SELECT ANONYMOUS REVIEWER REPORT and AUTHOR RESPONSE TO REVIEWERS

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


DOI 10.1162/99608f92.60e08ed5
We thank Dr. Meng, the four reviewers, the associate editor and the visualization editor for their insightful comments that led to significant improvement of the initial manuscript. In the following, we provide itemized response to each reviewer comment to the best of our abilities. We hope that this revised version of the manuscript is found suitable by the reviewers.

**Reviewer 1**

*This paper is well done. In terms of substance, I just have three comments:*

**R1.1) There is lots of modeling of India happening right now. Can they say more about how they compare to what others are saying and explain differences?**

**Response:** We thank the reviewer for appreciating our work. Yes, there has indeed been significant modeling effort with COVID-19 data from India. Not just for forecasting of cases but for optimal deployment of health care resources, network models that use regional contact and migration patterns and models for state-specific forecasting of cases and fatalities. We primarily focus on discussing other epidemiologic forecasting models for the number of infections as that is the domain our work belongs to. The forecasting models can be broadly classified into two categories: exponential growth models and compartmental epidemiologic models. We have now included a paragraph on existing India-specific models under the *Introduction* section. For the ease of readers, we also summarize these models, their pros and cons in a Table (see next page) which is included in the *Supplementary Materials*. One major limitation with these models is the lack of open-source code and clear statement of assumptions which make it challenging to reproduce the projections.

**R1.2) The paper provides confidence intervals, which is really useful as many other analyses don’t do this. But I have no idea where these confidence intervals come from and they aren’t emphasized nearly enough in the presentation. The confidence intervals are huge, and that should be acknowledged and discussed.**

**Response:** Thank you for this comment. It is indeed important to characterize the uncertainty in the projections. We implemented a Bayesian eSIR model that provides not only posterior estimates of underlying model parameters $\beta, \gamma$, and $R_0 = \frac{\beta}{\gamma}$ but also provides draws of predicted counts/proportions corresponding to the infected and the removed compartments at future time points. What we have reported are indeed credible intervals (CrI) and not confidence intervals. If $t_0$ denotes the last date with observed data, and $T$ denotes the last forecast date, the credible intervals for the projected cumulative case-counts from time $t_0 + 1$ to time $T$ were computed using the posterior distribution of the proportions corresponding to infected and removed compartments given the observed proportions of confirmed and removed compartments, i.e. $Y_{(t_0+1):T}^I | Y_{1:t_0}^I, Y_{1:t_0}^R$ and $Y_{(t_0+1):T}^R | Y_{1:t_0}^I, Y_{1:t_0}^R$. The cumulative case-counts at a time point are simple sums of these draws. Credible intervals we reported are simply the 2.5th and 97.5th sample quantile of the posterior draws after burn-in and thinning. As we go further in time, the uncertainties associated with the projections become very large. We have now provided the details of the algorithm and the MCMC implementation in the *Methods* section. We have also specifically mentioned about the wide credible intervals in the *Results* and the *Discussion* section. One advantage of the Bayesian method is that the exact posterior draws of the model parameters allow us to quantify uncertainty in the estimates of functions of parameters without relying on large sample approximations or delta theorem.
<table>
<thead>
<tr>
<th>Model type</th>
<th>Reference</th>
<th>Research question</th>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential model</td>
<td>Gupta and Shankar (2020)</td>
<td>Provide estimate of the infected population using death counts</td>
<td>Simple model helpful for scant data; modeled epidemic hotspots separately</td>
<td>Not accounted for population demographics (limited by data), non-pharmaceutical (NP) intervention effects; requires infection fatality rate</td>
</tr>
<tr>
<td>Poisson log-linear model</td>
<td>Das (2020)</td>
<td>Short-term prediction of future case counts; estimate $R_0$</td>
<td>Simple model helpful for short- &amp; medium-term forecasts using scant data; accounted for quadratic effect of time</td>
<td>Not accounted for population demographics (limited by data), hotspots, NP intervention effects; surveillance bias</td>
</tr>
<tr>
<td>Autoregressive-moving-average model</td>
<td>Deb and Majumdar (2020)</td>
<td>Analyze the trend pattern of incidence; estimate $R_0$</td>
<td>Accounted for quadratic effect of time, lockdown effect; captured time dependence incidence pattern</td>
<td>Not accounted for population demographics (limited by data), hotspots; surveillance bias</td>
</tr>
<tr>
<td>Susceptible-infected-recovered (SIR) model</td>
<td>Ranjan (2020)</td>
<td>Long-term prediction of future case counts; estimate $R_0$</td>
<td>Classical epidemiologic model used; accounted for social distancing effects</td>
<td>Not accounted for population demographics (limited by data), hotspots; surveillance bias; used first few weeks of data</td>
</tr>
<tr>
<td>SIR model</td>
<td>Dhanwant and Ramanathan (2020)</td>
<td>Long-term prediction of future case counts</td>
<td>Classical epidemiologic model used; split observed data into training and test data; training data used to learn the transmission rate; incorporated lockdown effect</td>
<td>Not accounted for population demographics (limited by data), hotspots; surveillance bias; lockdown training data not used to learn about transmission rate under lockdown</td>
</tr>
<tr>
<td>Age-structured SIR model</td>
<td>Singh and Adhikari (2020)</td>
<td>Study progress of the disease and impact of social distancing measures; estimate $R_0$</td>
<td>Extended epidemiologic model accounting for age distribution, social contact, social distancing effect</td>
<td>Not accounted for other population demographics (limited by data), surveillance bias; complex model given the scant count data and spotty individual-level data</td>
</tr>
<tr>
<td>Susceptible-Exposed-</td>
<td>Mandal et al. (2020)</td>
<td>Identify NP intervention</td>
<td>Extended epidemiologic model with an added compartment for</td>
<td>Not accounted for population demographics (limited by data);</td>
</tr>
<tr>
<td>Model</td>
<td>Authors</td>
<td>Description</td>
<td>Considerations</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>-------------</td>
<td>----------------</td>
<td></td>
</tr>
<tr>
<td>Influenza-Recovered (SEIR) model</td>
<td></td>
<td>Strategies that can help control the outbreak</td>
<td>Quarantine; accounted for other NP interventions, and connectivity between two places</td>
<td></td>
</tr>
<tr>
<td>Expanded SEIR model</td>
<td>Chatterjee et al. (2020)</td>
<td>Assess the impact on healthcare resources; study the effect of different NP interventions</td>
<td>Extended epidemiologic model with added sub-compartments for quarantined, recovered and death; accounted for different NP interventions; accounted for age groups</td>
<td></td>
</tr>
<tr>
<td>Expanded SEIR model</td>
<td>Senapati et al. (2020)</td>
<td>Assess the effect of different NP interventions; estimate $R_0$</td>
<td>Extended epidemiologic model with added sub-compartments for asymptomatic cases, quarantined, hospitalized, recovered and death</td>
<td></td>
</tr>
<tr>
<td>Expanded SEIR model</td>
<td>Sardar et al. (2020)</td>
<td>Assess long-term effect of 21-day lockdown; estimate $R_0$</td>
<td>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</td>
<td></td>
</tr>
</tbody>
</table>
**R1.3) These models can be very sensitive to parameter choices. It would be good to show more sensitivity to these choices.**

**Response:** The reviewer is absolutely right. This is the reason why we made all our code and assumptions public so that other researchers can reproduce our work and design their own interventions and prior specifications. In our sensitivity analyses, we have considered scenarios where we have focused on understanding the impact of:

(a) under-reporting of cases (assume that 10 times the number of reported cases exist in the population),

(b) case-clustering in urban areas (thus the population used is that of metro area hot spots instead of assuming the entire nation as one homogeneous unit)

(c) different prior specifications for basic reproduction number $R_0$, we varied the prior mean for $R_0$ with prior standard deviation set at 1.

The following Table R2 presents the sensitivity of the projected case counts and the estimated model parameters under the above three scenarios. We focused on one intervention setting and studied the effect of changing various model/data assumptions.

<table>
<thead>
<tr>
<th>Sensitivity Analysis Scenario</th>
<th>Predictions May 1</th>
<th>Predictions May 15</th>
<th>Posterior Estimates [Upper CrI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under-reporting*</td>
<td>25,248</td>
<td>62,797</td>
<td>$R_0$</td>
</tr>
<tr>
<td></td>
<td>[104,411]</td>
<td>[343,465]</td>
<td>$\beta$</td>
</tr>
<tr>
<td></td>
<td>[24,818]</td>
<td>[57,499]</td>
<td>$\gamma$</td>
</tr>
<tr>
<td></td>
<td>[59,525]</td>
<td>[189,010]</td>
<td></td>
</tr>
<tr>
<td>Case-clustering**</td>
<td>20,251</td>
<td>42,252</td>
<td>$R_0$</td>
</tr>
<tr>
<td></td>
<td>[135,034]</td>
<td>[315,348]</td>
<td>$\beta$</td>
</tr>
<tr>
<td></td>
<td>[25,757]</td>
<td>[68,705]</td>
<td>$\gamma$</td>
</tr>
<tr>
<td></td>
<td>[165,287]</td>
<td>[438,770]</td>
<td></td>
</tr>
<tr>
<td>Prior mean for $R_0 = 2$</td>
<td>34,587</td>
<td>253,935</td>
<td>$R_0$</td>
</tr>
<tr>
<td></td>
<td>[213,556]</td>
<td>[1,854,319]</td>
<td>$\beta$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\gamma$</td>
</tr>
</tbody>
</table>

* 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 $R_0 = 2$

Following are our reasons for considering these specific choices for sensitivity analyses:

(a) Due to reasons including but not limited to a lower number of tests being conducted across India (discussed in detail in our Medium article 3) and questions regarding the possibility of false negatives from the tests in use (NPR), it is important to ask what proportion of the true cases is being represented by the confirmed cases. Therefore, multiplying the number of reported cases by 10 and using that as an input to the eSIR model simply attempts to look at a scenario: what if only 1 in 10 cases are being captured in India. We note in Table R2 above that the predictions indeed go up and the effect is more palpable with long term projections.

(b) By observing the pattern of growth of COVID-19 across Indian states and union territories, it is quite evident that not all regions experienced increase in number of cases at the same time. Some early ‘hubs’ were observed, including the states of Kerala, Maharashtra and Karnataka, and the national capital region of Delhi. Therefore, it might not be prudent to assume 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 $Y_{1:20}$.
and $Y_{t_0}^{R}$ may lead to extremely small proportions, which may in turn yield underestimated outputs from the eSIR model. Hence, changing the total population size from that of India to that of representative (large) cities from the hub states is a simple but intuitive way to potentially do away with the aforementioned underestimation. One can note this significantly reduces the width of the credible interval.

(c) Multiple studies on the COVID-19 situation in India have indicated that a value of $R_0$ around 2 is a good ballpark estimate for the prior mean of $R_0$ to characterize the growth of the pandemic in India [Das (2020), Ranjan (2020), Deb and Majumdar (2020), Singh and Adhikari (2020), Sardar et al (2020)]. In most of our analyses, the posterior mean for $R_0$ was seen to be within 1.8-2.4 as well, irrespective of whether a higher/lower starting (prior) mean was used. We see in Table R2 that a prior mean of 4.0 sways the estimate significantly. As more data accumulate, we will expect the effect of the prior on the posterior estimates to diminish.

**Reviewer 2**

Congratulations on a wonderful manuscript and thank you for your diligent and timely analysis. In this paper you describe a set of analyses exploring the impact of non-pharmaceutical interventions on the spread of COVID-19 in India and the importance of maintaining social distancing measures to curb the spread of this disease. I particularly appreciate the discussion on the socio-economic impact of these interventions—we public health officials do not take these interventions lightly. There are a few sections that I wanted to highlight.

R2.1) In the final paragraph of the introduction you write “We convened virtually after being quarantined in our homes with alternating waves of fear and inspiration surrounding us. We decided to channel our collective energy to study the defining public health and economic crisis of our time and use our data science expertise to search for answers and solutions that can help COVID-19 related policymaking in India.”

I must admit that when I read this, my eyes welled up with tears. All too often science is seen as an aseptic environment devoid of emotional connection despite the fact that we come to science because of our commitment to humanity. I commend you for including these sentiments in this manuscript and wish to convey my gratitude at your commitment to science in these uncertain times. We are all in this together and I have felt these same things in my own work.

**Response:** We are very thankful to the reviewer for their thoughtful comments and for appreciating the effort and sentiment that we have put in as a team. However, based on other reviewer and editorial board suggestions, we have decided to refrain from making altruistic personal statements in the main text and kept this article strictly scientific. We do mention the exceptional context when this work was done in the acknowledgement section.

R2.2) You describe the limitations of testing in the second paragraph of the introduction. This has been the most frustrating aspect of this pandemic, but I thank you for providing information on the total number of tests conducted. If you happen to have the data at the state level, it would be nice to provide an average testing per capita by state and the confidence interval around that estimate just to drive home the issues in the distribution of testing.

Relatedly, can you provide some information on India’s priorities in testing? Is testing limited to those who are hospitalized? Or high risk? Or with recent travel history? Or confined only to those who have been in contact with a known positive? These details can provide a better understanding of the bias inherent in using the number of confirmed cases in your analysis.

**Response:** Although there exists some crowd-sourced state-level testing data (for some states only), we have found these data to be very spotty. Also, in this article, we are not really doing any state-level modeling. So, we
have decided not to include any state-level information on the average testing per capita. Our app covind19.org has tabs for state level forecasting and national testing data. The testing data from India has been limited.

India’s priorities in testing have changed a few 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 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 in his/her contact. More recently, the testing strategy was revised on April 9 to additionally include testing of all symptomatic people in hotspots/cluster and in large migration gatherings/evacuee centers.

In response to your question we delved a little deeper into the issue of testing bias and gathered data from various sources. Figure R1 shows that even though the number of tests in India has increased in the last 10 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 US and UK. We also plot Iceland and South Korea on this Figure as they have done a remarkable job in administering a large number of tests per detected case and serve as good examples for the world. We also looked at the proportion of the population tested in 61 countries around the world. 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. We have now added a paragraph on testing patterns in India in the Discussion section and have added these figures in the Supplementary Materials of the paper.

Figure R1. Daily testing patterns in selected countries.

Figure R2. Testing numbers and proportions for 61 countries around the world affected by COVID-19

R2.3) Can you provide some justification for using a similar $R_0$ value as reported from the Chinese context of the pandemic? The varying contact rate by population density is a major contributor to the spread of this virus. While we certainly don’t have the data for social networks, is there any reason to believe that the contact rate would be similar between the two countries?
Response: This is an excellent point and we apologize as we were not clear in our prior model specifications. India has a very unique contact network, particularly for the elderly, living with children and grandchildren. It is not easy to isolate the elderly in separate homes. Additionally, there are large slums and community dwellings with shared sanitation facilities. One of the major limitations in our model is the inability to take account of this region-specific contact network and the spatial complexity. It will indeed make the model more tuned to the Indian context. We discuss this issue in the Limitations section of the revised manuscript.

We would like to clarify that we are not using a similar $R_0$ value for India as reported from the Chinese context [Wang et al. (2020)]. The prior $R_0$ mean value for India, in absence of any interventions, is assumed to be 2.0, which is a sensible estimate confirmed by $R_0$ estimates from other studies on the early-phase data from India [Das (2020), Ranjan (2020), Deb and Majumdar (2020), Singh and Adhikari (2020), Sardar et al. (2020)]. It is important to note here that for China, this value was estimated to be around 3.5-4.0 [Wang et al. (2020)]. What we are assuming is that the effect of non-medical interventions (travel ban, social distancing, lockdown) in India and China would be similar on a proportional scale. That is, if an intervention decreases the reproduction rate in China by 25% (i.e. $\pi(t) = 0.75$ in our eSIR model), then a similar intervention would decrease the $R_0$ in India by the same proportion (i.e. we use the same $\pi(t)$ in our model). In the absolute scale of magnitude, with initial values of 4.0 and 2.0 respectively for China and India, this would result in an effective $R$ of 3.0 and 1.5 after the intervention. Given the similar population densities in Wuhan and India, the assumption on similar effect of interventions across the two countries does not seem too restrictive. Figure R3 shows how we apply $\pi(t)$ to the forecasting period under various types of release/return pattern once the lockdown is lifted. The comparative line that stays flat is the perpetual social distancing + travel ban without any lockdown.

![Figure R3](image.png)

Figure R3: A visual description of the effective $R_0$ under various patterns of return after the lockdown.
R2.4) If you have access to hospitalization data (number of probable or confirmed cases in hospital) this might serve as a better proxy than number of cases. I know these data have been hard to come by, but perhaps you could at least mention the utility of using hospitalization data, since this would make the modeling assumptions more robust if we had access to these numbers.

Response: As the reviewer indicated, getting hospital data from India is challenging. There exists limited amount of crowdsourced data but only for large metros. We have now mentioned the potential gain from having easily accessible hospitalization data and the need for having accurate cause of death listed in death record in the Limitations section. As multiple papers have suggested [Ferguson et al. (2020)] one can estimate infection rates from reliable mortality data. In absence of large number of tests, these data can help.

R2.5) Throughout the manuscript you use the term “significant”. I would recommend changing this to “substantial” given the weight the term “significant” often holds in statistics. (Second-to-last sentence in first paragraph of results “shift the projection curve significantly”; third-to-last sentence in sensitivity analyses “again showing the significant variability in these numbers”)

Response: Thank you for mentioning this important point. We have made this change throughout.

R2.6) You use the abbreviation CI to mean both “confidence interval” and “credible interval”. I’d strongly suggest using “CrI” for the abbreviation for credible interval instead of “CI”.

Response: We have now used CrI following your recommendation.

R2.7) I admired the attempt to look at the temperature variability. This is an important analysis, but I wonder how much of the transition to spring/fall while the epidemic is ramping up is impacting the results. It will be interesting to see whether these correlations continue to strengthen throughout April and May as both the seasonal transition wanes and the pandemic establishes itself. This analysis also relies heavily on the assumption that the rate of testing for COVID-19 is stable, when we know the rate of testing has been increasing. I don’t know that there’s a way to adjust for this (in fact I’m not certain that data are available ubiquitously on number of test being done), but this is an important data source that will help us continue to do this work. If it is available, perhaps doing a similar analysis looking at the correlation between number of tests done per capita (or, alternatively, the percent positivity) and the daily average temperature might clear up some of these questions.

Response: Thank you for this suggestion. We realize this analysis is not robust and subject to many confounding biases and data limitations. Based on suggestions from other reviewers and the editorial board, we have decided to remove the temperature analysis from this article.

R2.8) I appreciate your commitment to open data and am very happy that you have provided all of your work in an open source format. Thank you again for your thoughtful and timely analysis.

Response: Thank you for your kind remarks. The whole goal of this modeling exercise was to make the products accessible to everyone and make the models and assumptions explicit and transparent. Our app covind19.org has many updates now and is visited by thousands of people.

Reviewer 3

I congratulate the authors on a comprehensive analysis, compelling visualizations, and releasing their code and producing a Shiny app. Kudos to the research team. I was asked to comment on the manuscript with a very limited timeframe for review, so will focus specifically on the epidemiological implications, as this is my area of expertise. My main concern revolves around the presentation of the results and speculations in the discussion. I think we
have to be very cautious with who may read and misinterpret these findings, as well as form policy recommendations without evaluating them in the model.

**Response:** We thank the reviewer for providing a detailed review of our work and for appreciating our effort. We have provided point-by-point response to all their comments and accordingly modified our manuscript, which we hope will address their concerns.

**Methods**

**R3.1) I would caution authors about drawing firm conclusions from public data without consideration of surveillance bias: underreporting and misdiagnosis.** The authors state this upfront in their introduction, and mention in passing in the methods “underlying unobserved prevalence” in the eSIR models, but I do not know what this parameter was set to or how this was estimated. This would require some kind of serosurvey with a very accurate test – data that do not exist to my knowledge. The sensitivity analysis briefly mentions a 10-fold increase in unreported cases, but based on what? What were the results?

**Response:** We thank the reviewer for raising these concerns and apologize for not being explicit about our modeling and assumptions. We have provided a clear description of our model, algorithm and priors in the revised manuscript now (please see **Methods** section). We also provide a more in-depth look at the testing data from India (please see **Figures R1-R2** and response **R2.2**).

The issue that the reviewer mention regarding lack of testing, underreporting and misdiagnosis leading to surveillance bias is critical and acute for India. We have carried out sensitivity analysis for our model assumptions (Please see **Table R2** in response to Reviewer 1, R1.3). We note that the projected case-counts do go up if we assume only 1 out of 10 cases were reported/captured and the gaps in forecasting increase as we move further in time. However, we note that the estimate of basic reproduction number $R_0$ is more robust to under-reporting since both cases and recovered are under-reported in this situation ($R_0$ changes from 1.80 to 2.28 in rows 3 versus 1 in **Table R2**). Another point we would like to mention is that this type of bias affects each hypothetical intervention scenarios in a similar way, thus the relative comparison will be largely valid in qualitative terms.

We have mentioned the following in the **Results** section:

“We believe that our point estimates are at best underestimates due to potential surveillance bias (under-reporting and/or mis-diagnosis of case-counts) and our model not taking into account the population density, age-sex and contact network structure of the whole nation. However, we note that the estimate of basic reproduction number $R_0$ is more robust to such underreporting issues as counts in all compartments are assumed to be underreported to some extent. 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.”

**R3.2) Re: temperature analysis, aggregating data to a per-month level runs the risk of losing lots of valuable detail contained in the raw daily data.** Why not use the daily or weekly trends? Surely historic temperature data are available at this level by contacting the weather bureaus of various jurisdictions. Relatedly, I am confused as to why the authors now switch to a global analysis. This seems to be a totally separate paper. Instead, it would make more sense to focus on India specifically.

**Response:** The reviewer is right; the temperature analysis does not belong to this article in a natural way. Based on suggestions from the reviewers and the editorial board, we have decided to remove the temperature analysis from this article as there were many limitations in that very coarse analysis.
R3.3) From my understanding, the SIR model assumes a homogeneity within each “group” of people that are moved between states. This may be problematic as there can be great variability even within somewhat similar demographic groups, such as adherence to policy, disease natural history and progression, etc. Please comment.

Response: The eSIR model that we are using is purely count-based, and therefore it is indeed true that it 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 more or less likelihood of coming into contact with one another. To account for all such potential factors, we need more nuanced modeling and granular level data. One potential way is to stay in the count-based framework and break up the infected component into further sub-compartments (a simple example being replacing the $\beta \theta^\text{i1} \theta^\text{i2}$ term in the differential equations by something like $\sum_{i,j} \beta_{ij} \theta^\text{i1}_{ij} \theta^\text{i2}_{ij}$ where the sum ranges over a set of clusters). Another way that has been pursued in a different study could be to inform the SIR modeling via introducing contact networks at the initial stage (https://medium.com/@veerab_12080/regional-contact-networks-and-the-pandemic-spread-of-covid-19-in-india-28b3b3aa2161). However, it is to be noted here 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 (https://www.kaggle.com/sudalairajkumar/covid19-in-india), the quality and detail of the information provided are quite heterogenous and thus, how to best utilize such data remains a question. Therefore, as much as we agree with the reviewer’s comment, we refrained from using such a framework in our analysis. We also believe that in the beginning phase of the pandemic with limited days’ worth of data, simpler models are better. Reliable estimation of parameters from complex models with sparse data in sub-compartments is hard in absence of additional external data to identify the parameters. As more individual level data are available we can introduce more compartments.

We have now included a discussion of this issue in the Limitations section.

R3.4) I will not repeat in detail here as the authors acknowledge as a limitation, but important criteria are missing from the model such as demographic (e.g. age) and socioeconomic considerations for population mixing. This has become de facto in the epidemiologic infectious disease modeling world.

Response: We agree with the reviewer’s comment. As mentioned in the previous comment, the lack of reliable, granular level data from India as yet precludes us from considering demographic and socio-economic factors in this article. Future opportunities for improving our model include incorporating contagion network, age-structure, age-specific contact structure, using the SEIR model, incorporating test imperfection, and estimating true fatality/death rates. As pointed out, we have mentioned these issues in the Limitations section.

R3.5) Since there is a geographical aspect to this work, why not allow starting parameters to vary as a function of geography? The various disease transmission rates will vary in differences population density.

Response: To incorporate variations due to geography, one can consider a state-level modeling using population count data from 2010 data. Note that we do not have any granular region-/district-level data yet. State-level forecasts for some states are provided in our RShiny app (covind19.org); however, state-level data starts at a much later date (March 15). Our predictions in this article are based on data until April 14, which gives us only 4 weeks’ worth of observed data, and a long time-window to extrapolate for our forecasts. Including state-level projections is beyond the scope of this paper as the data quality varies drastically across states and is not feasible to carry out for many of the states and union territories.
Results

R3.6) I have a concern about the way the results are presented. First, presenting the results as “no intervention” is an unrealistic scenario, yet could easily be misconstrued by the public/media as hyperbole of the pandemic’s current impact, which shouldn’t be underestimated. Relatedly, why focus on absolute quantification of impact? Even with huge CI ranges, these numbers can be incorrect. Instead, would it not make more sense to include relative comparisons between the group? If the authors intend to keep the absolute numbers, I would also include the relative contributions.

Response: We agree that relative comparisons between groups will make more sense because absolute numbers are widely varying and most likely incorrect. We have now provided relative numbers comparing “social distancing” with “no intervention” to provide a glimpse of what benefit simple social distancing can provide. Next, we provided relative numbers comparing “lockdown” scenarios with “social distancing” to emphasize the added benefit of lockdown measures. Please note that we have not included the “no intervention” scenario in the long-term forecasts.

R3.7) The authors even acknowledge that “point estimates are at best underestimates”. This begs the question why present these? One can just present the plausible range, but again due to the uncertainty and wide CIs I think a focus on relative impact is more appropriate. I appreciate the statement “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” in the Discussion. This should be bolded and underlined. This summarizes many of my concerns quite well. Perhaps promote this disclaimer to the Introduction instead of burying in the Discussion.

Response: As mentioned in the previous comment, we have changed our narrative to focus on relative comparison. This acknowledgment about underestimation is made to reflect that our primary analysis does not attempt to account for underreporting of cases. While this is an extremely important consideration, no reliable estimate for the extent of underreporting in India exists at this time and an adequate estimation of such a metric is beyond the scope of this manuscript. Please also note our responses R1.3 and R2.2.

We have bold-faced and underlined the statement you mention about focusing on qualitative takeaway messages and not focus on exact numerical counts.

R3.8) I take issue to this claim “purely 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”. First, one could arbitrarily extend or shrink the lockdown to demonstrate what could be inferred from a back of the envelope calculation. Second, the focus of this model is quantifying the impact of COVID-19 disease: but there are other health and non-health outcomes that the epidemiologist considers. The following sentence “We note that longer lockdown periods are accompanied by increasing costs to individuals - notably economic - and must be considered” unfortunately gives short shrift to this. If this must be considered, why was it not considered? Economic is one impact (I am not an economist so cannot comment), but what about mental health issues or chronic disease exacerbation issues? What about disparities in the burden of COVID-19 disease? Please see my later comment in the Discussion section.

Response: We thank the reviewer for this comment. It is indeed true that this statement about extended lockdown periods was misplaced. Perpetual lockdown will definitely reduce case-counts further but is not a viable societal strategy. Even from a public health or epidemiologic perspective these stringent social distancing measures only focus on reducing coronavirus infections, perhaps at the cost of increasing other morbidities and mortality during this time. In US we now have suggestive data that morbidity and mortality due to traffic accidents and other
infectious diseases have gone down during this time but the toll from mental health disorders, substance abuse, chronic diseases have increased [CDC; Leon et al. (2020)]. We now mention this in the Discussion section and argue that a more holistic analysis using a cost-benefit framework, quantifying the total number of excess deaths during this period is important to conduct.

R3.9) Can the results of the sensitivity analyses be shared, especially since the authors state that “the exact quantitative projections are quite sensitive to such choices”

Response: Please see response to reviewer R1.3 and Table R2. We have now discussed this issue extensively in the manuscript and placed the table for sensitivity analysis in the Results section.

R3.10) I agree with the authors that “It is extremely important to update these models as new data arise.” Hear hear. This is not a one-and-done endeavor as the authors know. It is also important that the models remain transparent based on historic data so we can appreciate the accuracy of projections as better data become available. This will inform our bias correction approaches to surveillance data in the future.

Response: We could not agree more with the reviewer. Our Shiny app updates these models on a daily basis. Our GitHub repository stores all past projections. To check the calibrating properties of the model we truncated the data to certain dates and tried assessing the quality of the predictions with essentially one more week of data for a future date (Table R3 and Figure R4). We do notice the projected case-counts change significantly with more data but always underestimate the observed counts. This phenomenon is also due to more testing being done each week. We are in the process of using a correction factor each week based on what we have observed. However, the observed number is always within the prediction credible interval. This again reveals the large uncertainty. We have included this Model Calibration piece in the new Results section and the following Table.

Table R3. Comparison of model projections using observed data up to different dates assuming a 21-day lockdown with moderate return.

<table>
<thead>
<tr>
<th>Observed/Projection</th>
<th>Projected Counts [Upper Credible Interval]</th>
<th>Posterior Estimates [95% CrI]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>April 15</td>
<td>May 1</td>
</tr>
<tr>
<td>Observed</td>
<td>12,370</td>
<td>37,262</td>
</tr>
<tr>
<td>Use Data up to April 1</td>
<td>1,944</td>
<td>3,807</td>
</tr>
<tr>
<td></td>
<td>[14,178]</td>
<td>[28,777]</td>
</tr>
<tr>
<td>Use Data up to April 7</td>
<td>5,344</td>
<td>8,330</td>
</tr>
<tr>
<td></td>
<td>[36,222]</td>
<td>[61,270]</td>
</tr>
<tr>
<td>Use Data up to April 14</td>
<td>11,736</td>
<td>20,251</td>
</tr>
<tr>
<td></td>
<td>[68,836]</td>
<td>[135,034]</td>
</tr>
</tbody>
</table>

Note: All prediction scenarios assume a prior mean of $R_0 = 2$
R3.11) In this section “COVID-19 case counts and daily temperature” I believe the authors mean to reference Figure 7 (not Figure 6).

Response: We have removed the temperature analysis from the revised manuscript.

R3.12) Maps depicting cases by state/territory in the Shiny app should be adjusted by population density, otherwise we are just visualizing where the population is.

Response: We think the reviewer is referring to the “Cumulative case counts by state/union territory” plot in our Shiny app. The target of this plot is to show which parts are hot-spots and how new hot-spots are coming up over time. The government of India declared high risk red zones based on raw case-counts [Economic Times]. We now have state-level forecasts in the app which accounts for the state level population density. Please check the “State-level predictions” tab in the app.

Discussion

R3.13) The discussion seems to bring up points that weren’t the focus of the study, such as the healthcare system capacity, BCG vaccination, and impact on vulnerable groups, to name a few. I would refrain from these speculations unless it was a study aim and modeled. In fact, I think this is an opportunity with this model. There are lots of COVID-19 models out there, but few that consider the differential impact on vulnerable and stigmatized groups. Further, missing from the discussion was a comparison of this model to other models out there, such as the Imperial College model. The authors acknowledge the other models but do not really compare and contrast their approach with the others. What are their strengths and weaknesses? Are results complementary or not?

Response: We agree that the hypotheses about BCG vaccination, temperature and alike do not directly bear relevance to this study and so we have removed those points and streamlined the presentation to focus on the epidemiologic forecasting and intervention comparison. However, we think our comment on the healthcare capacity and vulnerable populations is important to drive home the point why we care about predicting and controlling the surge of COVID-19 infections through such stringent interventions.

Regarding the reviewer’s other comment about comparing with other models, numerical calculation and comparison is beyond the scope of this paper as none of the models have usable codes available in public domain. We have now provided an overview of India-specific models in the Introduction section. Please see Table R1 and response R1.1.
R3.14) The recommendations section strikes me also as inappropriate. I am not saying that our work as epidemiologists should not connect with policy, but unless the interventions were specifically modeled this is pure speculation. Policy should not be informed on the basis of a single model: it should come from observational data and other competing models. (On a personal note, I happen to agree with many of the authors recommendations, but this manuscript did not evaluate the overwhelming majority of suggestions thus my suggestion to omit, or place I a separate article as a commentary).

Response: The reviewer is absolutely correct. We have now removed the recommendation section as they cannot be derived or deduced from the models and were more generic in nature. We mention that takeaway messages from the models inform policymaking. We incorporate some of our suggestions regarding the need for more accurate data in the Discussion section.

R3.15) The following section “Role of data science and the troubling existence of many models” also seems like a standalone commentary rather than a component to the present work.

Response: We have now removed this entire section “Role of data science and the troubling existence of many models”. Instead we provide a summary of the many existing models for India in the Introduction section and fold in some of the comments regarding model uncertainty and data science in the Discussion section. This strategy is governed by suggestions from other reviewers who deemed some of the statements made in that section on data science and modeling variation to be crucial.

Reviewer 4

R4.1) Abstract: Objective: need to better motivate analysis. why is ‘average monthly temperature” important to assess? Temperature is not a possible intervention.

Design: need to clarify what Hopkins dashboard data is providing, need to define kaggle - ‘COVID-19 incidence counts data from Kaggle’

What assumptions are used to populate the eSIR model?

Outcome: define ‘cases’; death would be an important stand-alone outcome to predict.

Results: stated outcome results “e.g. predicted cumulative number of COVID-19 cases in India’ do not match those defined and stated in the outcome section.

Interventions modelled need to be defined in the methods or outcome section of abstract

Interventions stated do not match those in the objective section.

Conclusions: this statement is unsupported “Our analysis shows we need to have some measures of suppression in place after the lockdown for the best outcome.’ Needs more context.

This statement is also unsupported: ‘However, the lockdown comes at a tremendous price to social and economic health through a contagion process not dissimilar to that of the coronavirus itself.’ Needs more context.

Response: We thank the reviewer for their thorough reading of our manuscript. We have restructured the abstract completely to conform to Harvard Data Science Review (HDSR) format. In summary, we have removed the temperature analysis as this was truly out of place. We provide more details on data sources and the definition of “cases” as reported confirmed COVID-19 infections. We have made sure the results section is consistent with the abstract. We have provided more details on the specification of intervention and the model we used. We have removed any conclusions/recommendations that are not directly studied through our predictive and model
building exercise. We have quantified what we mean by “Our analysis shows there is need for some measure of suppression after lockdown” by comparing the relative reduction in number of infections with a cautious return versus an unconstrained return with no interventions (back to normal).

R4.2) Software: “Anyone can visualize the observed data for India and create predictions under hypothetical scenarios with quantification of uncertainties.”

What are the assumptions of the model? What are the ‘uncertainties’ mentioned. Needs more context.

Response: We have modified this statement to say that we report predictions and associated 95% Bayesian credible intervals. We have provided more details on the models, assumptions and intervention scenarios. In the Methods section, we have now clearly described how credible intervals are calculated for each quantity we report.

R4.3) Body of article

Introduction, first paragraph: “instituting the right public health interventions at the right time including”.. unsupported statement.

Last paragraph first page of introduction, define ‘cases’.

Define curve – related to cases or to deaths?

D and e outcomes are not the objects of study, but rather potential implications of your analysis. Why isn’t testing adequacy on your list? Why isn’t case fatality on your list of outcomes? What is the objective of India’s policy interventions? Avoidance of death or of ‘cases’ however defined.

If this is an India-centric analysis hoping to improve public policy, why is this journal the right fit for the analysis? What is the value of your model and app over other models and apps already available?

The temperature analysis needs to be better motivated.

Response: We have streamlined the revised manuscript, removing recommendations and statements that are generic, purely driven by common sense but not scientifically tied to the epidemiological modeling piece. We have clarified that we only consider number of reported confirmed cases of COVID-19 infections as our outcome. We have delved deeper into the issue of testing (see response to reviewer R2.2). The number of deaths in India up to April 14 is fairly small, thus we do not model or predict the case-fatality rate. With more data being accumulated, we hope to be able to model death as a separate compartment in the SIR model. Right now, the “R” or “Removed” compartment of the SIR model contains both death and recovered with no distinction. We also extensively review existing models for India (see Table R1 and response R1.1) and include it in the manuscript. It is to be noted that numerical calculation and comparison with other models is beyond the scope of this paper as none of the models have usable codes available in the public domain. And, as mentioned in a previous comment, we have removed the temperature analysis.

Your comment regarding this analysis being focused on India and primarily for informing policy then why HDSR is a suitable outlet is a complex one. Our initial report and articles that are not peer-reviewed already provided critical information to the government, generated approximately 200 media clips reaching 1 billion people. The lockdown was introduced and extended. Thus, in a sense, the purpose of this work is served and the mission fulfilled. However, we thought it was important to highlight the importance of rigorous data science work even with imperfect and limited data to inform policy. We felt our app and data science product that is visited and used daily by thousands of people represent open-source, reproducible and transparent data science. We are deeply appreciative that the review panel concurred and thought this is an important and timely analysis and of potential interest to HDSR readers.
R4.4) Methods

How does Hopkins define ‘cases’ how does that match your measure of cases or of true cases. Why is ‘cases’ the relevant measure for policy makers? Others that might be interested?

Since temperature is measured daily, why do this analysis using an aggregated monthly measure – this data transformation may entail assumptions that need to be articulated in your analysis.

Response: “Cases” are defined as reported confirmed COVID-19 infections. Due to reasons including but not limited to a lower number of tests being conducted across India (Please note response to reviewer R2.2 and Figure R1-R2) and questions regarding the possibility of false negatives from the tests in use (NPR), it is important to ask what proportion of the true cases is being represented by the confirmed or reported cases. There are other outcomes that are of interest to policymakers, in particular how many cases will need hospitalization, ICU admission or ventilators. In absence of ground level data in India, we can grossly estimate these numbers based on the fact that approximately 20% of symptomatic infections in other countries required hospitalization and 5% needed admission to the intensive care unit (ICU). The number of deaths in India up to April 14 is fairly small, thus predictions of death-counts will have larger uncertainties and may not be as useful for policymakers. We have added a discussion of outcomes relevant for policymakers in the Introduction section in the revised version. As mentioned before, we have removed the temperature analysis.

R4.5) Statistical model

I find this assumption “we used 1.34 billion as the population of India, thus treating the country as a homogeneous system for the outbreak” particularly unsatisfying as India is a very large and heterogeneous country in many respects. This suggests the country level analysis is also making many assumptions about the data generating process and therefore predicted outcomes. Moreover, ‘removed’ is a complicated aggregated measure and one that does not lend itself easily to interpretation particularly by policy makers.

Response: The reviewer is absolutely right – assuming India to be homogeneous is not a realistic assumption for our model. To incorporate variations due to geography, one can consider a state-level modeling using population count data from 2010 census data in each state. We now mention that state-level forecasts for some states are provided in our RShiny app (covind19.org); however, state-level data starts at a much later date (March 15). The problem with state level modeling is the very sparse counts in some states as well as gaps in daily reporting over a period (sometimes the case-counts are updated after 2-4 days). Our predictions in this article are based on data until April 14, which gives us only 4 weeks’ worth of observed data, and a long time-window to extrapolate for our forecasts. Including a detailed discussion of the state-level projections is beyond the scope of this paper as it is not feasible to carry out the modeling for many of the states and union territories due to data limitations. In addition to differences in population and demographics, India has a very unique contact network, particularly for the elderly, living with children and grandchildren. It is not easy to isolate the elderly. There are large slums and community dwellings with shared sanitation facilities. One of the major limitations in our model is the inability to take account of this region-specific contact network. It will indeed make the model more tuned to the Indian context. We discuss this issue in the Limitations section. Please see response to Reviewer 3, R3.3. And, as mentioned before, we only consider number of reported confirmed cases of COVID-19 infections as our outcome. Right now, we are not making any prediction about the “Removed” compartment of our eSIR model. When we have more death data, dividing this compartment into recovered and deceased will be the next step.

R4.6) Parameter choices for short-term forecasts “basic reproductive number R0 (the expected number of cases generated by one infected person assuming that the whole population is susceptible) are 2.0 and 1.5 respectively [the change in R0 was created based on what we saw in Wuhan16]. The value of 2.0 was estimated based on the
early phase data in India. For the current scenario of lockdown, our chosen mean for R0 prior starts with 2.0 during the period of no intervention, drops to 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).

These are critical assumptions of the model – what if R0 is different in different areas (urban vs. rural) or provinces? What if it is endogenous to potential interventions. How does your model account for different assumptions regarding R0?

How sensitive is your model results to these assumptions? Does the software allow for the user to change R0? And other assumptions?

Response: The reviewer is right; R0 can be different in different areas depending on region-specific contact network. However, due to unavailability of reliable granular region-/district-level data from India yet, it is difficult to account for these in our model. We have discussed these in the Limitations section. We have performed sensitivity analyses of some of our assumptions, including evaluation of robustness of results to R0 prior mean specification. Please see Table R2 and response R1.3 for details. The eSIR R package is flexible and allows the user to change R0 and other parameters to reflect user assumptions (https://github.com/lilywang1988/eSIR).

R4.7) Parameter choices for duration of lockdown analysis – why are these durations relevant to policymakers?

Response: India started with a 21-day lockdown from March 25. We wanted to see the relative benefits (in controlling growth of COVID-19 case-counts) of lockdowns of an extended duration compared to a 21-day lockdown in a simulation experiment. These comparisons can be useful for Indian policymakers in making decisions about whether to extend lockdown depending on the situation at the end of 21-day period. We considered extensions by 1, 3 and 5 weeks. As we were finalizing this manuscript, India announced a 2-week extension on April 14 that led to a 40-day lockdown. It was again extended until May 18, making it a 56-day national lockdown. The relevance of the time-window was based on what the government was considering as possibilities. Many other countries in the earlier cycle of the pandemic have opted for a 2-3 month lockdown.

R4.8) Why was the temperature analysis chosen over others enumerated in the manuscript?

Response: We have removed the temperature analysis in this revised version as it did not belong with the rest of the article.

R4.9) Open source software – how has your model been validated? What advantages does it have over other models? What disadvantages does it exhibit?

Response: In this manuscript and in our Shiny app, we adapt the eSIR model (Wang et al., 2020) and the corresponding R package (https://github.com/lilywang1988/eSIR) to model case-count data from India. This software has been validated by the authors using data from Hubei and Italy. We have summarized some India-specific models, their pros and cons in a Table (see Table R1 and response to reviewer R1.1) which is included in the Supplementary Materials. One major limitation with these other models is the lack of open-source code and clear statement of assumptions which make it challenging to reproduce their projections. We have also commented on the calibration properties of our prediction models as more data accumulate (see response to reviewer R3.10 and Table R3 and Figure R4).

R4.10) When the manuscript mentions ‘uncertainties’ related to the software, do you mean risks or actual uncertainties. What predictor variables are ‘uncertain’ in your model?

Response: We mean uncertainty in the predictions of case-counts as indicated by wide credible intervals corresponding to our estimates. Large uncertainty in our predictions arise from many unknowns including but not
limited to not-so-precise prior distributions for key parameters related to disease transmission in our model, potential biases from a lower number of tests being conducted across India, under-reporting and misdiagnosis of fatalities. We have now provided the details of how these credible intervals are obtained in the Methods section. We have also specifically mentioned about the wide credible intervals and large uncertainties in the Discussion section.

R4.11) Lack of clarity in the language used to describe the app relative to the analysis detracts from the potential relevance of the work described.

Response: Thank you for this point. We have tried our best to align the language across the manuscript.

R4.12) Results

Suppression analysis needs to be better motivated explained in the introduction and methods sections and abstract. ‘optimal duration’ of lockdown analysis also requires more work – what is the objective that defines ‘optimality’ is it ‘cases’ or ‘deaths’. Not obvious which policymakers are optimizing.

Response: As we mentioned in response to your earlier comment R4.1, we are only modeling and predicting future case-counts related to COVID-19 infections. The coronavirus related death-counts in India are fairly small to make useful predictions of deaths using our model. We have removed the word “optimal” from the manuscript as it is known that the more you extend the lockdown, (given that it works), there will be further reduction in case-counts. We were using the word “optimal” more loosely in terms of striking a balance between reduction of case-counts and minimizing collateral damages due to other economic and social costs. The reviewer is absolutely right and we have been more cautious with our choice of terms/words in the revised manuscript.

R4.13) Sensitivity analysis – nice to include it, but should be discussed and motivated in the methods section.

The sensitivity of the model to these assumptions requires more explanation and consideration.

Response: We have carried out multiple sensitivity analyses of our model assumptions and have now included more details in the manuscript. Please see Table R2 in response to Reviewer 1, R1.3.

R4.14) Discussion Contains a number of unsupported assertions, that require more explanation. They also suggest that the analysis might be better conducted using additional covariates to be relevant to policymakers.

Response: The reviewer is absolutely correct. We have now removed the recommendation section as they cannot be derived or deduced from the models and were more generic in nature. We mention that takeaway messages from the models inform policymaking and one should not dwell on the exact numbers but focus on relative comparisons. We incorporate some of our suggestions regarding the need for more accurate data, including ecological level covariates in the Discussion section.

R4.15) BCG vaccine hypothesis description is nice but it is unclear whether the model as articulated can test this hypothesis. How does this relate to the ‘uncertainties’ of the model. What percentage of the India population is vaccinated for TB?

Response: We agree that the hypotheses about BCG vaccination, temperature and alike do not directly bear relevance to this present study and so we have removed those points and streamlined the presentation to focus on the epidemiologic forecasting of COVID-19 cases and intervention comparison.

R4.16) The healthcare and social recommendations as articulated are too confidently expressed given the methodological shortcomings of this analysis as presented.

The economic recommendations are wholey unrelated to the analysis as presented.
Response: The reviewer is absolutely correct. We have now removed these recommendations as they cannot be derived or deduced from the models and were more generic in nature. This was perhaps a byproduct of the initial broader audience we were trying to reach through this work. For an HDSR contribution we have tried to stay focused on the modeling and intervention forecasting in the revised manuscript. Thank you for this suggestion.

R4.17) Value of data models – this section needs a lot of work. Up until this section of the manuscript the authors have not articulated the value of their model over others. Nor do they discuss tradeoffs presented by models more generally. The reader will already appreciate that ‘all models are wrong, some are useful.’

Response: We have now removed this entire section “Role of data science and the troubling existence of many models” based on other reviewer suggestion. We have tried to retain some of the general takeaways from this section in the Discussion section. Additionally, we have summarized some India-specific models, their pros and cons in a Table (see Table R1 and response R1.1) which is included in the Supplementary Materials, and is briefly described in the Introduction section. One major limitation with the other India-specific models is the lack of open-source code and clear statement of assumptions which make it challenging to reproduce their projections. We hope that readers can now appreciate the value of this work over other models from India with respect to using a validated eSIR model with publicly available software (https://github.com/lilywang1988/eSIR), providing open source code (https://github.com/umich-cphds/cov-ind-19), an interactive app that updates our prediction models on a daily basis (covind19.org), our GitHub repository that stores past projections, and detailed descriptions of assumptions and parameter choices of our models in the revised manuscript.

Associate Editor

Ray et al. use an eSIR model to predict COVID-19 incidence under different scenarios, including varying lengths of lockdown and varying levels of post-lockdown activities. They consider a Dirichlet prior on the initial vector of latent probabilities, log normal priors for R0 and λ, and Gamma priors for κ, λI, λR. In addition, the authors create a publicly available interactive app with visualizations of their projections. Ray et al. show that a lockdown has a high chance of reducing the total number of COVID-19 cases in the short term. They recommend a lockdown period of 42–56 days with additional post-lockdown social distancing measures to effectively control COVID-19 incidence. In the discussion, the authors also make several healthcare, social, and economic policy recommendations.

This is an important paper that contains timely results for informing COVID-19 policy decision-making in India. Given the current difficulties in controlling the migration of millions of day laborers—many of whom are infectious and will accelerate the spread of the disease—the urgency of publication of this manuscript cannot be overemphasized.

My comments are divided into general and specific points:

Response: We thank the associate editor for their positive and constructive feedback. Please see responses below.

General Comments

AE.1). The use of the eSIR method to project the number of cases is sound. It will be more credible if the authors can cross-validate their priors, especially of R0, with the many other papers in this growing literature. For example, Alimohamadi, Taghdir, and Sepandi (2020) perform a meta-analysis of papers that estimate R0 which can provide the mean and standard deviation for the prior for R0 (https://www.jpmph.org/journal/view.php?doi=10.3961/jpmph.20.076).
Response: Multiple studies on the COVID-19 situation in India have indicated that a value of $R_0$ around 2.0 is a good ballpark estimate to characterize the growth of the pandemic in India [Das (2020), Ranjan (2020), Deb and Majumdar (2020), Singh and Adhikari (2020), Sardar et al (2020)]. We have set the prior mean of $R_0$ at 2.0 (SD=1.0) in the analysis that we present in the main paper. Moreover, we have conducted a sensitivity analysis (see Table R2 and response to reviewer R1.3) where we vary the prior mean for $R_0$. In most of our sensitivity analyses, the posterior mean for $R_0$ was found to be around 1.8-2.4 except when the prior mean is set at a very large value of 4.0 (Table R2). We have also discussed the choice of $R_0$ and the effect of the intervention in response to reviewer R2.3. Following are the posterior distributions of the projected case-counts and latent proportions under various prior choices. We have now included Table R2 in the revised paper and these plots in the Supplementary Materials.

Prior mean of $R_0 = 2$

Prior mean of $R_0 = 3$

Prior mean of $R_0 = 4$

It would be ideal if the authors could provide more details for their computations, similar to what some of the authors have done in a prior paper (e.g. Figure 9 of Wang et al., 2020). The results can be placed in the Supplementary Materials, but should be made explicit to help readers to understand the models.
Response: We have now provided more details of the methods, computational algorithms in the Methods section and provided more supplementary figures related to the Bayesian estimation. For example, we present posterior density and trace plots for the underlying model parameters $\beta$, $\gamma$, and $R_0$. We then provide posterior distributions for the predictions $Y$ and the latent proportions $\theta$ for the I and R compartments over time. Finally, we provide estimates and posterior distribution of the daily prevalence of active cases over time i.e. $\frac{d\theta_I}{dt}$. The advantage of the Bayesian framework is that we can obtain estimates and uncertainty quantification on all these parameters through the exact posterior draws. We have emphasized this aspect of the model in our revised version.

Posterior density and trace plots for $\beta$

![Trace of beta](image1)

![Density of beta](image2)

Posterior density and trace plots for $\gamma$

![Trace of gamma](image3)

![Density of gamma](image4)

Posterior density and trace plots for $R_0$
Posterior distribution for predictions $Y$ and the latent proportions $\theta$ in I and R compartments over time.

Posterior distribution of the daily prevalence of active cases over time or $\frac{d\phi_I}{dt}$.
AE.3. The comparison of the outcomes of different scenarios is interesting, informative and useful. However, the confidence bands are wide, which makes the results somewhat less compelling. Is there anything that can be done to tighten these bands?

Response: We thank the associate editor for mentioning this. It is indeed true that as we move further in time, the uncertainty in the forecasts are quite large. This is quite understated in the media coverage of all these models and projections. One option is to comment only on short term projections, calibrate the models as more data come in and be skeptical about longer term projections. We have focused on uncertainty quantification of the model and calibration with growing data (please see response to reviewer R3.10 and Table R3) in the revised version of the manuscript. Tighter and sharper priors reduce the uncertainty in the model parameters, but the daily forecasts are still quite variable is our experience. We have also tried to communicate that the utility of the models is in terms of relative comparison of interventions and the takeaway messages are not in terms of the exact numbers. Please note response to reviewer R3.7.

AE.4. The sensitivity analysis lacks evidence showing how the simulations change with different assumptions. It would be useful if the authors can generate graphs similar to Figure 5 (and Figure 9 of Wang et al., 2020) for the various assumptions. While the authors assert that the “broad qualitative outcomes” will not change, readers will still want to know how sensitive the predictions are to changes in the assumptions.

Response: We have now provided summary results from our sensitivity analysis in the main text (Table R2 in this response) and figures corresponding to posterior distribution of model parameters in the Supplementary Materials. Please note responses AE.1 and AE.2.

AE.5. The authors offer many healthcare, social, and economic recommendations. Many are sensible but obvious, and some are controversial or impractical. For example, healthcare recommendation (e), “Ensure the healthcare facilities have adequate supply of medications that are currently being recommended” makes sense (albeit trivially obvious), but the authors then go on to cite papers providing conflicting views on the efficacy of the anti-malaria drug chloroquine. What are we to make of this? I would suggest dropping this discussion; it’s a needless distraction mired with political overtones that detract from this paper’s focus. Also, the authors don’t provide any clear solutions as to how the government is to ensure an adequate supply of Remdesivir when the world is facing an acute shortage of the drug. I recommend that the authors exercise restraint in making these broad recommendations lest they be dismissed as ivory tower academics with no grounding in reality.

Response: We agree that the recommendations do not follow from our analyses in this paper. As such, we have removed them and tried to focus on the epidemiologic modeling and forecasting piece. Our only recommendations are now on need for more granular level data and testing (see Discussion section).

AE.6. Despite the many recommendations offered by the authors, they are silent on what to do about the plight of day laborers currently involved in a mass migration back to their homes, many of them on foot either because they cannot afford any other form of transportation or such transportation is no longer available due to the lockdown. Needless to say, these individuals do not wear masks and are likely spreading the disease as they make their way home. Although the authors allude to this issue indirectly in economic recommendation (k), some explicit recognition of this current public health challenge seems important if the paper is to have the impact it deserves. Indeed, the very presence of this large, mobile, and apparently uncontrollable subpopulation undermines the authors’ recommendation of a 42- to 56-day lockdown—have they factored in the impact of this subpopulation into their simulations?

Response: We thank the associate editor for this very important and genuine point. We have removed the broader recommendations segment from the revised manuscript. As for factoring in the impact of specific sub-populations
in our analyses and let the transmission parameters depend on geography, contact networks and other demographic factors in the model, we have added an entire paragraph in the **Limitations** section detailing our inability to do so due to unavailability of granular data. Please also note reviewer responses R3.4 and R4.5.

**AE.7** Finally, the tone of this paper is quite uneven throughout the manuscript, perhaps because different authors contributed different sections. Many passages are written in the usual style of a scientific paper that focuses on a narrow set of findings and implications. Other passages are considerably more florid and heartfelt (e.g., “In these frightening times, we find inspiration in the power of the common people and the magic of human kindness.”). I sympathize with the authors’ passion, but this is more of a stylistic question for the HDSR editor as to whether a more consistent scientific style throughout the article is desirable. Perhaps these issues will be ironed out by the copy editor, but in my experience, most copy editors are loathe to propose such significant changes.

**Response:** Our apologies. The associate editor is absolutely correct that this was an effect of who we tried to reach with the initial work. However, for an HDSR audience such broader generic statements that are not deduced from science are neither relevant, nor appropriate. We have modified our narrative to keep the paper scientific and have removed the “florid and heartfelt” sections/phrasing. We briefly explain the context of this paper in the **Acknowledgement** section so that the main text remains relevant for a HDSR audience.

**AE.8 Specific Comments**

1. **Page 4:** Figure 1 contains graphs for several countries in addition to India but the authors never refer to those countries so I would either delete them or refer to them in some meaningful way.

**Response:** Thanks for pointing out, we have only kept India in Figure 1.

2. **Page 6:** R0 under no intervention was estimated based on early phase data in India; R0 under moderate intervention was estimated based on the data from Wuhan. Is there a reference for the assumed value of 0.8 for R0 under lockdown? This seems to be a particularly important parameter and it would be useful for readers to know the rationale behind the assumed value.

**Response:** We thank the associate editor for this question. Please note our response to reviewer R2.3. In summary, the estimate of $R_0$ is based on early phase data from India, but the intervention effect or $\pi(t)$, the factor which reduces $R_0$ with each intervention is obtained from the data from Wuhan (see below for a reproduced Figure from Pan et al, JAMA, 2020). One limitation of the SIR model is that the $\pi(t)$ for a future time point cannot be estimated in a data adaptive way as it is a part of the forecasting scenario. The changing effective $R_0$ for the past data can be estimated in time intervals as done in the figure below by dividing the intervention time intervals into blocks.

---

![Figure 4: The Effective Reproduction Number ($R_t$) Estimates Based on Laboratory-Confirmed Coronavirus Disease 2019 (COVID-19) Cases in Wuhan, China](https://example.com/figure4.png)

The effective reproduction number $R_t$ is defined as the mean number of secondary cases generated by a typical primary case at time $t$ in a population, calculated for the whole period over a 5-day moving average. Results are shown since January 1, 2020, given the limited number of diagnosed cases and limited diagnosis capacity in December 2019. The darkened horizontal line indicates $R_t=1$, below which sustained transmission is unlikely so long as antitransmission measures are sustained, indicating that the outbreak is under control. The 95% credible intervals (CrIs) are presented as gray shading. Daily estimates of $R_t$ with 95% CrIs are shown in Table 3 in the Supplement.

---

**Figure 4 from Pan et al. 2020**
3. Page 8: The predictions have wide confidence intervals due to the large number of unknowns and the high levels of uncertainties in parameter estimates. How do the projections from 4/8–4/15 (one week from 4/7) compare with the actual reported number of COVID-19 cases? How close or far apart are the predicted numbers and the actual numbers? This comparison can provide some measure of validation and perhaps support the authors’ point that their forecasts are underestimates of the actual situation.

Response: This is an excellent point and also raised by reviewers 3 and 4. Please see our response R3.4, Table R3 and Figure R4.

4. Page 8: Figure 4. What does the horizontal dashed line represent?

Response: Thanks for pointing out; we have clarified this in the revised figure. This is the 95% Bayesian upper CrI.

5. Page 9: Last line, “Figure 6” should be “Figure 7”. Figure 7 needs quite a bit of additional work (or perhaps the authors would prefer to omit it entirely if they can’t address all of these issues in a timely fashion?):

Response: Fig 7 relates to the temperature analysis, which we have completely removed from the revised article.

The sub-figures on the left should have a fixed color bar and a heatmap measured in terms of incidence rate. Are the authors looking at case counts or incidence rate? They mentioned the former in the section title and figures, but write about incidence rates later. If they do look at case counts, isn’t that misleading, especially if one is looking at the finding patterns between incidence rates and temperatures?

Response: This comment relates to Fig 7 (temperature analysis), which we have completely removed from the revised article.

It would be more informative if the color bars for the sub-figures on the right are consistent across all the sub-figures (i.e. same min-max ranges). The readers will then be able to see how the temperatures vary both across latitudes and time.

Response: This comment relates to Fig 7 (temperature analysis), which we have completely removed from the revised article.

It is unclear what the term “north east-west directions” means. I presume that the authors mean “from Northeast to Northwest”? However, the heatmap also shows increases in incidence rate in the Southern Hemisphere. I wonder if the author’s use of this terminology refers to “top-left to top-right versus” instead of anything related to geographical positions.

Response: This comment relates to Fig 7 (temperature analysis), which we have completely removed from the revised article.

The authors compute the correlation between the average monthly temperature and the total monthly incidence, and used the correlation to conclude that there is no correlation between case counts and daily temperature. This conclusion is different from that written in the abstract (“We find some suggestive evidence that the COVID-19 incidence rates worldwide are negatively associated with temperature ...”).

Response: We have removed the temperature analysis from the revised article.

Fundamentally, this analysis may be flawed. Even with a constant R0 and temperature, the incidence rate is expected to change over time. Thus, it is only fair to correlate temperature with the incidence rates after x number of days since cases reached y, similar to what is presented in Figure 2.

Response: We agree, and we have removed the temperature analysis from the revised article.
An ideal analysis will be a multivariate regression of temperature and other factors against the eSIR’s beta but I understand that it is going to be a huge undertaking and may not be feasible given the time constraint.

**Response:** We have completely removed the temperature analysis from the revised article.

6. Page 10: As shown by the authors’ results, government intervention has significant impact on incidence. Instead of computing one correlation coefficient for all countries, I think it would be interesting if the authors can stratify the dataset by levels of intervention (e.g., lockdown, partial lockdown, or social distancing and travel bans only) before performing a regression analysis to investigate the correlation between temperature and incidence. This approach might yield results which are significant.

**Response:** We have completely removed the temperature analysis from the revised article.

7. Figure 6b: What is the intuition behind a higher peak for the 28-day lockdown versus a 21-day lockdown?

**Response:** We have re-run all our models with data up to April 14 and larger number of iterations and we do not see this phenomenon anymore. The previous observation could have been with limited data and inherent variation at further time points with these models that we have noted all along. Please see the revised Figure below.

![Predicted number of daily COVID-19 infections under varying lockdown lengths](image)

8. Figure 5 versus Supplementary Figure 2: Supplementary Figure 2 suggests that a return to normal levels of activities post-lockdown is no worse than social distancing and travel ban measures without lockdown. This contradicts the point made on page 8 that returning to normal activities post-lockdown will lead to a surge beyond social distancing and travel ban measures without lockdown. What is the intuition behind this?

**Response:** Again, please note the revised figures below, now run with observed data up to April 14, it is clear that the two curves for pre-intervention behavior and perpetual social distancing do in fact cross each other in both quick adherence (**new Figure 5a**) and slow adherence (**new Supplementary Figure 3a**). The reason being you quickly lose the benefit of a 21-day lockdown. We do not see this quick crossing under longer durations of lockdown. However, they do eventually cross as we lose the benefits of lockdown without any form of suppression in place and perpetual social distancing makes up the lost ground.
9. Figure 5 versus Supplementary Figure 2: The number of projected cases by 6/15 is ~2.7M under quick adherence and normal return, and is ~2.4M under slow adherence and normal return. This seems to suggest that slow adherence is actually beneficial in the short term. It would be useful if the authors can provide some explanation to avoid misinterpretation.

Response: Again, the long-term forecasts are quite variable. When we re-ran the models with updated data and larger number of iterations, we do not see this property anymore.

10. Figures 5 and 6: I recommend drawing arrows and labels to represent the gap.
Response: Thank you. We implemented this change.

11. Supplementary Figure 4: On page 7, the authors state that they assume the “mean for R0 prior remains at 0.8 for the duration of the lockdown and returns to 1.5 three weeks after the lockdown period ends.” Should this be 1.5 weeks?

Response: The prior mean value for R0 remains at 0.8 for the duration of the lockdown. Post-lockdown, it gradually returns to a value of 1.5 over a span of three weeks.

12. It’ll be useful if the authors can provide more information regarding the estimated models, particularly the convergence plots for the MCMC algorithm.

Response: Please note response to your comment AE.2. We have provided details of the algorithm in the Methods section, and trace plots and posterior density plots in the Supplementary Materials as you suggested.

Dataviz Editor

DVE.1) Fig 1 The title says India, but other countries are shown as well. I think it would be a good idea to point out that the vertical scales are logarithmic and not all the same. It would also be worth mentioning that the time scale is different in the first plot.

Response: We have now removed the other countries. The case-counts in Figure 1 are on the original scale and not on the log scale.

DVE.2) In several plots (Figs 2-6) it would be better to put the names of the curves to the right at the curve ends, instead of having a legend in alphabetic order to the left.

Response: We have done this. Please see an example below. We hope this is what you suggested.

DVE.3) Fig 3 would possibly be better with a different aspect ratio, taller rather than broader. If the intention is to compare the two graphics, then that is difficult at the moment.

Response: We prefer to have time on the horizontal scale as that seems more natural.
DVE.4) Fig 4 Why two y-axes? Better to drop the right-hand one and put the labels to the right.

Response: We implemented this change. Please see modified figure below.

DVE.5) Figs 5 and 6 need better explanation. They seem to be too complicated, including numbers of cases and rates on the same graphic.

Response: Based on the reviewer comments we have replaced the large numbers by relative change from the weakest to strongest intervention. This simplifies the caption inside the figure.

DVE.6) What does Fig 7 show? It is referred to as Figure 6 in the text. I'm not convinced by the temperature analyses; the data seem too uncertain.

Response: We have removed the temperature analysis from the revised manuscript and thus this figure.

DVE.7) In Supplementary table I the numbers should be vertically aligned.

Response: We tried our best to make the Tables look pretty by rounding up the numbers. Please let us know if you have other suggestions.
Editor comments from Dr. Xiao-li Meng

E.1) Your article clearly has generated a lot of interests, comments, as well as criticisms, all are expected -- see attached. Reading through all of them, I got reminded that an important task before diving into the details of the revision is to think through what you really want to accomplish via this article. Currently, it reads as a summary documentation of what you have done, and the major findings and actions generated by the teamwork. The reviewers' various comments suggest that this is not very effective, even though everyone understands that you were under stringent time constraints. The reviewers want more depth, comparisons, details, cautions, and nuanced discussions.

Response: We have tried to make the paper more consistent, careful and focused on the modeling and epidemiological forecasting and removed most of the general and generic statements and florid/emotional language. As the editor can note from the previous 30 pages of response to reviewers, we have carried out many additional analyses including a sensitivity analysis, a calibration analysis and provided more methodological and computational details. We hope the editorial board and the reviewers find our effort to be in the right direction.

E.2) I can see at least three different emphases for a paper as such:

1) Methodologically oriented: showcase how to conduct sound data analysis under so many constraints, from data quality to timeline and to our understanding of the virus and human behavior, etc.

2) Action and Policy oriented: how to provide reasonable evidence under stringent time and other constraints to inform policies and actions, knowing that nothing is perfect, but we need to do our best to minimize at least disasters;

3) Collaboration and Impact oriented: demonstrating how (data) scientists can work together to make a difference for matters of great importance, while being willing to take the risks and responsibilities, including being very wrong.

You can see from the reports, reviewers have different preferences on what your article should be, with for example Review II leaning towards 2) or 3), and Reviewer III more towards 1). But ultimately you need to decide which way you want to emphasize more, or you want to cover all of them. Emphasizing (1) requires you provide as many details as possible (HDSR is very flexible on length and format, so you shouldn't feel too constrained in that regard), so others can follow your approaches and to improve upon. Emphasizing (2) means you need to be much clearer which recommendations are based on solid evidences you build from (1), and which ones are mostly your judgment call (with the understanding of the time constraints). Emphasis (3) seems to me best describes what you currently have, and in some sense, it is most refreshing, but it is also the hardest to do well, since it essentially requires a good mix of both (1) and (2).

Response: We have taken mostly the style of (1) and retained a flair of (2) that forecasting interventions can provide critical input for policymaking but is only one component of it. We have removed all general recommendations that are not corollaries of the modeling work we have undertaken.

E.3) In Reviewer III's email to the AE, s/he wrote "My greatest concern is ensuring that 1) the results are presented appropriately with all of the caveats of these data, and 2) the discussion refrains from making recommendations not substantiated by their research. I am somewhat of a purist in this sense that unless the work explicitly evaluates a policy recommendation, that it should refrain from making claims about such in the discussion." Given what you told me about the history of the paper and your teamwork, it seems that being "purist" is not an option here, since you have already made recommendations to the policy makers and they already have acted upon them (correct?). I think this makes it particularly important for you to present the article
in essentially two parts, each part with clearly stated goals (you can do so in the abstract), so readers will have
the right expectation before getting into the details. (As you know, a sure way to disappoint or even annoying
reviewers and readers is to set the wrong expectation.)

The first part is about building and testing the evidences, where you document all your assumptions, models,
algorithms, codes, and sensitivity studies; I want to emphasize that since HDSR needs to practice what it preaches
about doing the best kind of data science, issues such as uncertainty assessments and checking robustness to
assumptions all should be an integrated part of the analysis and reporting, not deferred to appendix (other than
secondary details). And when performing sensitivity studies, please conduct statistically principled and cost
effective (both computationally and information wise) full factorial experiments, instead of varying one-factor at
a time (as typically done outside of the statistical literature). The second part is about how you go from the
evidences to your recommendations, including how you communicate them (to policy makers and readers in
general) effectively with caveats. I think Reviewer III's main concern is that when you make a recommendation
not being tested, you have much greater potential of being very wrong. But if you express things clearly that you
are aware of the risks and yet are still willing to take them for such and such reasons, then policy makers and
readers at least have a chance to not blindly buy into your recommendations for they may not agree with your
reasons. This second part also permits you to provide various thoughts and discussions, without distracting from
your contributions in the first part. This is essentially what Reviewer III's meant a "stand along commentary",
which I interpreted as that the Reviewer III wants you to clearly separate which part is data science research and
which part is opinion and is of more speculative/guessing nature.

Response: We have adopted the more conventional data science style in the revised manuscript and largely
focused on the first part about data science and modeling for the HDSR audience. Since India had already
announced the 21-day lockdown, then extended to 40 days and now to 53 days with some relaxation, in many
senses our recommendations were considered and implemented. We have shared briefly in the Discussion and
Acknowledgement sections how this project originated, the media coverage it received and the actions that were
taken as we were working on this article to give a sense of that story but the main text is focused on the models.

E.4) I realize that the reviewers are asking for a lot, as they should. I also recognize that we shouldn't let perfect
be enemy of good, especially in cases as such that there is absolutely no perfect answer, not even approximately
perfect. But I do think that if you take all the review comments together and address them seriously (including
spell out the reasons why some of the suggestions are not possible to address at the present time), there is a very
good chance that your paper will stand out as a model for others to follow, both for its methodologies (not about
new ones, but about how to apply existing ones in skillful ways) and for its balanced and clear presentation on
findings and limitations. That of course is also what HDSR wants to showcase -- doing serious data science under
realistic constraints.

Response: We thank the editor for the encouragement and the support. We have tried our best to address the
reviewer comments. The places where we could not were really due to the limitation and paucity of granular level
data.

E.5) As for your presentation style, please consult existing ones in HDSR, particularly the following one, which
was originally also written in your current "medical article" format: https://hdsr.mitpress.mit.edu/pub/fxz7kr65  You will see they turned that into a scholarly research article, permitting more in-depth methodological and investigative discussions. If you want to preserve a form of the current abstract, if in your judgement that's the best way to reach for example medical community, you can do so in the "Media Summary" (without exceeding 400 words), which should be written in plain language (for media to quote directly and accurately). This article was also initially written by different authors with
inconsistent styles, which induced rather negative feedback from multiple reviewers. In the end, one author took the time as needed to go through it line by line, and integrate different styles to make the whole piece much easier to read. Given the broad readership of HDSR, a clear and engaging writing-style is important. It does not have to be written in rigid "scholarly style" (as you may know I don't write in that style myself), since engaging readers is the most important consideration. But because of that, having a coherent style and flow is indeed crucial, as otherwise some readers will be distracted or even get frustrated. I can also ask @Paige to send you the (draft) authors guidelines if you need, but the example above should give you a general sense of HDSR's style, including how to number the sections (e.g., 1, 1.2, but not 1.2.3 -- no sub-subsections).

Response: We thank the editor for his useful suggestions regarding HDSR writing and formatting style. We have tried to follow this style as best as we could and have streamlined the content and presentation style of the revised manuscript extensively. We hope the revised manuscript meets the expectations of the editorial board and is found suitable for publication in HDSR.