B) 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 $R_0\pi(t)$. To better understand the introduction of this effect modifier, we follow an example given by L. Wang et al. (2020). Suppose at a time t, $q^{S(t)} \in [0,1]$ is the chance of an at-risk person being in home isolation, and $q^{I(t)} \in [0,1]$ is the chance of an infected person being in hospital quarantine. Consequently, the chance of disease transmission when an at-risk person meets an infected person is $\beta\left\{1-q^{S(t)}\right\}\theta_t^S\left\{1-q^{I(t)}\right\}\theta_t^I := \beta\,\pi(t)\theta_t^S\theta_t^I$, with $\pi(t) := \left\{1-q^{S(t)}\right\}\left\{1-q^{I(t)}\right\} \in [0,1]$. In effect, this $\pi(t)$ modifies the chance of a susceptible person meeting with an infected person which is termed as a transmission modifier. In this article, the functional form of $\pi(t)$ is a continuous function that reflects a combination of steadily increased community-level awareness and responsible quarantine and preventive measures, and the country-wide lockdown measures initiated by the government. This predefined transmission modifier can be smoothly incorporated into the differential equations as well as the MCMC algorithms. Its functional form can be quite flexible in reflecting the changing pattern of human intervention that affects the transmission rate of the epidemic within the population.

Implementation of the eSIR model—We implemented the proposed algorithm in R package rjags (Plummer et al.) and the differential equations were solved via a fourth-order Runge–Kutta approximation. To ensure 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 burnin 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. One could reduce the computation time, but consequently might risk the quality of data. This implementation provides not only posterior estimation on parameters and prevalence of all three compartments in the SIR model, but also predicted proportions of the infected and the removed cases at future time points. To obtain 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 (The World Bank).

Uncertainty Quantification—One major advantage of a Bayesian implementation is that uncertainty associated with all parameters and functions of parameters can be calculated from exact posterior draws without relying on large-scale approximation or the delta method. The credible intervals (CrI) for the prevalence are computed using the posterior distribution of proportions given the observed confirmed and removed prevalence, i.e., $Y^I_{(t_0+1):T}|Y^I_{1:t_0}, Y^R_{1:t_0}$ and $Y^R_{(t_0+1):T}|Y^I_{1:t_0}, Y^R_{1:t_0}$, where t_0 denotes the last observed date, and T denotes the last forecast date.

More specifically, suppose we want to compute the 95% posterior CrI for the observed proportion of confirmed cases on the first day of forecast, i.e., a CrI for the random variable $Y^I_{t_0+1}$. Then, from the M solution paths of the posterior, we have the draws $\left\{Y^{I(m)}_{t_0+1}, 1 \le m \le M\right\}$. We construct a 95% posterior CrI for $Y^I_{t_0+1}$ by simply computing the 2.5^{th} upper and lower percentiles from this set of M draws. The cumulative prevalences are sums of the draws from the I and R compartments at a given time and thus the confidence

interval for the sum can be calculated in a similar way. Case-counts can be obtained from prevalences by using population size. Similar techniques apply to $\theta^I_{t_0+j}$ for any $1 \le j \le T - t_0$ and transmission parameters like β and γ. For instance, a 95% posterior CrI for β can be constructed by calculating the 2.5th upper and lower percentiles of $\{\beta^{(m)}, 1 \le m \le M\}$. Therefore, we could simply define $R^{(m)} = \frac{\beta^{(m)}}{\gamma^{(m)}} \; \forall \; 1 \le m \le M$, and compute the 95% posterior CrI for the effective reproduction number R from $\{\beta^{(m)}, 1 \le m \le M\}$.

2.3. Parameter choices for short-term forecasts

We made projections of the cumulative number of cases over a time horizon to assess the short-term impact of lockdown as well as the long-term impact of lockdown and post-lockdown activities. For the short-term forecast on April 30, we assumed lockdown is implemented until April 14 with either a 1- or a 2-week delay in people's adherence/compliance to lockdown restrictions. We compared these projections with two hypothetical scenarios: (A) no non-pharmaceutical intervention (i.e., a constant disease transmission rate over time since the first case was reported in India), (B) a moderate intervention with social distancing and travel bans only (i.e., a decreased transmission rate compared to no intervention). The prior mean for R_0 (the expected number of cases generated by one infected person assuming that the whole population is susceptible) was set at 2.0. This was estimated based on the early phase data in India and is consistent with other models (S. Das, 2020; Deb & Majumdar, 2020; Ranjan, 2020; Sardar et al., 2020; Singh & Adhikari, 2020). For the no intervention and the moderate intervention scenarios, we chose the transmission rate and the removal rate such that the means for the prior distribution of the basic reproductive number R_0 are 2.0 and 1.5 respectively (SD=1). The change in R_0 from 2.0 to 1.5 as an effect of intervention was created based on what we saw regarding the effect of interventions and the relative reduction of R_0 in Wuhan (C. Wang et al., 2020). Given the similar population size and comparable population densities in China and India, the assumption on similar effect of interventions on the pandemic across the two countries does not seem too restrictive. For the current scenario of lockdown, our chosen mean for the prior of R_0 starts with 2.0 during the period of no intervention, drops to 75% of its original value or 1.5 during the period of moderate intervention, and further drops to 0.8 during the 21-day lockdown period, and moves back up to 1.5 after the lockdown ends as described in Figure 3 (assuming a gradual, moderate resumption of daily activities). The drop in R_0 from 2.0 to 0.8 during lockdown represents a 60% reduction, which is proportionally slightly less than the ~65% drop estimated in R_0 from the COVID-19 outbreak in Wuhan following the introduction of *cordon sanitaire* (Lin, 2020; Pan et al., 2020).

2.4. Parameter choices for long-term forecasts

For the longer-term forecast until June 15, we considered three hypothetical post-lockdown scenarios: (i) people return to normal activities due to the urgent desire for reconnecting after lockdown; (ii) people return to moderate activities as they did during the period with social distancing and travel ban intervention; and (iii) people make a cautious return out of fear for the coronavirus and partake in subdued activities. For these three scenarios, we assume the prior mean on R_0 moves back up from 0.8 to 2.0, 1.5 and 1.2 respectively three

weeks after lockdown ends on April 14. We compared these post-lockdown scenarios with another hypothetical scenario involving perpetual social distancing and travel ban without any lockdown (we fixed the prior mean on R_0 at 1.5 over the entire forecasting interval). The time-dependent changes to R_0 values across our simulation scenarios are depicted in Figure 3.

2.5. Parameter choices for duration of lockdown analysis

To assess the long-term impact of lockdown duration, we considered four scenarios: 21-, 28-, 42, and 56-day lockdown periods. In all scenarios, we assume the prior mean on R_0 remains at 0.8 for the duration of the lockdown. Post-lockdown, the prior mean on R_0 gradually returns to a value of 1.5 over a span of three weeks (analogous to the "moderate return" scenario). The changes to R_0 values across our simulation scenarios for studying length of lockdown are depicted in Appendix Figure 7.

2.6. Open-source software

We are committed to data transparency and reproducible research. Daily updates of our India projections, based on cases, recovered and deaths reported the day before by covid19india.org, a crowd-sourced database using state bulletins and official handles, can be found in our interactive and dynamic Shiny app (covind19.org). Apart from the scenarios described in this article, anyone can create predictions under other hypothetical scenarios with quantification of uncertainties. Open source code underlying this app are available at https://github.com/umich-cphds/cov-ind-19.

3. Results

As we interpret the results from our model, let us use caution in not over-interpreting the numbers. Any statistical model is wrinkled with many assumptions. Similarly, the predictions themselves have large uncertainty (as reflected by the large upper credible limits). A rigorous quantitative treatment often allows us to analyze a problem with clarity and objectivity, but we recommend *focusing more on the qualitative takeaway messages from this exercise rather than concentrating on the exact numerical projections or quoting them with certainty*.

3.1. Short-term forecast of cumulative case counts in India

Under national lockdown (March 25 – April 14), our predicted cumulative number of COVID-19 cases in India on April 30 are 19,625 and 19,503 (upper 95% CrI 130,326 and 129,422) assuming a 1- or 2-week delay (i.e., either a quick or a slow adherence), respectively, in people's adherence to lockdown restrictions and a gradual, moderate resumption of daily activities post-lockdown (Figure 4, Appendix Figure 3). In comparison, the predicted cumulative number of cases under "no intervention" and the "intervention involving social distancing and travel bans without lockdown" are 222 thousand and 53 thousand (upper 95% CrI of nearly 1.4 million and 0.3 million) respectively. Under quick adherence, these figures correspond to a relative reduction of 91% and 63% of cases due to lockdown with moderate return compared to "no intervention" and "social distancing and travel ban". The *relative reduction* in cases between two scenarios (often from the least to

the most intense intervention) is calculated as the difference between an estimate (on a particular day, e.g., April 30) under the social distancing and travel ban scenario and under the lockdown with moderate return scenario and then divided by the estimate under the social distancing and travel ban scenario.

We are reporting only the upper credible limit here and elsewhere since the lower credible limits are very small and uninformative due to the large uncertainty in our predictions arising from many unknowns. We also believe that our *point estimates are at best underestimates* due to potential surveillance bias (under-reporting and/or misdiagnosis of case-counts) and our model not taking into account the population density, age-sex composition and regional contact network structure of the whole nation. Increase in testing and community transmission may lead to a spike in a single day and that may increase the projections substantially upward. Regardless of the exact numbers it is clear that the 21-day lockdown will likely have a strong relative effect on reducing the predicted number of cases in the short term when compared to weaker interventions.

3.2. Long-term impact of lockdown on the outbreak in India

We took a close look at what might be coming in the next two months, based on what we have seen in other countries and an epidemiologic model that has been gainfully employed to assess the effect of interventions in Hubei province (L. Wang et al., 2020). We estimate that *roughly 388 thousand* (upper 95% CrI 2.4 million), *7.5 million* (upper 95% CrI 104 million), and *18.5 million* (upper 95% CrI 196 million) cases are prevented on May 15, June 15, and July 15, respectively, by instituting a 21-day lockdown with *quick adherence and a cautious return* compared to perpetual social distancing and travel ban (without lockdown) (Figure 5). This corresponds to relative reductions in cases of 93%, 96%, and 91%, respectively, compared to perpetual social distancing.

Without some measures of suppression after lockdown is lifted, the impact of lockdown in bringing down the case-counts (the now ubiquitous term, “flattening the curve”) can be negated by as early as the first week of June. In fact, in Figure 5a, the pre-intervention (“normal”) curve first passes the social distancing and travel ban curve on June 5. In particular, if people immediately go back to pre-intervention (“normal”) activities post-lockdown, a surge in the predicted case-counts is expected in the long-term beyond what we would have seen if there were only social distancing and travel ban measures without any lockdown (27 million when post-lockdown activity returns to pre-intervention levels versus 26 million under social distancing and travel ban without a lockdown period on July 31; Figure 5). Longer lockdowns would delay this crossover, but a normal (pre-intervention) return post-lockdown would surpass social distancing and travel ban (if these scenarios continued perpetually). Long-term forecasting under slow adherence (2-week delay) can be seen in Appendix Figure 4.

We present posterior density and trace plots for the underlying model parameters β, γ and R_0, posterior distributions for the predictions Y and the latent proportions θ for the I and R compartments over time, and estimates and posterior distribution of the daily prevalence of active cases over time or $\frac{d\theta_t^I}{dt}$. These are contained in Appendix Figure 5.

3.3. Relative impact of duration of lockdown on predicted case-counts

We took the quick adherence epidemiologic models and compared the 21-day lockdown with hypothetical 28-, 42-, and 56-day lockdown scenarios (Figure 6). When comparing a 21-day lockdown with a hypothetical lockdown of longer duration, we find that 28-, 42-, and 56-day lockdowns can approximately prevent 733 thousand (upper 95% CrI 6.8 million), 1.4 million (upper 95% CrI 9.8 million), 1.6 million (upper 95% CrI 10.3 million) cases by June 15, respectively. These numbers correspond to a relative reduction in cases of 45%, 86% and 96%, respectively. A 28-day lockdown does not appear to have a substantial impact on cumulative case counts when compared to a 21-day lockdown. From an epidemiologic perspective, there appears to be some evidence that suggests a 42- or 56-day lockdown would have a more meaningful impact on reducing cumulative COVID-19 case counts in India. Our models suggest that some form of post-lockdown suppression (e.g., extension of social distancing measures, limits of gathering size, etc.) is necessary to observe long-term benefits of the lockdown period. We note that longer lockdown periods are also accompanied by increasing costs to individuals such as economic costs, mental health issues, other public health exacerbation costs, and must be considered in policymaking.

Lockdown duration study under the slow adherence (2-week delay) scenario can be found in Appendix Figure 6. The implied R_0 plots can be found in Appendix Figure 7.

3.4. Sensitivity analyses

We did explore some alternative assumptions and conducted thorough sensitivity analysis before settling on the models presented above. In one example, we assumed that there are actually 10 times the number of reported cases to date to reflect potential underreporting of cases due to lack of testing. We note that our predictions of case-counts indeed go up and the effect of under-reporting of cases is more palpable with long term projections (see Table 2, under-reporting). In another scenario, we assumed these cases occurred in metropolitan areas to reflect a potential intensification of case clustering. In our primary analyses, we assumed that the cumulative case-counts across the country represent equal contributions from all the regions, and using the whole population of India as a scaling factor to compute initial inputs for $Y^I_{1:t_0}$ and $Y^R_{1:t_0}$. This may lead to extremely small proportions, which may in turn yield underestimated outputs from the eSIR model. Changing the total population size from that of India to that of representative (large) cities from the hub states (the states of Kerala, Maharashtra and Karnataka, and the national capital region of Delhi) is a simple but intuitive way to potentially do away with the aforementioned underestimation. We note that this substantially reduces the width of the credible intervals (see Table 2, case-clustering). In yet a third scenario, we hypothesized that the prior mean of R_0 is set at 3.0 or 4.0 instead of 2.0 (i.e., a single infected individual would infect 3 or 4 susceptible individuals, on average, instead of 2). In most of our analyses (Table 2), the posterior mean for R_0 was seen to be between 1.8 and 2.4, irrespective of whether a higher/lower starting (prior) mean was used. We observe that a prior mean of 4.0 for R_0 sways the posterior R_0 estimate substantially (posterior mean 3.38). As more data accumulate, we will expect the effect of the prior on the posterior estimates to diminish. The posterior distributions of the prevalence in each

compartment and latent proportions under these changing scenarios are available in Appendix Figure 8.

In summary, these sensitivity analysis scenarios did not appreciably change our conclusions in broad qualitative terms, though the exact quantitative projections for case-counts are quite sensitive to such choices. We note that the estimate of basic reproduction number R_0 is more robust to underreporting issues as counts in all compartments of our eSIR model are assumed to be underreported. Since underreported case-counts affect all our hypothetical intervention scenarios in a similar way, the relative comparison of interventions and the associated conclusion remain valid in a qualitative sense. In all cases, our confidence in these projections decreases markedly the farther into the future we try to forecast. It is extremely important to update these models as new data arise.

3.5. Model Calibration

To check the calibrating properties of our model, we truncated the data to certain dates and tried assessing the quality of the case-count predictions with essentially adding one more week of data for predicting active cases at a future date (Table 3 and Appendix Figure 9). We do notice the projected case-counts change significantly with more data and improve (become closer to the observed) with more data. Our projections always underestimate the observed counts. This phenomenon is also due to more testing being done each week. However, the observed number of infected cases is always within the 95% prediction credible interval provided by our model. This again reveals the large uncertainty in our predictions.

4. Discussion

Our projections using current daily data on case counts until April 14 in India show that the lockdown, if implemented correctly in the end, has a high chance of reducing the number of COVID-19 cases in the short term and buying India invaluable time to prepare its healthcare and disease monitoring system. In the long-term, we need to have some measures of suppression in place after the lockdown is lifted to prevent a massive surge in the number of cases that can quickly overwhelm an already over-stretched Indian healthcare system resulting in increased fatalities. Specific vulnerable populations will be at higher risk of severity and fatality from COVID-19 infection: older persons and persons with pre-existing medical conditions (e.g., high blood pressure, heart disease, lung disease, cancer, diabetes, immunocompromised persons) (Coronavirus Disease 2019 (*COVID-19*): *Are You at Higher Risk for Severe Illness?*, 2020; Q&A on Coronaviruses (COVID-19), 2020). Appendix Table 1 provides a description of the approximate number of individuals in these high-risk categories in India. Beyond the fragile population characterized by health and economic indicators, we must remember that healthcare workers and first responders at the front line of this pandemic are amongst the most vulnerable (C. Wang et al., 2020). Though we have focused on forecasting and modeling of COVID-19 case-counts in this paper, we recognize that this is only one component of the problem. Long term surveillance and management of the COVID-19 crisis is needed with not just public health in mind but also to take care of the economic, social, and psychological impact that it will have on the people of India.

4.1. Limitations

Our statistical modeling and forecasts are not without limitations. We have limited number of data points and a wide time-window to extrapolate for the long-term forecasts. The uncertainty in our predictions is large due to many unknowns arising from model assumptions, population demographics, the number of COVID-19 diagnostic tests administered per day, testing criteria, accuracy of the test results, and heterogeneity in implementation of different government-initiated interventions and community-level protective measures across the country. We have neither accounted for age-structure, contact patterns or spatial information to finesse our predictions (Klein et al., 2020; Mandal et al., 2020; Singh & Adhikari, 2020) nor considered the possibility of a latent number of true cases, only a fraction of which are ascertained and observed (C. Wang et al., 2020). Increase in frequency and scale of testing, and community transmission of the SARS-CoV-2 virus may lead to a spike in a single day and that can shift the projection curve substantially upwards. COVID-19 hotspots in India are not uniformly spread across the country, and state-level forecasts (S. Das, 2020) may be more meaningful for state-level policymaking. We are assuming that the implementation and effects of public health interventions and policies are the same everywhere in India by treating India as a homogeneous unit.

The eSIR model treats the entire group of people within a single compartment as homogenous and exchangeable. We also assume that all subjects who were not infected are susceptible. Certainly, this overlooks the possibilities of people moving between states and different subsets of infected and susceptible populations having greater or lesser likelihood of coming into contact with one another. To account for all such potential factors, we need more nuanced modeling. One potential way is to break up the interaction component into further sub-compartments; however, sparsity of current data in each sub-compartment is an issue. Another way that has been pursued in a recent study is to inform the SIR modeling via introducing contact networks at the initial stage (Bhattacharyya et al., 2020). However, it is worth noting that any such approach would need more granular and reliable data containing individual details of the confirmed cases including their location and travel history. Even though such data are available from some self-reporting-based repositories (Novel Corona Virus 2019 Dataset, 2020), the quality and detail of the information provided are quite heterogeneous and, thus, how to best utilize such data remains a question. We see the tremendous role of data and its transparency of collection and reporting in finessing our predictions. Accurate and consistent reporting of case counts and deaths due to COVID-19 are extremely critical. Future opportunities for improving our model include incorporating contagion network, age-structure, estimating SEIR model, incorporating test imperfection, and estimating true fatality/death rates. Future research would benefit from easily accessible hospitalization data, accurate recording of deaths in death records, and availability of ecological-level covariate data. Regardless of the caveats in our study, our analyses show the impact and necessity of lockdown and of suppressed activity post-lockdown in India. *Though the exact numerical projections are perhaps far from the truth, the qualitative inference on the relative effects of the interventions are still valuable and valid since all projections are subject to similar biases*.

One ideological limitation of considering only the perspective of controlling COVID-19 transmission in our model is the inability to account for excess deaths due to other causes during this period (chronic disease and mental health related diseases in particular), or the flexibility to factor in reduction in mortality/morbidity due to some other infectious or flu-like illnesses, traffic accidents or health benefits of reduced air pollution levels. A more expansive framework of a cost-benefit analysis is needed as we gather more data and build an integrated landscape of changes in population attributable risks due to various disease categories.

4.2. Testing

A reviewer of the paper suggested giving a sense of the testing data from India and how that may affect our conclusions. India's priorities in testing have changed multiple times over the last few weeks. On March 17, India proposed testing all people who recently traveled internationally and developed symptoms (fever, cough, difficulty in breathing etc.) of COVID-19 within 14 days of return, or all symptomatic contacts of laboratory confirmed positive cases, or all symptomatic health care workers managing respiratory distress. On March 20, India revised this testing strategy to include all symptomatic healthcare workers, all hospitalized patients with Severe Acute Respiratory Illness (fever *and* cough and/or shortness of breath), and asymptomatic direct and high-risk contacts of a confirmed case to be tested once between day 5 and day 14 of coming into contact. More recently, the testing strategy was again revised on April 9 to additionally include testing of all symptomatic people in hotspots/cluster and in large migration gatherings/evacuee centers. These testing strategies are obtained from the Indian Council of Medical Research (https://www.icmr.gov.in).

We looked a little deeper into the issue of testing bias using data from Our World in Data (Hasell et al.). Appendix Figure 10a shows that, even though the number of tests in India has increased in recent days, the proportion of daily discovered positive cases still remain stable (staying at about 4%) and do not show an obvious increasing trend yet like in other countries such as the US and the UK. We also plot Iceland and South Korea on this Figure as they have performed remarkably in administering a large number of tests per detected case and serve as examples for the world. We also looked at the proportion of the population tested in 61 countries around the world (Appendix Figure 10b). While most advanced countries have tested around 1–3% of the population, India has tested roughly 0.06% population and it will take weeks if not months for India to reach testing 1–3% of the population. In the absence of rapid and large-scale testing, informative proxies or surrogates can be tracked through syndromic surveillance, temperature reporting, monitoring hospital admissions and medical claims due to respiratory and flu like illnesses. This additional information will strengthen the prediction models. In absence of these models we rely on sensitivity analysis for underreporting as presented in Table 2.

4.3. Our data science product

Finally, in our strong commitment to reproducibility and dissemination of our research, we have made the code for our predictions available at GitHub (https://github.com/umich-cphds/cov-ind-19) and created an interactive and dynamic *RShiny* app (covind19.org) to visualize

the observed data and create predictions under hypothetical scenarios with quantification of uncertainties. These forecasts are updated daily as new data come in. We hope these products will remain our contribution and service as data scientists during this tragic global catastrophe, and the model and methods will be used to analyze data from other countries.

5. Conclusion

Our epidemiologic and mathematical calculations make a convincing case for enforcing national lockdown in the largest democracy in the world, acting early, before the growth of COVID-19 infections in India starts to accelerate. We observe the public health benefit of extending the lockdown by 3–5 weeks in our projections. Measures of suppression are needed post-lockdown to acquire maximal long-term benefits from the lockdown. We also illustrate the critical role of epidemiologic forecasting in aiding policy decisions through this modeling exercise. We highlight the importance of conducting model checks, sensitivity analysis and uncertainty quantification.

We realize that these draconian public health measures come at a tremendous price to social and economic health that can last for months or even years after the restrictions on social mobility are lifted. Our general message to the public is to proceed with prudence and caution and adhere to effective public health interventions until there are rapid and reliable home testing kits [there is none yet] (Food and Drug Administration, 2020), FDA approved drugs [Solidarity Trial] (World Health Organization, 2020), and a vaccine [5 clinical trials ongoing] (Craven, 2020).

Acknowledgement

The authors will like to thank the University of Michigan Advanced Research Computing Services, in particular Professor Ravi Pendse, for enabling daily updates to our models and allocating us abundant computational resources. We will also like to thank Professor Matthew Fox from the Boston University School of Public Health for his valuable comments on our *RShiny* App. This work originated at a time when the authors in this study team came together to provide their service as quarantined data scientists. We thank our families and friends for their support during these unprecedented times. The initial report made a profound impact through media outlets; we remain grateful to journalists across the world who reported our findings. In the middle of a global pandemic, we found inspiration in the heroic battles by frontline health workers, we found hope in science, we found strength in the power of common people and we found solace in the magic of human kindness. We wish the government and the people of India the very best in their fight against the coronavirus. We are all in this together.

Disclosure Statement

No conflicts of interest to declare. The research was supported by the University of Michigan Precision Health Initiative, University of Michigan School of Public Health, University of Michigan Institute for Health Care Policy and Innovation, Michigan Institute of Data Science and the University of Michigan Rogel Cancer Center.

Appendix

Appendix Table 1.

Proportion of population in specifically vulnerable subgroups at potentially high risk of COVID-19 severity risk in India.

Metric	Number (in millions)†	Year	Source
Uninsured	1,104	2014	Prinja et al. 2019

Metric	Number (in millions)†	Year	Source
Population over 65	92.5	2020 (est.)	CIA World Factbook
Hypertension (men)*	175.7	2015/16	Gupta & Ram 2019
Hypertension (women)*	132.6	2015/16	Gupta & Ram 2019
People with cardiovascular disease*	78.7	2016	Prabhakaran et al. 2018
Population with COPD*	75.9	2016	Lancet 2018
Population with asthma*	45.5	2016	Lancet 2018
Develop cancer by age 75 (men)**	70.3	2018	NICPR
Develop cancer by age 75 (men)**	62.3	2018	NICPR
Population with diabetes (adult)	122.8	-	IDF
Access to inpatient department facilities***	731.4	2012	IMS Institute 2013
Access to outpatient department***	1,104	2012	IMS Institute 2013

† based on 2020 est. of 1.38 billion from UN Department of Economic and Social Affairs

* age-standardized

** risk

*** defined as within 5-kilometer distance of home or work

Abbrev.: COPD, chronic obstructive pulmonary disease; IDF, International Diabetes Federation; NICPR, National Institute of Cancer Prevention and Research

Appendix Table 2.

Intervention landscape of countries severely affected by COVID-19

Country	Date of 1st case¶	Interventions/Lockdown	Crude fatality rate†	Active cases†
China	Nov 17, 2019*	Lockdown in Wuhan on 01/22, extended to neighboring cities in Hubei province on 01/23. Wuhan lockdown to be lifted on 04/08.[1]	4.1%	1,116
South Korea	Jan 19, 2020**	Tested widely for the virus, isolated cases and quarantined suspected cases. Figures indicate that this has helped suppress transmission of the virus. The country appears to have reined in the outbreak without some of the strict lockdown strategies deployed elsewhere in the world.[2]	2.0%	3,125
United States	Jan 20, 2020	On 01/31, restricted travel from China; expanded restrictions to other countries on 02/29. On 03/03, CDC lifted all restrictions on testing. On 03/15, CDC recommends no gatherings of 50 people or more. Stay-at-home directives issued at state-level.[3]	3.7%	431,557
France	Jan 24, 2020	On 03/17, France imposed a nationwide lockdown, prohibiting gatherings of any size and postponing the second round its municipal elections. The lockdown was one of Europe's most stringent. While residents were told to stay home, officials allowed people to go out for fresh air but warned that meeting a friend on the street or in a park would be punishable with a fine.[3]	10.8%	79,279
Germany	Jan 28, 2020	Germany has a National Pandemic Plan, with three stages. In the first stage (containment) health authorities are focusing on identifying contact persons, who are put in personal quarantine and are monitored and tested. In the second stage (protection) the strategy will change to using measures to protect vulnerable persons from becoming infected. The final stage (mitigation) will try to avoid spikes of intensive treatment in order to maintain medical services.[4]	2.2%	76,420
United Kingdom	Jan 31, 2020	Citizens advised to stay at home. Violators to be fined with the exception of a few special circumstances, in order to contain the spread of the disease. The government is supporting and coordinating with research institutes to explore treatment and curative options.[5]	12.7%	61,825

Country	Date of 1st case¶	Interventions/Lockdown	Crude fatality rate†	Active cases†
Italy	Jan 31, 2020	Flights to China suspended and a national emergency declared on 01/31 after two cases confirmed in Rome. Schools and universities closed on 03/04. By 03/09, the entire nation placed under lockdown, with restaurants, bars closing on 03/11 and factories closing on 03/22. All nonessential production halted.[6]	12.7%	96,877
Spain	Feb 1, 2020	State of emergency declared on 03/14, allowing authorities to authorities to confine infected people and ration goods. Originally planned to last until 03/29; has been extended to 04/12. Schools, bars, restaurants and shops selling nonessential items have been shut since March 14 and most of the population is house bound.[7]	10.2%	85,386
Iran	Feb 19, 2020	On 02/28, the Iranian authorities closed schools, canceled Friday prayers and moved to restrict visitors from China. On 03/15, the official leading Iran's response to the new coronavirus acknowledged Sunday that the pandemic could overwhelm health facilities in his country, which is battling the worst outbreak in the Middle East.[8]	6.2%	28,495

¶Date of 1 case data obtained from JHU CSSE time series data on LUV1D-19 (except for China & South Korea)

†Microsoft bing COVID-19 tracker (https://bing.com/covid as of 1:00 PM EST April 10, 2020)

*https://www.livescience.com/first-case-coronavirus-found.html

**https://www.worldometers.info/coronavirus/

1https://www.cnn.com/2020/03/24/asia/coronavirus-wuhan-lockdown-lifted-intl-hnk/index.html

2https://www.npr.ors/sections/soatsandsoda/2020/Q3/26/821688981/how-south-korea-reisned-in-the-outbreak-without-shuttins-everythins-down

3https://www.nytimes.com/article/coronavirus-timeline.html

4https://en.wikipedia.org/wiki/2020_coronavirus_pandemic_in_Germany

5https://www.sov.uk/sovernment/publications/coronavirus-action-plan/coronavirus-action-plan-a-suide-to-what-you-can-expect-across-the-uk

6https://www.axios.com/italy-coronavirus-timeline-lockdown-deaths-cases-2adb0fc7–6ab5–4b7c-9a55-bc6897494dc6.html

7https://www.reuters.com/article/us-health-coronavirus-spain/shellshocked-spain-reports-record-832-new-coronavirus-deaths-idUSKBN21F0NR

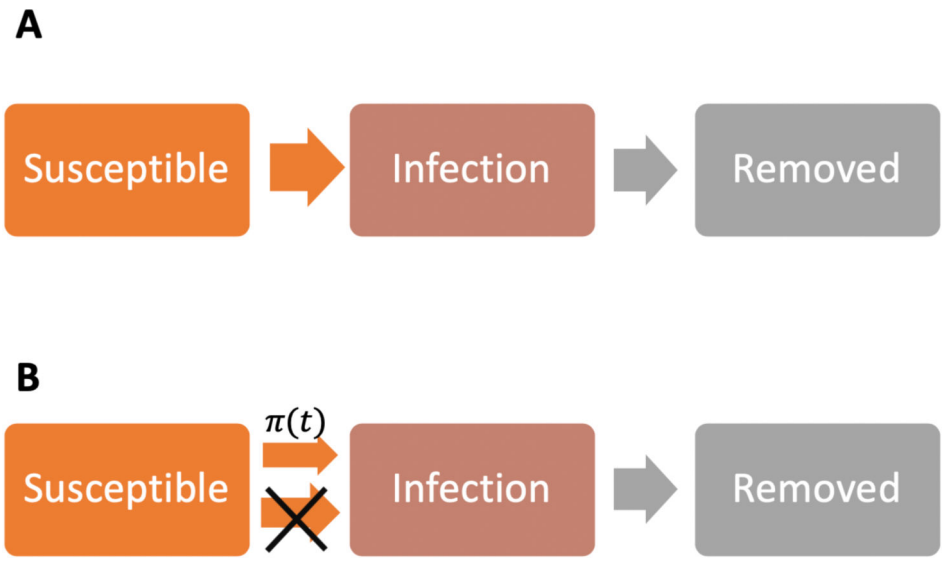
8https://www.nytimes.com/aponline/2020/03/15/world/middleeast/ap-ml-virus-outbreak-mideast.html?searchResultPosition=10

Appendix Table 3.

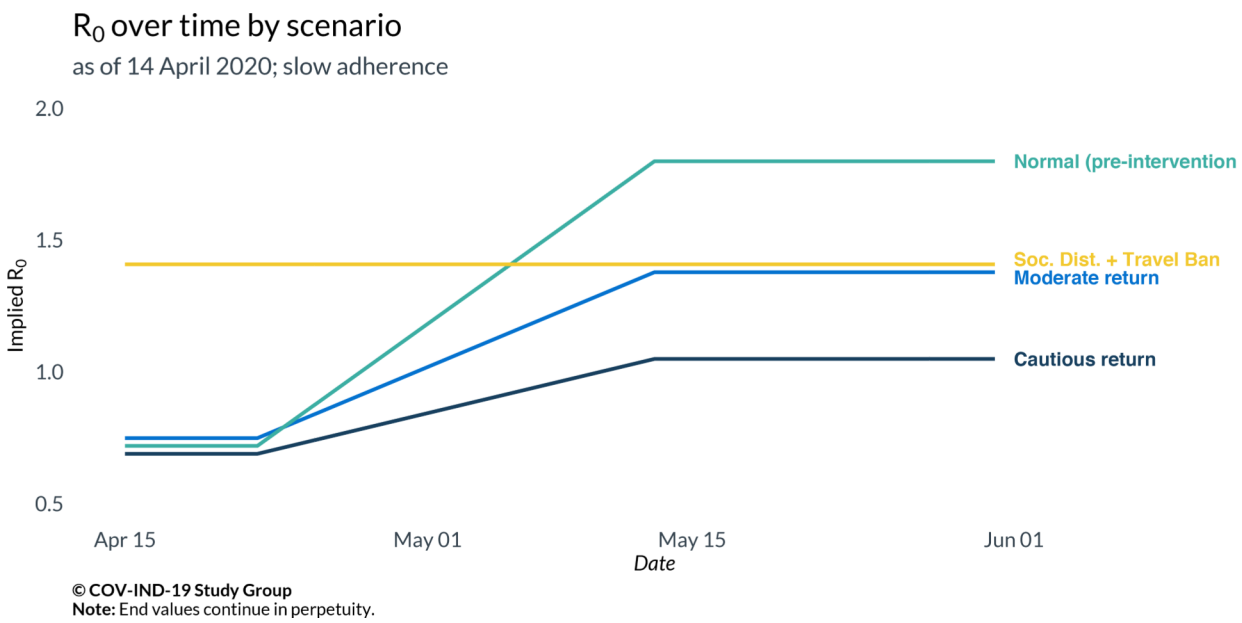
Comparison of types of infectious disease models, specifically those used to study COVID-19 in India

Model type	Reference	Research question	Strengths	Weaknesses
Exponential model	Gupta and Shankar (2020)	Provide estimate of the infected population using death counts	Simple model helpful for scant data; modeled epidemic hotspots separately	Not accounted for population demographics (limited by data), non-pharmaceutical (NP) intervention effects; requires infection fatality rate
Poisson log-linear model	Das (2020)	Short-term prediction of future case counts; estimate R_0	Simple model helpful for short- & medium-term forecasts using scant data; accounted for quadratic effect of time	Not accounted for population demographics (limited by data), hotspots, NP intervention effects; surveillance bias
Autoregressive–moving-average model	Deb and Majumdar (2020)	Analyze the trend pattern of incidence; estimate R_0	Accounted for quadratic effect of time, lockdown effect; captured time dependence incidence pattern	Not accounted for population demographics (limited by data), hotspots; surveillance bias

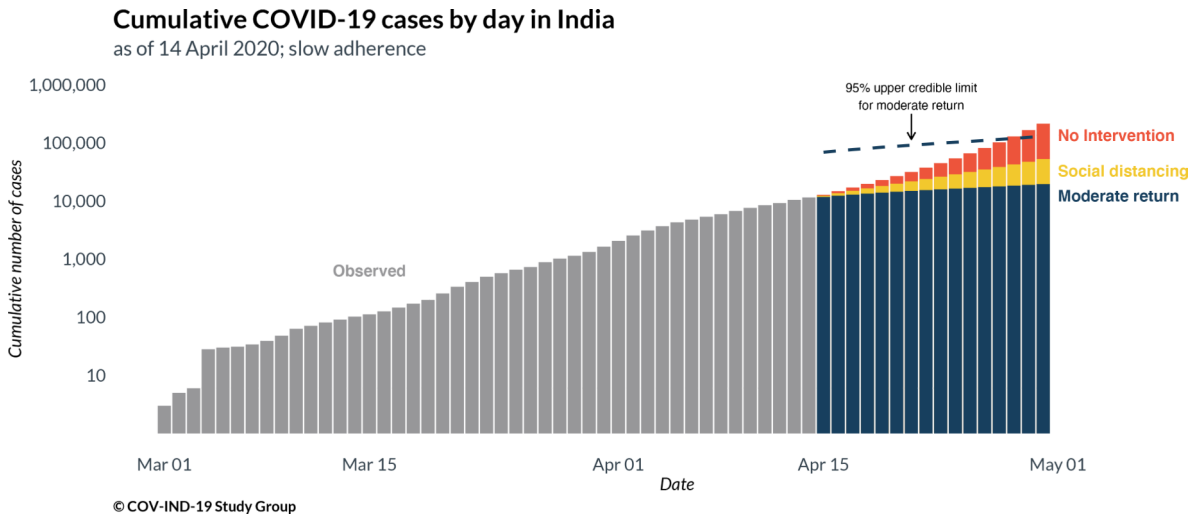
Model type	Reference	Research question	Strengths	Weaknesses
Susceptible-infected-recovered (SIR) model	Ranjan (2020)	Long-term prediction of future case counts; estimate R_0	Classical epidemiologic model used; accounted for social distancing effects	Not accounted for population demographics (limited by data), hotspots; surveillance bias; used first few weeks of data
SIR model	Dhanwant and Ramanathan (2020)	Long-term prediction of future case counts	Classical epidemiologic model used; split observed data into training and test data; training data used to learn the transmission rate; incorporated lockdown effect	Not accounted for population demographics (limited by data), hotspots; surveillance bias; lockdown training data not used to learn about transmission rate under lockdown
Age-structured SIR model	Singh and Adhikari (2020)	Study progress of the disease and impact of social distancing measures; estimate R_0	Extended epidemiologic model accounting for age distribution, social contact, social distancing effect	Not accounted for other population demographics (limited by data), hotspots; surveillance bias; complex model given the scant count data and spotty individual-level data
Susceptible-Exposed-Infectious-Recovered (SEIR) model	Mandal et al. (2020)	Identify NP intervention strategies that can help control the outbreak	Extended epidemiologic model with an added compartment for quarantine; accounted for other NP interventions, and connectivity between two places	Not accounted for population demographics (limited by data); surveillance bias; complex model given the scant count data; lockdown effect not studied; studied four cities only
Expanded SEIR model	Chatterjee et al. (2020)	Assess the impact on healthcare resources; study the effect of different NP interventions	Extended epidemiologic model with added sub-compartments for quarantined, recovered and death; accounted for different NP interventions; accounted for age groups	Not accounted for other population demographics (limited by data), hotspots; surveillance bias; complex model given the scant count data and spotty individual-level data; hospitalization-related parameters based on UK data; lockdown effect not studied
Expanded SEIR model	Senapati et al. (2020)	Assess the effect of different NP interventions; estimate R_0	Extended epidemiologic model with added sub-compartments for asymptomatic cases, quarantined, hospitalized, recovered and death	Not accounted for other population demographics (limited by data), hotspots; surveillance bias; complex model given the scant count data; lockdown effect not studied
Expanded SEIR model	Sardar et al. (2020)	Assess long-term effect of 21-day lockdown; estimate R_0	Extended epidemiologic model with added sub-compartments for asymptomatic cases, lockdown, hospitalized, recovered and death; accounted for transmission variability between symptomatic and asymptomatic groups; modeled hotspots and overall India	Not accounted for other population demographics (limited by data); surveillance bias; complex model given the scant count data



Appendix Figure 1:
The SIR model with (A) or without (B) considering human intervention by introducing a transmission rate modifier $\pi(t)$.

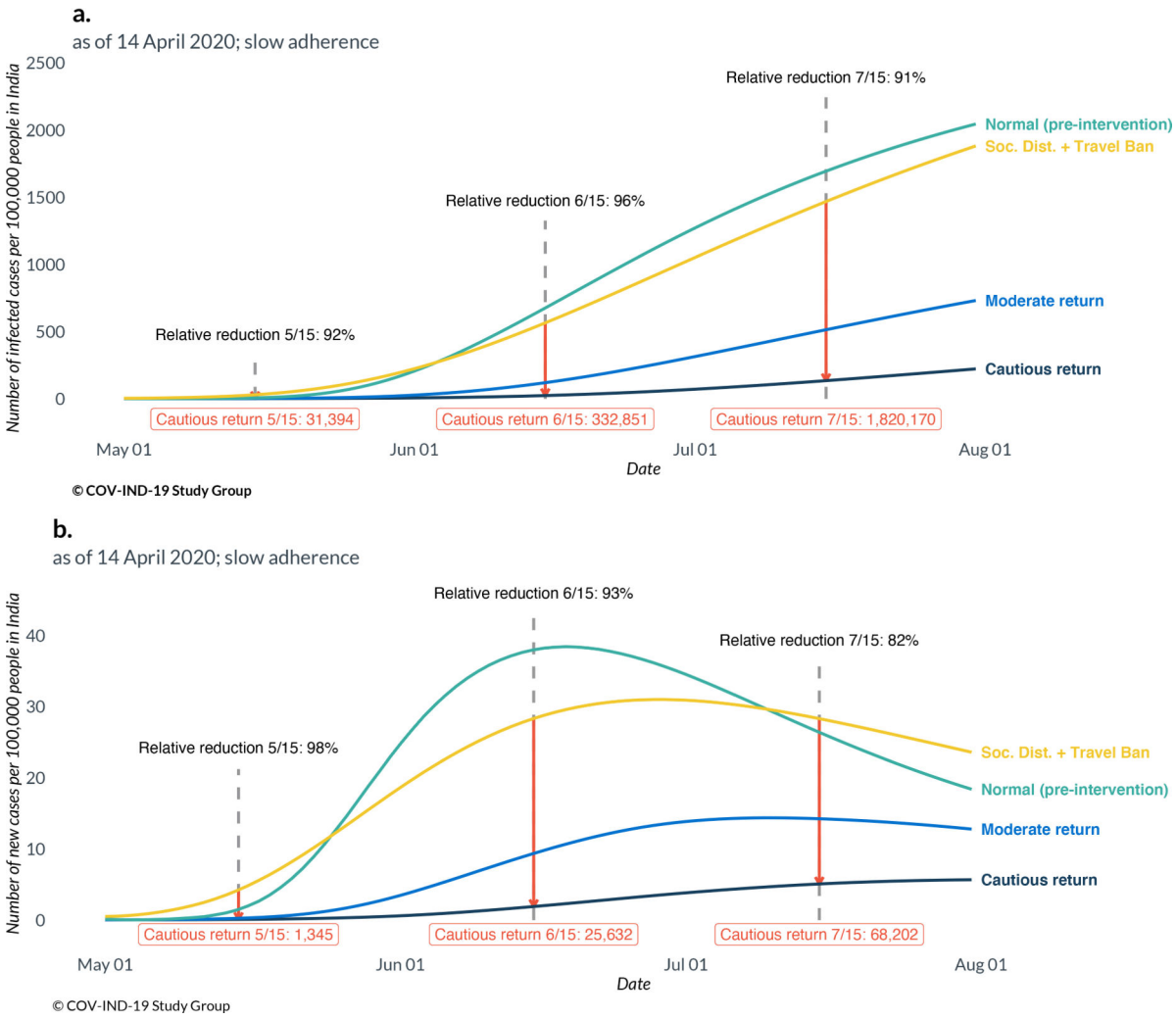


Appendix Figure 2.
Implied R_0 schedules corresponding to the hypothetical scenarios under slow adherence.



Appendix Figure 3.

Short-term daily growth in cumulative case counts in India assuming a 2- week delay in people's adherence to restrictions. Observed data are shown for days up to April 14. Predicted future case counts for April 15 until April 30 are based on observed data until April 14 using the eSIR model.



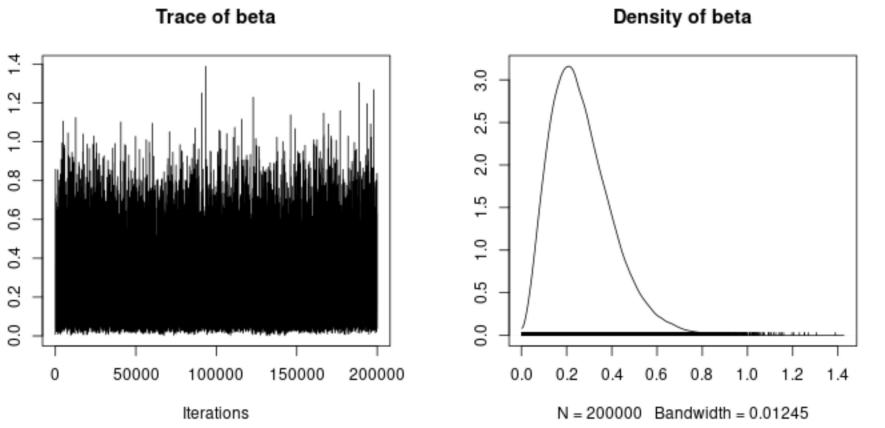
Appendix Figure 4.

Long-term daily growth in case counts in India per 100,000 people assuming a 2-week delay and how that is affected by different non-pharmaceutical intervention strategies. Predicted cumulative (a) and incident (b) case counts from April 30 to July 31 from the eSIR model are shown, based on observed data until April 14.

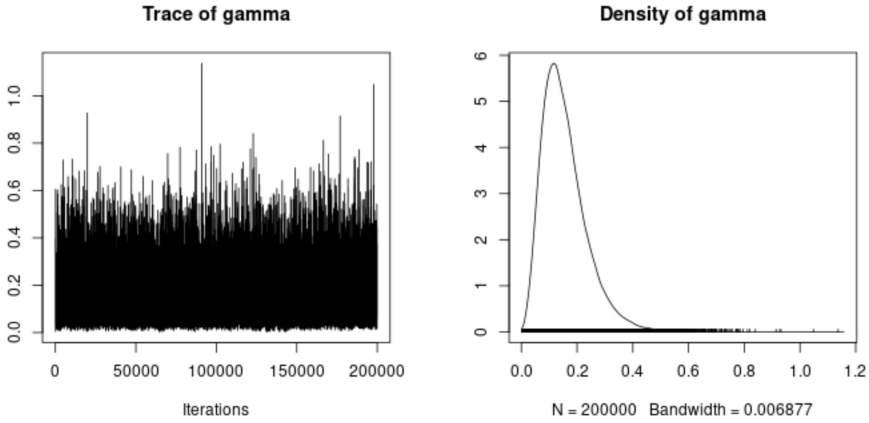
a. Predicted number of COVID-19 cases in India under hypothetical scenarios

b. Predicted number of new COVID-19 cases in India under hypothetical scenarios

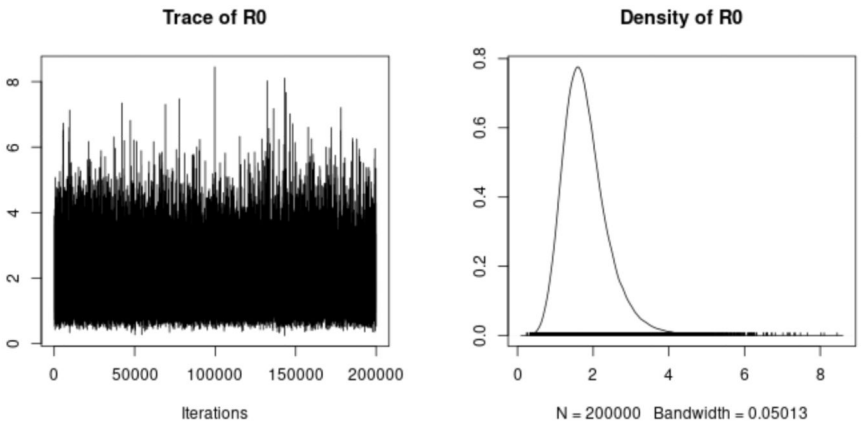
a.



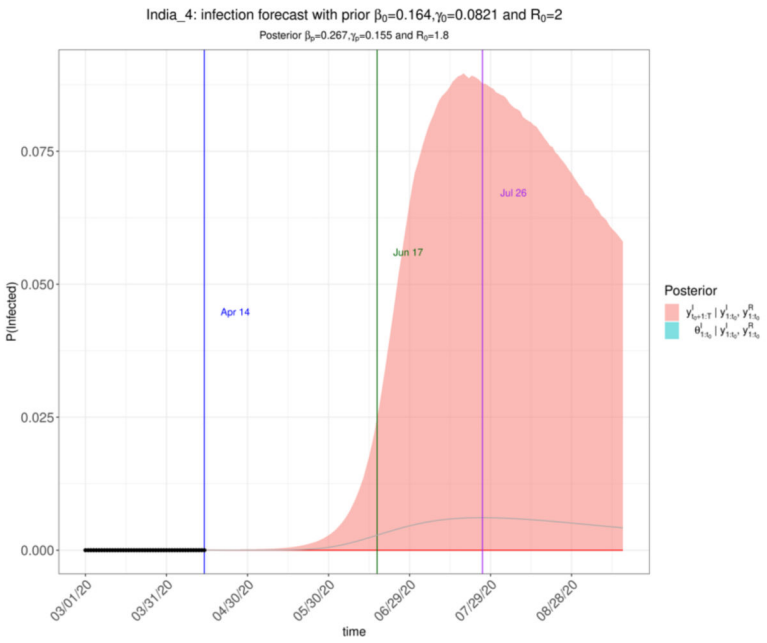
b.



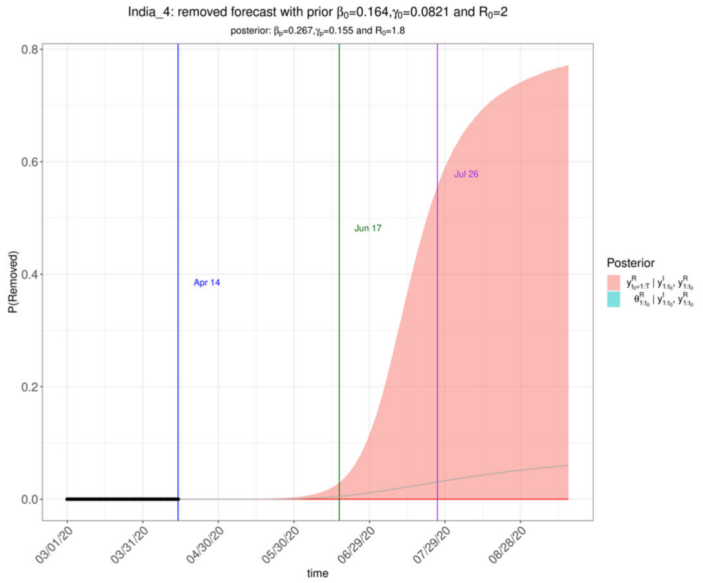
c.



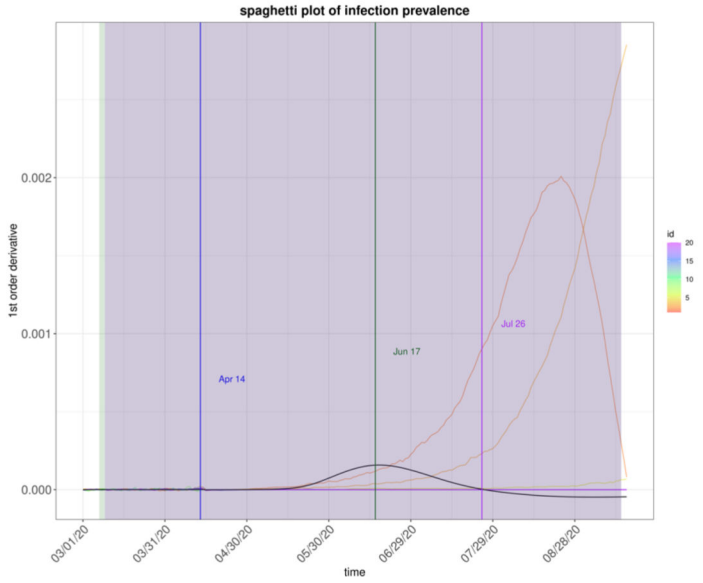
d.



e.



f.



Appendix Figure 5.

Trace plots and posterior density plots for the underlying model parameters β (**a**), γ (**b**), and R_0 (**c**), posterior distributions for the predictions Y and the latent proportions θ for the I (**d**) and R (**e**) compartments over time, and estimates and posterior distribution of the daily prevalence of active cases over time or $\frac{d\theta_t^I}{dt}$ (**f**). These plots correspond to the 21-day lockdown with moderate return scenario under quick adherence.

a. Trace plots and posterior density plots for β

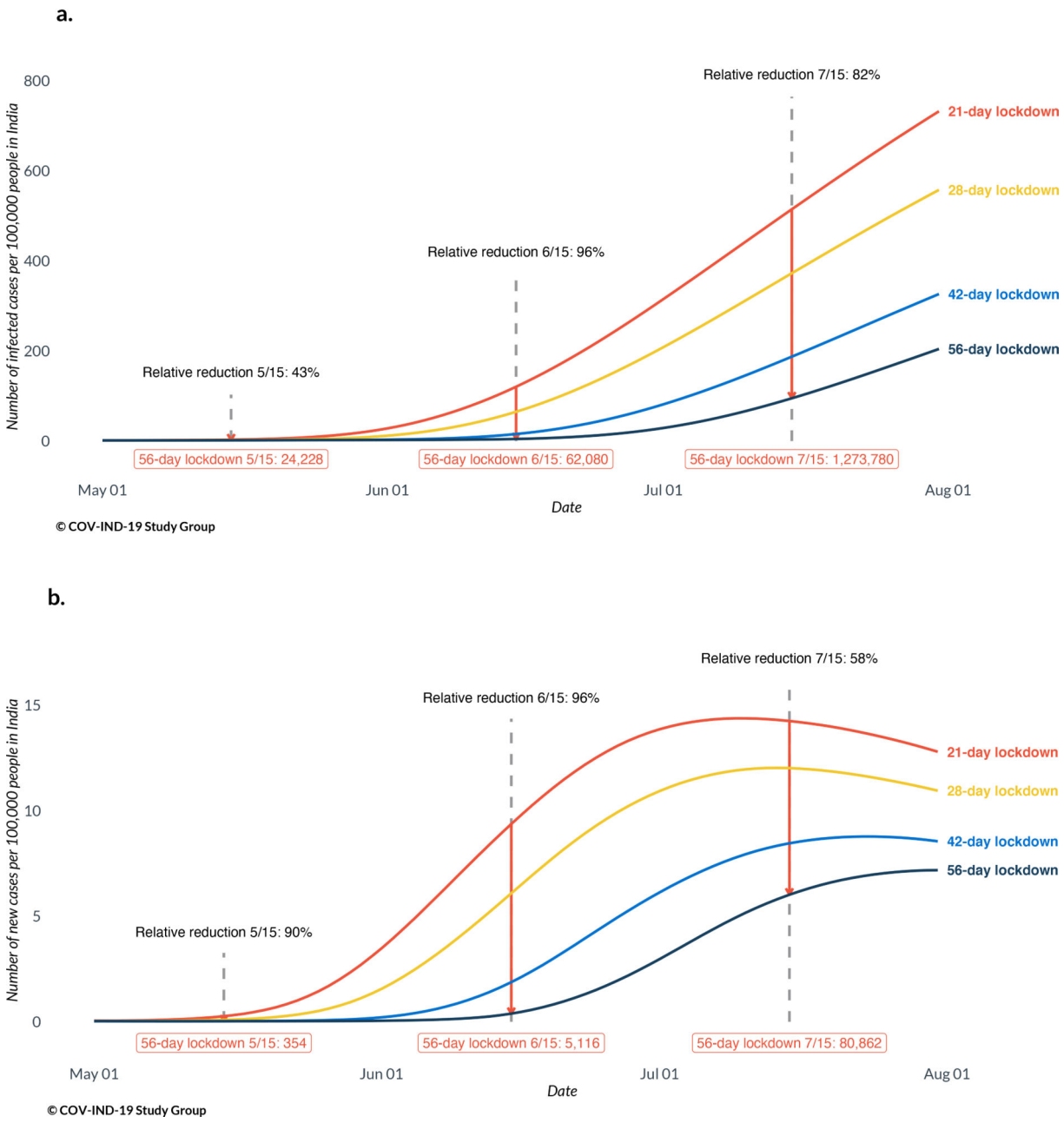
b. Trace plots and posterior density plots for γ

c. Trace plots and posterior density plots for R_0

d. Posterior distribution for the predictions Y and the latent proportions θ for the I (infected) compartment

e. Posterior distribution for the predictions Y and the latent proportions θ for the R (removed) compartment

f. Estimates and posterior distribution of the daily prevalence of active cases over time or $\frac{d\theta_t^I}{dt}$

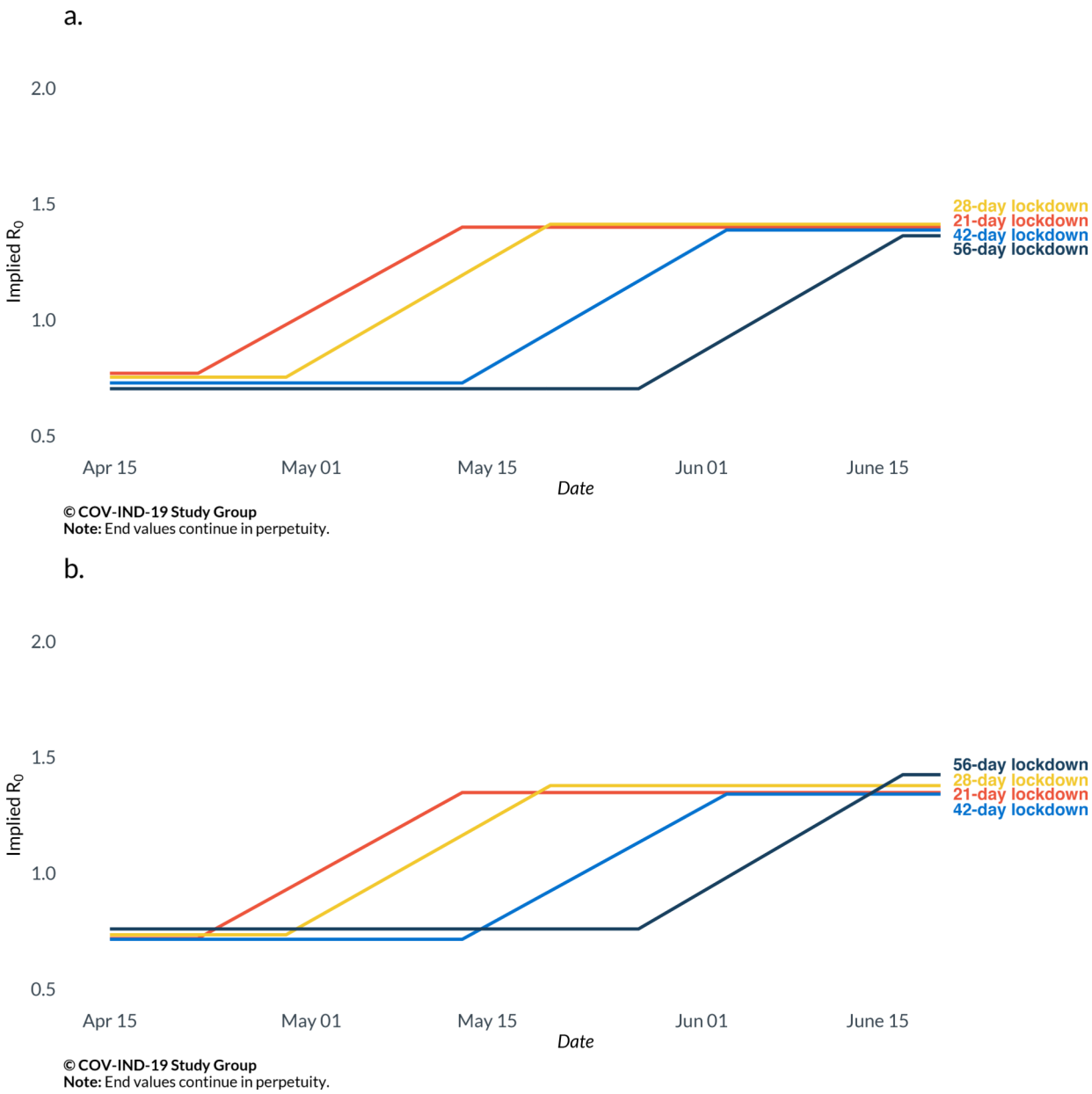


Appendix Figure 6.

Cumulative (a) and incidence (b) graphs for forecasting models assuming a 2-week delay under 21-, 28-, 42-, and 56-day lockdown scenarios using observed data through April 14.

a. Predicted number of COVID-19 infections under varying lockdown lengths

b. Predicted number of daily COVID-19 infections under varying lockdown lengths



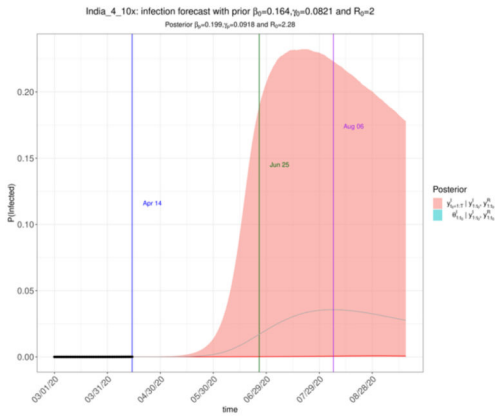
Appendix Figure 7.

Implied R_0 schedules corresponding to quick and slow adherence for the hypothetical lockdown duration scenarios.

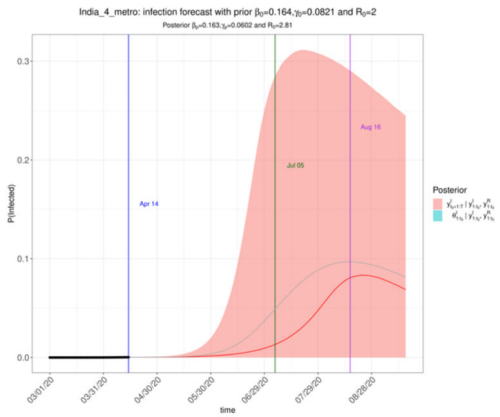
a. R_0 over time by scenario

b. R_0 over time by scenario

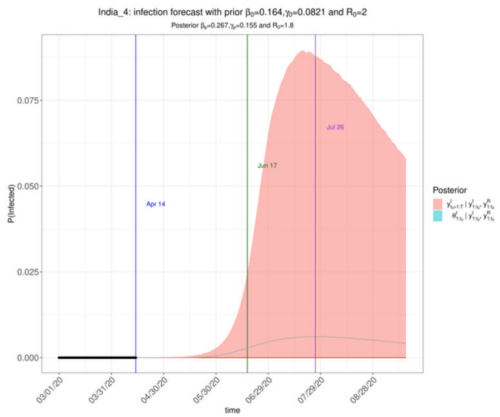
a.



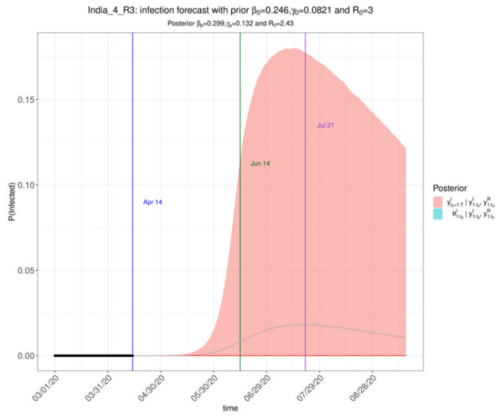
b.



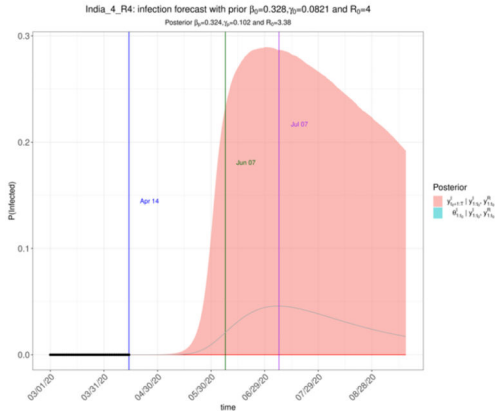
c.



d.



e.



Appendix Figure 8.

Posterior distributions of the projected case-counts and latent proportions under sensitivity scenarios.

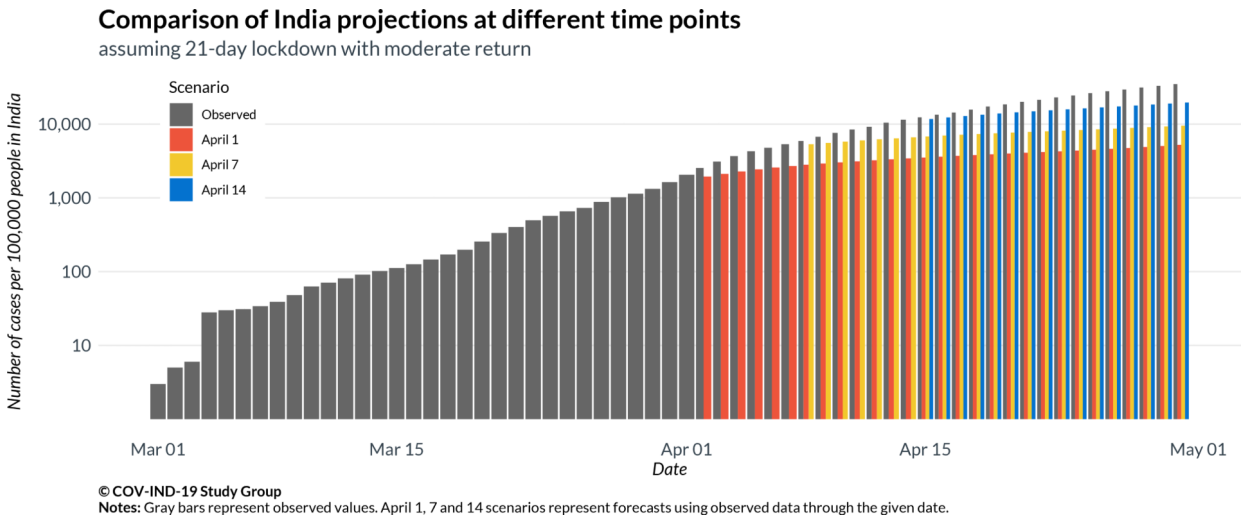
a. Scenario with 10 times the number of reported cases (e.g., under-reporting)

b. Scenario using metro population (e.g., to mimic case-clustering)

c. Scenario with prior mean of $R_0 = 2$

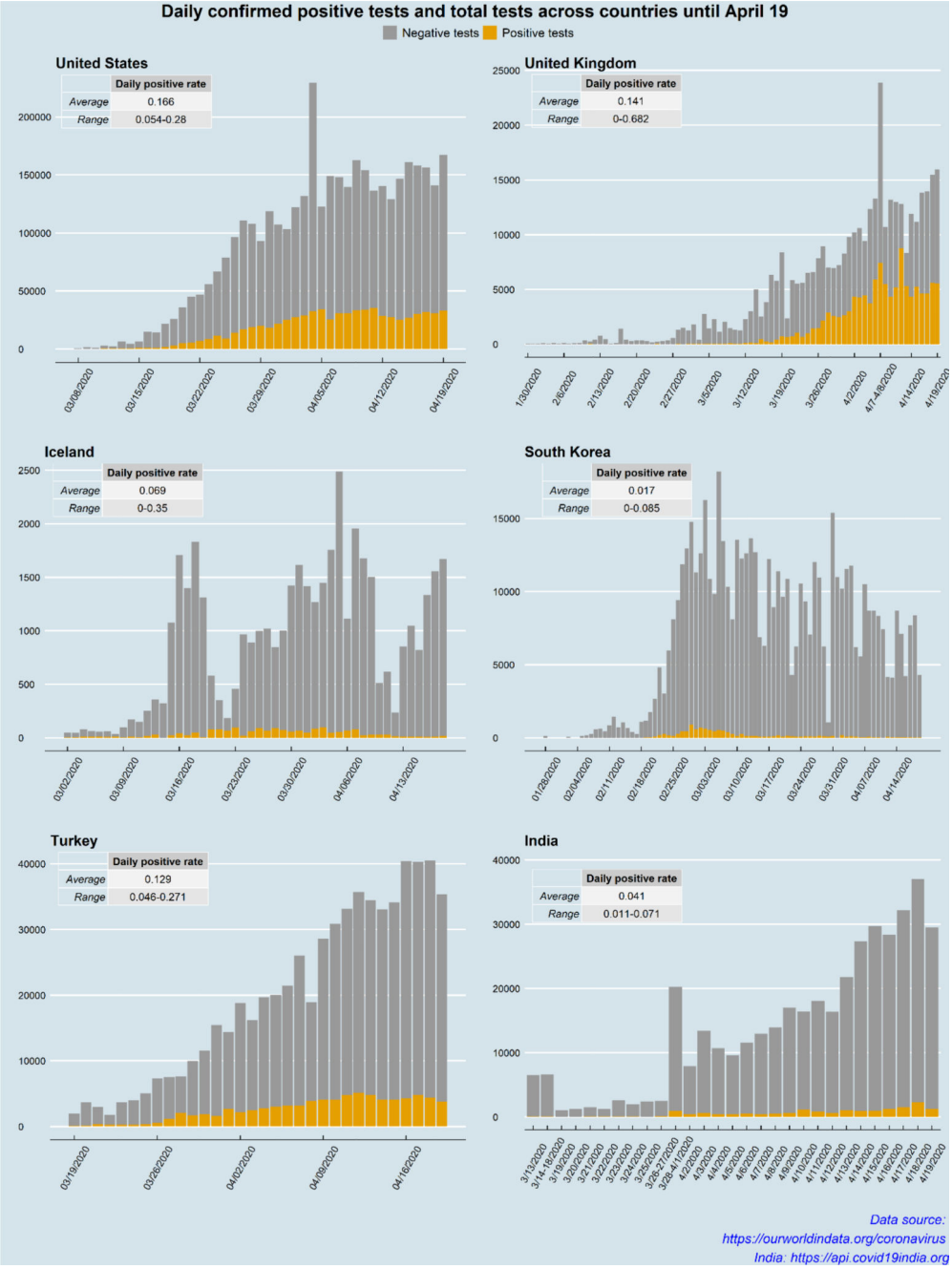
d. Scenario with prior mean of $R_0 = 3$

e. Scenario with prior mean of $R_0 = 4$

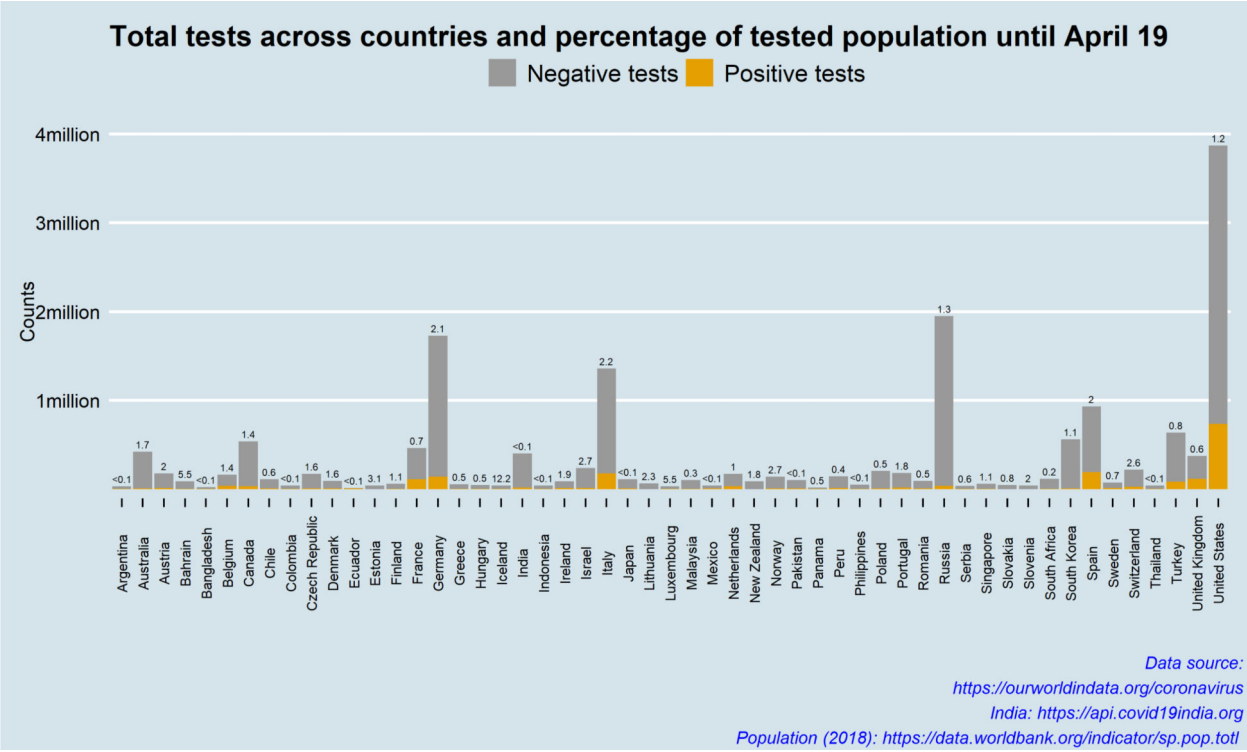


Appendix Figure 9.

Model Calibration: Relative comparison of predictions using observed data up to a certain date (April 1, 7 and 14). Observed data (gray) is provided through April 30.



Appendix Figure 10a.
Daily testing patterns in selected countries



Appendix Figure 10b.

Testing numbers and proportions for 61 countries around the world affected by COVID-19

References

Bhattacharyya R, Mohammed S, Baladandayuthapani V, Banerjee S, & Nanda U (2020, 4). Regional Contact Networks and the Pandemic Spread of COVID-19 in India. Medium.

Chatterjee K, Chatterjee K, Kumar A, & Shankar S (2020). Healthcare impact of COVID-19 epidemic in India: A stochastic mathematical model. Medical Journal Armed Forces India. 10.1016/j.mjafi.2020.03.022.

Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Qiu Y, Wang J, Liu Y, Wei Y, Xia J, Yu T, Zhang X, & Zhang L (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. The Lancet, 395(10223), 507–513. 10.1016/S0140-6736(20)30211-7.

Coronavirus Disease 2019 (COVID-19): Are You at Higher Risk for Severe Illness? (2020). Centers for Disease Control and Prevention https://www.cdc.gov/coronavirus/2019-ncov/specific-groups/high-risk-complications.html.

covid19india. (2020). India COVID-19 Tracker. https://www.covid19india.org.

Craven J (2020, 4 10). COVID-19 vaccine tracker. Regulatory Affairs Professionals Society. https://www.raps.org/news-and-articles/news-articles/2020/3/covid-19-vaccine-tracker.

Das P (2020, 4 1). “Epidemiologic models show we need aggressive measures in the early phase...lockdown buys us time.” The Times of India. https://timesofindia.indiatimes.com/blogs/melange/epidemiologic-models-show-we-need-aggressive-measures-in-the-early-phase-lockdown-buys-us-time/.

Das S (2020). Prediction of COVID-19 disease progression in India: Under the effect of national lockdown. ArXiv. http://arxiv.org/abs/2004.03147.

Deb S, & Majumdar M (2020). A time series method to analyze incidence pattern and estimate reproduction number of COVID-19. ArXiv. http://arxiv.org/abs/2003.10655.

Dhanwant JN, & Ramanathan V (2020). Forecasting COVID 19 growth in India using Susceptible-Infected-Recovered (S.I.R) model. ArXiv. http://arxiv.org/abs/2004.00696.

Dong E, Du H, & Gardner L (2020). An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases 10.1016/S1473-3099(20)30120-1.

Ellis-Petersen H (2020, 3 25). Overcome by anxiety: Indians in lockdown many can ill afford. The Guardian. https://www.theguardian.com/world/2020/mar/25/overcome-by-anxiety-india-in-lockdown-many-can-ill-afford.

Food and Drug Administration. (2020, 3 20). Coronavirus (COVID-19) Update: FDA Alerts Customers About Unauthorized Fraudulent COVID-19 Tests. https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-alerts-consumers-about-unauthorized-fraudulent-covid-19-test-kits.

Ghosal D (2020, 3 23). India faces spike in coronavirus cases, says study, in test for health system. Reuters. https://www.reuters.com/article/health-coronavirus-india-estimates/india-faces-spike-in-coronavirus-cases-says-study-in-test-for-health-system-idINKBN21A32Q.

Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, Liu L, Shan H, Lei C, Hui DSC, Du B, Li L, Zeng G, Yuen K-Y, Chen R, Tang C, Wang T, Chen P, Xiang J, … Zhong N (2020). Clinical characteristics of coronavirus disease 2019 in China. New England Journal of Medicine, 382(18), 1708–1720. 10.1056/NEJMoa2002032. [PubMed: 32109013]

Gupta S, & Shankar R (2020). Estimating the number of COVID-19 infections in Indian hotspots using fatality data. ArXiv. http://arxiv.org/abs/2004.04025.

Hasell J, Ortiz-Ospina E, Mathieu E, Ritchie H, Beltekian D, Macdonald B, & Roser M (n.d.). Our World in Data COVID-19 Testing dataset. Our World in Data. https://ourworldindata.org/covid-testing.

Hospital beds (per 1,000 people). (2020). The World Bank. https://data.worldbank.org/indicator/SH.MED.BEDS.ZS.

Indian Council of Medical Research. (2020). SARS-CoV-2 (COVID-19) Testing: Status Update 14 April 2020 9:00 PM IST.

Johns Hopkins University Center for Systems Science and Engineering. (n.d.). JHU CSSE COVID-19 Github. https://github.com/CSSEGISandData/COVID-19.

Klein E, Lin G, Tseng K, Schueller E, Kappor G, & Laxminarayan R (2020). COVID-19 for India Updates. https://cddep.org/wp-content/uploads/2020/03/covid19.indiasim.March23-2-4.pdf.

Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KSM, Lau EHY, Wong JY, Xing X, Xiang N, Wu Y, Li C, Chen Q, Li D, Liu T, Zhao J, Liu M, … Feng Z (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. New England Journal of Medicine, 382(13), 1199–1207. 10.1056/NEJMoa2001316. [PubMed: 31995857]

Lin X (2020). Analysis of 25,000 lab-confirmed COVID-19 cases in Wuhan: Epidemiological characteristics and non-pharmaceutical intervention effects. https://cdn1.sph.harvard.edu/wp-content/uploads/sites/21/2020/03/COVID-19-03-16-2020-Lin.pdf.

Mandal S, Bhatnagar T, Arinaminpathy N, Agarwal A, Chowdhury A, Murhekar M, Gangakhedkar R, & Sarkar S (2020). Prudent public health intervention strategies to control the coronavirus disease 2019 transmission in India: A mathematical model-based approach. Indian Journal of Medical Research. 10.4103/ijmr.IJMR_504_20.

Microsoft bing COVID-19 tracker. (n.d.). Microsoft Corporation. Retrieved March 19, 2020, from https://www.bing.com/covid.

Mkhatshwa T, & Mummert A (2010). Modeling super-spreading events for infectious diseases: Case study SARS. ArXiv. http://arxiv.org/abs/1007.0908.

Noronha G (2020, 4 10). India could see a reduction in the number of coronavirus cases by next week: Study. The Economic Times https://economictimes.indiatimes.com/industry/healthcare/biotech/healthcare/india-couldsee-a-reduction-in-the-number-of-coronavirus-cases-by-next-week-study/articleshow/75086386.cms?from=mdr.

Novel Corona Virus 2019 Dataset. (2020). Kaggle. https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset.

Oke J, & Henegan C (2020, 4). Global COVID-19 Case Fatality Rates. CEBM Research.

Pan A, Liu L, Wang C, Guo H, Hao X, Wang Q, Huang J, He N, Yu H, Lin X, Wei S, & Wu T (2020). Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. JAMA 10.1001/jama.2020.6130.

Plummer M, Stukalov A, & Denwood M (n.d.). rjags: Bayesian Graphical Models using MCMC. https://cran.r-project.org/web/packages/rjags/index.html.

Q&A on coronaviruses (COVID-19). (2020). World Health Organization. https://www.who.int/news-room/q-a-detail/q-a-coronaviruses.

Ranjan R (2020). Predictions for COVID-19 outbreak in India using epidemiological models. MedRxiv. 10.1101/2020.04.02.20051466.

Ray D, Bhattacharyya R, Wang L, Salvatore M, Mohammed S, Halder A, Zhou Y, Song P, Purkayastha S, Bose D, Banerjee M, Baladandayuthapani V, Ghosh P, & Mukherjee B (2020, 3 21). Predictions and role of interventions for COVID-19 outbreak in India. Medium https://medium.com/@covind_19/predictions-and-role-of-interventions-for-covid-19-outbreak-in-india-52903e2544e6.

Root A (2020, 4). Nearly a third of Americans with COVID-19 are hospitalized. Here are the latest numbers. Barron's.

Ryu S, & Chun BC (2020). An interim review of the epidemiological characteristics of 2019 novel coronavirus. Epidemiology and Health, 42, e2020006 10.4178/epih.e2020006. [PubMed: 32023775]

Salvatore M, Ray D, Du J, Wang L, Das S, Kleinsasser M, Rix A, Barker D, & Mukherjee B (2020, 4). Unlocking the 40-day national lockdown in India: There is no magic key. Medium. https://medium.com/@covind_19/unlocking-the-40-day-national-lockdown-in-india-there-is-no-magic-key-de4e43177cb4.

Salvatore M, Wang L, Bhattacharyya R, Purkayastha S, Mohammed S, Halder A, Barker D, Kleinsasser M, Rix A, Banerjee M, Baladandayuthapani V, Ray D, & Mukherjee B (2020, 4 3). Historic 21-day lockdown, predictions for lockdown effects and the role of data in this crisis of virus in India. Medium. https://medium.com/@covind_19/historic-lockdown-prediction-models-to-study-lockdown-effects-and-the-role-of-data-in-the-crisis-a0afeeec5a6.

Sardar T, Nadim SS, & Chattopadhyay J (2020). Assessment of 21 days lockdown effect in some states and overall India: A predictive mathematical study on COVID-19 outbreak. ArXiv. http://arxiv.org/abs/2004.03487.

Senapati A, Rana S, Das T, & Chattopadhyay J (2020). Impact of intervention on the spread of COVID-19 in India: A model based study. ArXiv. http://arxiv.org/abs/2004.04950.

Sindhu J, Reddy KK, Satyanarayana N, Devaraya S, & Fathima A (2019). Hospital utilization statistics: Thirty-five year trend analysis, a measure of operational efficiency of a tertiary care teaching institute in South India. IOSR Journal of Dental and Medical Sciences, 18(4), 49–55. 10.9790/0853-1804024955.

Singh R, & Adhikari R (2020). Age-structured impact of social distancing on the COVID-19 epidemic in India. ArXiv. https://arxiv.org/abs/2003.12055.

The World Bank. (n.d.). India (2017). https://data.worldbank.org/country/india?view=chart.

Times of India. (2020, 3). Coronavirus: Does India have enough ventilators, hospital beds? The Times of India. https://timesofindia.indiatimes.com/india/coronavirus-does-india-have-enough-ventilators-hospital-beds/articleshow/74822427.cms.

Wang C, Liu L, Hao X, Guo H, Wang Q, Huang J, He N, Yu H, Lin X, Pan A, Wei S, & Wu T (2020). Evolving epidemiology and impact of non-pharmaceutical interventions on the outbreak of coronavirus disease 2019 in Wuhan, China. MedRxiv. 10.1101/2020.03.03.20030593.

Wang L, Zhou Y, He J, Zhu B, Wang F, Tang L, Eisenberg M, & Song PX (2020). An epidemiological forecast model and software assessing interventions on COVID-19 epidemic in China. MedRxiv. 10.1101/2020.02.29.20029421.

WHO Director-General's opening remarks at the media briefing on COVID-19 – 11 March 2020. (2020, 3 11). World Health Organization. https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-themedia-briefing-on-covid-19-−-11-march-2020.

World Bank. (n.d.). Data for India, United States, China. https://data.worldbank.org/?locations=IN-US-CN.

World Health Organization. (2020, 3 27). WHO Director-General's opening remarks at the media briefing on COVID-19 – 27 March 2020. https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19-−-27-march-2020.

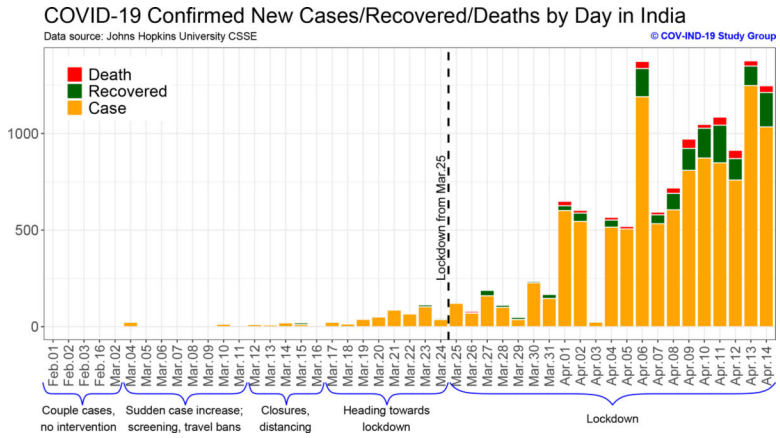


Figure 1.

Description of the cases, recovered and fatalities in India with landmark policy/ recommendations. Data used up to April 14.

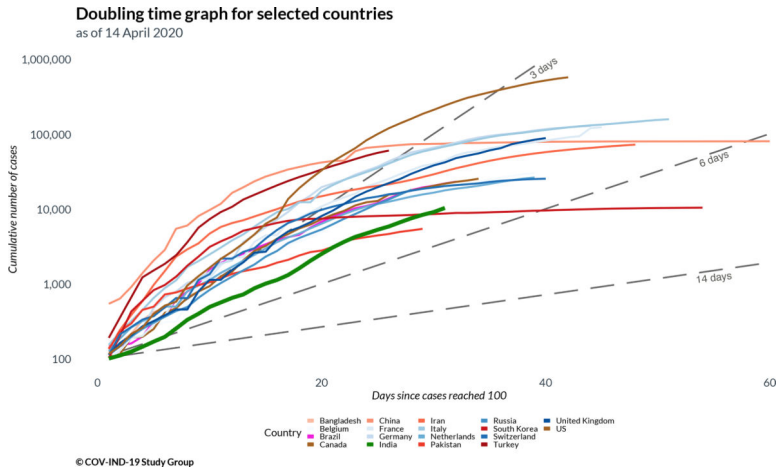


Figure 2.

Early phase of the epidemic and daily growth in cumulative COVID-19 case counts in India compared to other countries affected by the pandemic using data through April 14.

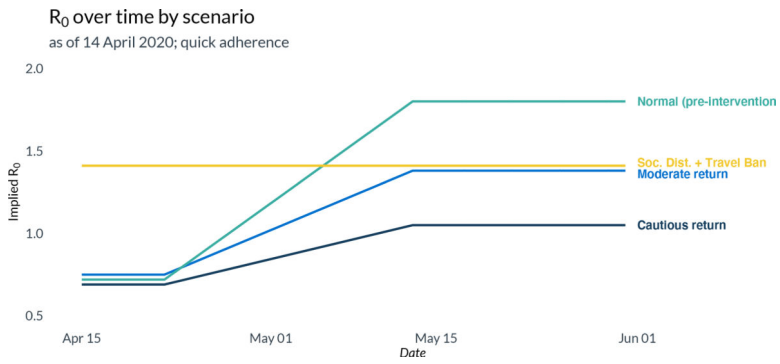


Figure 3.

Implied R_0 schedules corresponding the hypothetical scenarios under quick adherence. Corresponding plot for slow adherence is in Appendix Figure 2.

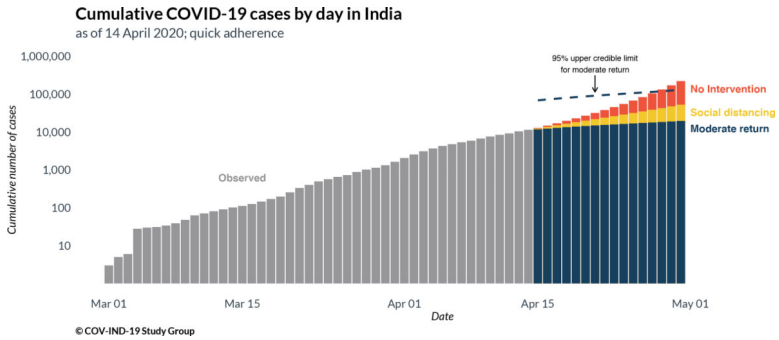


Figure 4.

Short-term daily growth in cumulative case counts in India assuming a 1-week delay in people's adherence to restrictions. Observed data are shown for days up to April 14. Predicted future case counts for April 15 until April 30 are based on observed data until April 14 using the eSIR model. The dashed horizontal line represents the upper 95% credible limit for estimates under "lockdown with moderate release" scenario. Corresponding graph following a 2-week delay schedule can be found in Appendix Figure 3.

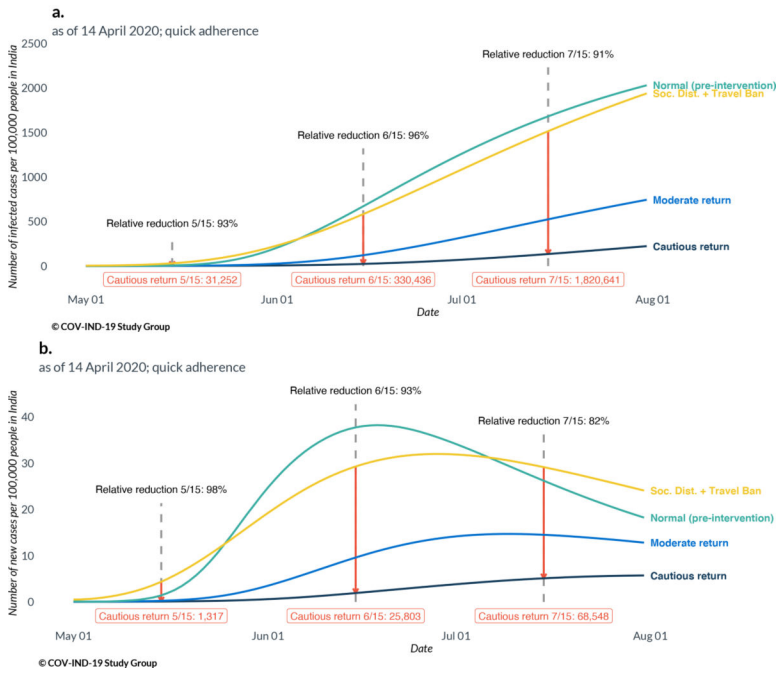


Figure 5.

Long-term daily growth in case counts in India per 100,000 people assuming a 1-week delay and how that is affected by different non-pharmaceutical intervention strategies. Predicted cumulative (a) and incident (b) case counts from May 1 to July 31 from the eSIR model are shown, based on observed data until April 14. Corresponding plots for slow adherence are in Appendix Figure 4.

a. Predicted number of COVID-19 cases in India under hypothetical scenarios

b. Predicted number of new COVID-19 cases in India under hypothetical scenarios

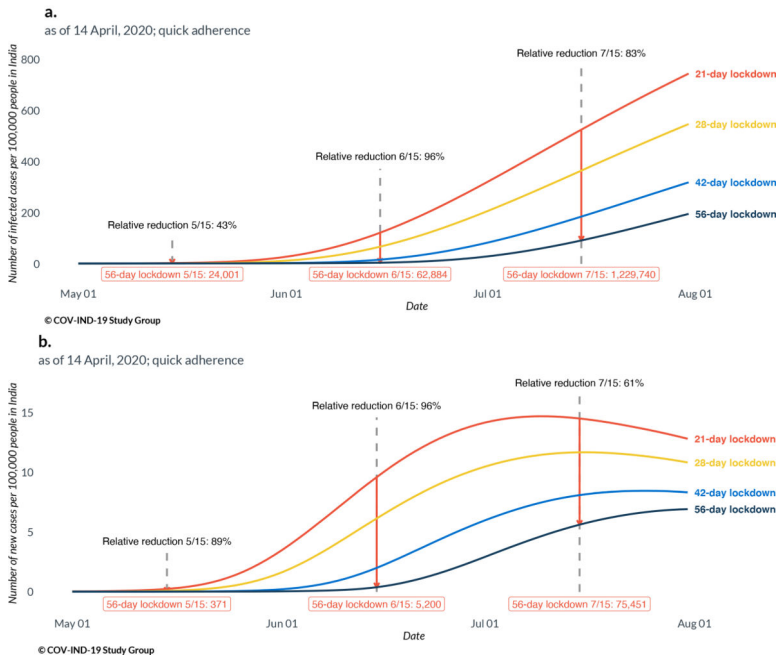


Figure 6.

Cumulative (a) and incidence (b) graphs for forecasting models assuming a 1-week delay under 21-, 28-, 42-, and 56-day lockdown scenarios using observed data through April 14. Corresponding plots for slow adherence are in Appendix Figure 4.

a. Predicted number of COVID-19 infections under varying lockdown lengths

b. Predicted number of daily COVID-19 infections under varying lockdown lengths

Table 1.

Timeline of COVID-19 interventions in India.

Date	Interventions
3 March 2020	• India issues travel ban on four countries - China, South Korea, Italy, and Iran
6 March 2020	• Union Health Ministry issues advisory to avoid mass gatherings
7 March 2020	• Mayor of Agra urges the Union government to close down historical monuments including Taj Mahal • Kuwait suspends flights to India
9 March 2020	• Qatar puts India on travel ban list
10 March 2020	• Manipur closes its border with Myanmar
11 March 2020	• India suspends all visas/e-visas granted to nationals of France, Germany, and Spain on or before today
12 March 2020	• **WHO declares the COVID-19 outbreak as 'pandemic'** • India suspends all visas excepting those for diplomatic, UN, or international bodies, official and employment purposes until April 15 • India reports 1st death
13 March 2020	• India reports 2nd death • Several academic institutions (e.g., JNU, IIT, IIM) cancel classes/convocations; some hostels close
16 March 2020	• Central government proposes social distancing measures until March 31 • India bans passengers from EU countries, UK, and Turkey until March 31 • Central government recommends closure of educational institutions until March 31
17 March 2020	• Taj Mahal is shut until March 31; ASI closes 3,000 monuments and 200 museums • Mandatory quarantine is imposed on passengers coming from UAE, Qatar, Oman, and Kuwait • **India is heading towards a countrywide lockdown mode**
19 March 2020	• India halts all incoming commercial international flights for one week • Some state governments ban public transportation • Prime Minister urges people of India to observe self-imposed curfew ('Janata Curfew') on March 22
20 March 2020	• Maharashtra announces lockdown in Mumbai, Nagpur and Pune • Jawaharlal Nehru University (JNU) in Delhi orders students to vacate hostels
21 March 2020	• Private labs can conduct COVID-19 tests, says Maharashtra government • Rajasthan government declares lockdown until March 31
22 March 2020	• 12 states including Telangana and Delhi announce lockdown until March 31 • International commercial passenger flights are disallowed to land in India for one week starting today • Railways suspends all train services until March 31
23 March 2020	• **Central government orders all states in India to impose lockdown** • Legal action is to be initiated against people violating lockdown measures
24 March 2020	• **Prime Minister of India announces lockdown for 21 days as country records 552 COVID-19 cases and 10 deaths**
28 March 2020	• Central government unveils stimulus package to help those hit by 21-day lockdown • Priorities are to construct COVID-19 hospitals, sample testing, contact-tracing and social distancing: Union Health ministry
2 April 2020	• Common exit strategy necessary for "staggered" relaxations after lockdown period ends, Prime Minister tells chief ministers
6 April 2020	• Prime Minister instructs union ministers to prepare a graded plan to gradually open departments that are not COVID-19 hotspots
8 April 2020	• Prime Minister and chief ministers decide on lockdown extension to April 11
9 April 2020	• Odisha extends lockdown until April 30 and becomes first Indian state to do so

Source: https://www.pharmaceutical-technology.com/news/india-covid-19-coronavirus-updates-status-by-state/

Table 2.

Comparison of estimated projections and posterior estimates of model parameters across different sensitivity analysis scenarios under 21-day lockdown with moderate return, using observed data until April 14. Prior SD for R_0 is 1.0.

Sensitivity Analysis Scenario	Predictions May 15		Posterior Estimates		
	May 1		R_0	β	γ
Under-reporting *	25,248 [104,411]	62,797 [343,465]	2.28 [1.05, 4.20]	0.20 [0.05, 0.39]	0.09 [0.03, 0.19]
Case-clustering **	24,818 [59,525]	57,499 [189,010]	2.81 [1.47, 4.70]	0.16 [0.07, 0.26]	0.06 [0.03, 0.10]
Prior mean for $R_0 = 2$	20,251 [135,034]	42,252 [315,348]	1.80 [0.87, 3.26]	0.27 [0.06, 0.59]	0.16 [0.04, 0.35]
Prior mean for $R_0 = 3$	25,757 [165,287]	86,750 [638,770]	2.43 [1.41, 4.07]	0.30 [0.09, 0.60]	0.13 [0.04, 0.30]
Prior mean for $R_0 = 4$	34,587 [213,556]	253,935 [1,854,319]	3.38 [2.09, 5.27]	0.32 [0.10, 0.63]	0.10 [0.03, 0.23]

* Observed case-counts are multiplied by 10, prior mean for $R_0 = 2$

** Assume that the cases happen in metro hotspots, use population size N=32 million instead of national population 1.34 billion, prior mean for $R_0 = 2$

Table 3.

Comparison of model projections using observed data up to different dates assuming a 21-day lockdown with moderate return.

	Projected Counts [Upper Credible Interval]		Posterior Estimates [95% CrI]		
Observed/Projection	**April 15**	**May 1**	R_0	β	γ
Observed	12,370	37,262	-	-	-
Used data up to April 1	1,944 [14,178]	3,807 [28,777]	1.85 [0.84, 3.47]	0.28 [0.05, 0.70]	0.16 [0.03, 0.40]
Used data up to April 7	5,344 [36,222]	8,330 [61,270]	1.74 [0.80, 3.22]	0.22 [0.05, 0.52]	0.14 [0.03, 0.32]
Used data up to April 14	11,736 [68,836]	20,251 [135,034]	1.80 [0.87, 3.26]	0.27 [0.04, 0.35]	0.16 [0.04, 0.35]

Note: All prediction scenarios assume a prior mean of $R_0 = 2$