Statistics and Applications {ISSN 2452-7395(online)}
Volume 18, No. 1, 2020 (New Series), pp 209-221

What if Lockdown is Removed? District Level Predictions for Maharashtra and Gujarat

Srimanti Dutta[1], Kalyan Das[2], Kashinath Chatterjee[1] and Arindom Chakraborty[1]

[1]*Department of Statistics Visva-Bharati*

[2]*Department of Mathematics IIT Bombay, Powai*

Received: 29 May 2020; Revised: 11 June 2020; Accepted: 12 June 2020

Abstract

Throughout the world everyday, thousands of people are getting infected and hundreds are dying in each country due to an pandemic caused by the outbreak of COVID-19. Starting from Wuhan, this virus has almost travelled each and every country infecting millions of people. In the absence of any vaccine till date and no confirmation of herd immunity (D'Souza and Dowdy (2020)), if any, the world is depending on some non-pharmaceutical interventions (NPIs) to curb the spread of COVID-19. India is no exception. After three consecutive lockdown spells, India has entered into lockdown 4.0 from May 18, 2020, with some relaxations. Lockdown has a huge negative impact on the economy. This study aims to predict the future in few districts of two selected states (Maharashtra and Gujarat) of India if lockdown is removed or partially implemented. A statistical model based on renewal process has been used for prediction. A hierarchical Bayesian method has been used for this data. Predictions have been made till June 30, 2020, for each district of the two states (Maharashtra and Gujarat) under this study.

Key words: Average reproduction number; Renewal process; Hierarchical Bayes; Infections.

AMS Subject Classifications: 62M20, 62F15

1. Introduction

After ages, the world is experiencing something so lethal and frightening. Starting from a fish market in Wuhan, a novel coronavirus named SARS-CoV-2 or COVID-19 has travelled the globe within a short span of time. Over 45 lacs people are already infected and the death toll crossed 3 lacs throughout the world. In comparison with other deadly viruses like Ebola, SARS or MERS, this virus is less fatal but is more contagious in nature (Mahase (2020), Fox (2020)). A recent discovery of a particular mutation, found in India, claimed that this virus has become more contagious (See Korber *et al.* (2020)).

Till date, no vaccine is available for SARS-CoV-2. A lot of clinical trials are going on throughout the world to discover some "effective" vaccines. From previous experience, we have observed that it took years to find any "effective" vaccine. In this situation, the

Correponding Author: Arindom Chakraborty
Email: arindom.chakraborty@visva-bharati.ac.in

world is relying more on non-pharmaceutical interventions (NPIs) to curb the spread of this virus. Closing of educational institutions, ban on travel, maintaining physical distance were taken as NPIs by different countries at different point of time. However, these NPIs were not sufficient enough to tackle the spread. Knowing most of the adverse consequences, many countries were compelled to impose lockdown as the last resort to battle against COVID-19. Unlike most of the European countries, in India, the lockdown was not maintained properly in most of the parts. As a consequence, infection spread throughout the country.

It is always a challenge to epidemiologists to predict the progress of infection caused by an unknown virus. A lot of people are working in this direction individually as well as in groups. Commonly used SER, SEIR, adaptive SEIR are more common among epidemiologists which are basically based on differential equations. As an alternative, statisticians are using some probabilistic models to capture the uncertainty. It is imperative to use Bayesian inference in all models as we have little or no experience in this SARS-CoV-2. The data available from Wuhan gives us some idea of the spread. However, it is to be noted that, fatality or attack rate depend on several factors like genetic profile, particular mutation of the virus and on many confounding factors. It is well known that, even if in Bayesian inference, we borrow prior information from some other virus. As an example for SARS-CoV-2, priors are chosen on the basis of data on SARS infection which happened a few years ago. This sort of assumptions always put any prediction model under severe threat of reliability. In a recent paper published by Luo (2020) claimed that 97% of active cases would have been solved by May 27, 2020, which received strong criticism.

Specific to the Indian scenario, the first work that got some acclamation was done by a group of biostatisticians and epidemiologists at the University of Michigan (The Covid-19 India Group, 2020). They have used adaptive SEIR model and predictions have been made on the basis of this model. Gompertz model has been considered by Lee, Lei and Mullick (2020). A lot of work has been done by the researchers from Imperial College of London (ICL) (Seth *et al.* (2020), Walker *et al.* (2020)). The effects of non-pharmaceutical interventions have been studied for European countries. Another impressive model has been developed by the researchers belonging to IHME, University of Washington, Seattle (UW) (IHME, 2020). In a recent work done by Chatterjee (2020), it has been shown that the models developed by ICL and UW are so far giving good results in terms of prediction. In this work, we have implemented a model induced from ICL model. A question that people are interested in is that when this pandemic will come to an end. From previous studies, we have some idea about the disease progression of several viruses like SARS or MERS. Researchers are looking forward to the time point where asymptote will be achieved. It may be noted that the Government is trying to implement restrictions, if any, in micro levels to minimize economic loss. For the policymakers, it becomes necessary to have an idea in micro-level. The model that we are using in the present work is quite general and can be implemented at any region provided sufficient data is available.

The paper is structured as follows. In Section 2, we describe the methods briefly. Results of the districts are mentioned in Section 3. It is followed by a discussion section.

2. Methods

To predict the time-varying reproduction number (R_t) under various levels of lockdown, the reproduction number at time t is taken as a scale multiple of the baseline reproduction number. The multiplicity factor is a constant function of the lockdown levels. Note that we insert four dummy variables for the above four lockdown levels. The number of infected cases at any day can be predicted using the reproduction number and the weighted average of previous days' affected figures with the discretized serial interval distribution probability of secondary infection as weights (See Fraser (2007), Cori *et al.* (2013), Nouvellete (2018), Cauchemez *et al.* (2008)). The mortality at any day can similarly be predicted using the case fatality ratio (CFR) and the weighted average of previously affected figures with the chance of mortality as weights. Mathematically speaking, there are three models working together. These are models for infections, deaths and average reproduction rate. Let $I_{s,t}$, $D_{s,t}$ and $R_{s,t}$ be, at time t, number of new infections, number of new deaths and average reproduction number for the state s. Moreover, let $g(t)$ be the serial interval distribution. As explained above, we then express the infection on t th day as

$$I_{s,t} = R_{s,t} \sum_{\tau=0}^{t-1} I_{s,\tau} g_{t-\tau} \tag{1}$$

We discretize the serial interval as:

$$g_u = \int_{u-.5}^{u+.5} g(\tau) d\tau; u = 2, 3, 4, ... \tag{2}$$

with $g_1 = \int_0^{1.5} g(\tau) d\tau$ where from the past experience, $g(t)$ is assumed to be a gamma distribution with mean 6.5 (average time from onset in a primary infection to onset in a secondary infection) and a relatively small coefficient of variation 0.62 (Seth et al. (2020)).

For mortality, the observed number of deaths may be assumed to follow a Negative Binomial law where the expected deaths are assumed to be the weighted average of the daily infection, weights being a mixture of two gamma distributions that account for the incubation period and time between the onset of symptoms and death. In case of $R_{d,t}$ (the average reproduction number for a state d), we use levels of lockdown as covariates. Let $R_{d,0}$ be the baseline reproduction number for dth district. Then $R_{d,t}$ is modelled as

$$R_{d,t} = AF_d * R_{d,0} * (2 * \phi^{-1}(-\alpha * \delta)), \tag{3}$$

where AF_d is the adjustment factor considered for district d based on its population, δ indicates the level of lockdown: 0 (no lockdown), 1 (partial lockdown) or 2 (complete lockdown) and ϕ^{-1} is the inverse logit or Sigmoid function. It may be noted that partial lockdown is a policy decision which can be quantified. The Government or the local administration may decide the extent of lockdown which may qualify as "partial". We further assume

$$\alpha \sim N(0, 0.5)$$

To model the baseline reproduction number for a district d, we assume

$$R_{d,0} \sim N(3.28, \kappa)$$

where $\kappa \sim N^{+}(0, 0.5)$, N^{+} denotes a half normal distribution with positive support.

To fit the model we use 6 sequential days of an equal number of infections: $I_{1,d} = .. = I_{6,d} \sim Exponential(1/\tau)$, where $\tau \sim Exponential(0.03)$.

MCMC samples are drawn from posterior distributions using Stan software and convergence criteria have been studied in details.

3. Results

Here we apply the proposed methodology for all the districts of Maharashtra and Gujarat. Daily infection and death data have been considered. In this study, for each district, data on 83 time points (from 2nd March, 2020 to 23rd May, 2020) are used. A few districts with no death or very few cases have been excluded from the study.

We have tried to predict the number of daily infections and daily deaths till 30th June, 2020 for each district under different levels of lockdown. Base reproduction rate and reproduction rate on 30th June, 2020 under different levels of lockdown, have also been reported.

In Figure 1 the values of reproduction rate R_t for different districts are given. Figure 1(a) depicts the baseline reproduction number R_0 values for each district belonging to Maharashtra. Aurangabad has the highest R_0 closely followed by Solapur. For Mumbai and Pune, R_0 values are over 3.3, which is also high.

As expected it is found that lockdown for a prolonged period has a positive effect in controlling R_t. After 31st May 2020 even if lockdown is completely removed, R_t values are less than 1.2 (Figure 1(b)) for all the districts of Maharashtra. Our purpose is to see which stage of lockdown pushes the value of R_t to smaller than 1. This may not be achieved, as in Figure 1(d), even if we impose complete lockdown for the entire month of June 2020. For districts like Pune and Mumbai, the situation remains almost identical even if lockdown is extended to its highest level from partial restrictions (Figure 1(c)). It may be noted that for a densely populated district like Mumbai, a reduction of a decimal place in R_t value may result in the reduction of a large number of infections.

In Figures 2(a)-2(l), predictions for four important districts (Mumbai, Nagpur, Pune and Thane) have been considered. Results for other districts are also available. Three different levels of lockdown have been considered for each of these four districts.

In the case of Mumbai, a surge may be observed if lockdown is withdrawn totally. Daily infection may cross 1000 marks within the first week of June and there will be exponential growth (Figure 2(a)). This may put the existing healthcare system into tremendous pressure. On the other hand, complete lockdown may reduce the rate of daily infection but may result in severe economic depression to this financial capital of India. In this tricky situation, the Government may opt for some intermediate solution where partial lockdown may be imposed. From Figure 2(c) it can be seen that the daily infection rate may increase slowly if partial lockdown is imposed. Similar features may be observed for the other three districts.

Pune may observe daily infected counts just over 100 if partial lockdown is maintained (Figure 2(h)). Removing lockdown may create a similar situation like Mumbai, where daily count may exceed 10,000 (Figure 2(g)). For Thane, it may take end of June to reach the daily count to be 10,000 if lockdown is removed completely (Figure 2(j)). On the contrary, complete lockdown may bring the rate of daily infection down to a great extent (Figure 2(l)). In the case of Nagpur, similar sort of inferences may be made (Figures 2(d)-2(f)).

In Figures 3(a)-3(l), we have plotted the predictions for a few districts (Aurangabad, Nashik, Akola and Palghar) which are less affected, compared to the districts we have already considered, till date. Here also it is seen that removal of complete lockdown may bring the less affected districts to highly affected districts. A partial lockdown may help the administration to maintain the situation which may be controlled with existing health care set up.

In terms of fatality, so far, COVID-19 is less fatal compared to MERS or SARS. A few probable confounding factors like the effect of malaria or BCG vaccine or any other factor or combination of factors may have contributed to this cause. For most of the districts, an increasing trend in the daily number of deaths may be seen. Unlike regions of Europe or states of the US, the death toll is relatively very low for the districts of India.

For Mumbai, the removal of lockdown may result in more than 100 daily deaths at the end of June (Figure 4(a)). A complete lockdown may bring the rate down significantly (Figure 4(c)), whereas a partial lockdown may result in less than 10 deaths per day.

Among all states in India, Gujarat has the highest baseline reproduction number for which the number of infected people will be higher. From Figure 5(a), it can be seen that for districts like Ahmedabad and Surat this value is above 3.5. Lockdown for more than two months has done something significant in reducing values to a great extent. Figure 5(c) shows that the expected reproduction number on 30th June will be less than 1 for most of the districts if partial lockdown is imposed. For the two effected districts Ahmedabad and Surat, the R_t may cross the threshold of 1. Further stringent intervention i.e. complete lockdown may further reduce R_t values.

Like Mumbai, Ahmedabad is also severely affected due to high population density. Removing lockdown completely would be very fatal and it is evident from Figure 6(a). A partial lockdown may help in curbing the spread (Figure 6(b)) whereas complete lockdown may contain the spread (Figure 6(c)). However, districts like Ahmedabad and Surat (Figures 6(d)-6(f)) complete lockdown may have a huge negative impact on economy. Some mitigation policies may be adopted. Vadodara, another important district shows a similar pattern (Figures 6(g)-6(i)). Daily death predictions for two districts, Ahmedabad and Surat are given in Figures 7(a)-7(f). If lockdown is removed completely, the daily death toll may cross 10 from 1st June 2020 and then an exponential growth may be anticipated. However, in a partial lockdown situation death rate will increase in a more constant manner. In the case of Surat, a similar trend is found.

It is to be mentioned here that in reality, the death figures may be much higher. We are assuming that the fatality rate is the same for all the age groups. From Wuhan data, it is found that the death rate for individuals above 60 is much higher compared to other age groups. Moreover, patients with complex diseases like COPD, CVD, diabetes, hypertension have a higher probability of dying.

4. Discussion

In this study, we have considered different levels of lockdown. The proposed model is very general in nature and hence may incorporate several other covariates. For example, the effect of another important NPI *viz.* the role of intensive testing may also contribute to a reduction of R_t values which eventually reduce the number of infections. This study is under investigation and we plan to communicate it soon. While predicting, we have assumed that if there is any non-pharmaceutical intervention (NPI), it needs to be adhered to completely.

It may be noted that, after lockdown 4.0, complete lockdown may not be applicable any more. Relaxing lockdown norms will help the economy but at the same time will increase the chance of infection with increased mobility. We feel that this work may give an idea to the policymakers to chalk out future plan district-wise. As mentioned earlier, this can be done for subdivision, block or even a small region provided sufficient data is at our disposal.

The model we have used is based on a renewal process. A comparison may be of interest to see the change in inferences if models are changed. Another important issue in modelling the spread of any viral disease is the reproduction rate model. A comparison can also be made using some other types of modelling strategies like using Weibull model.

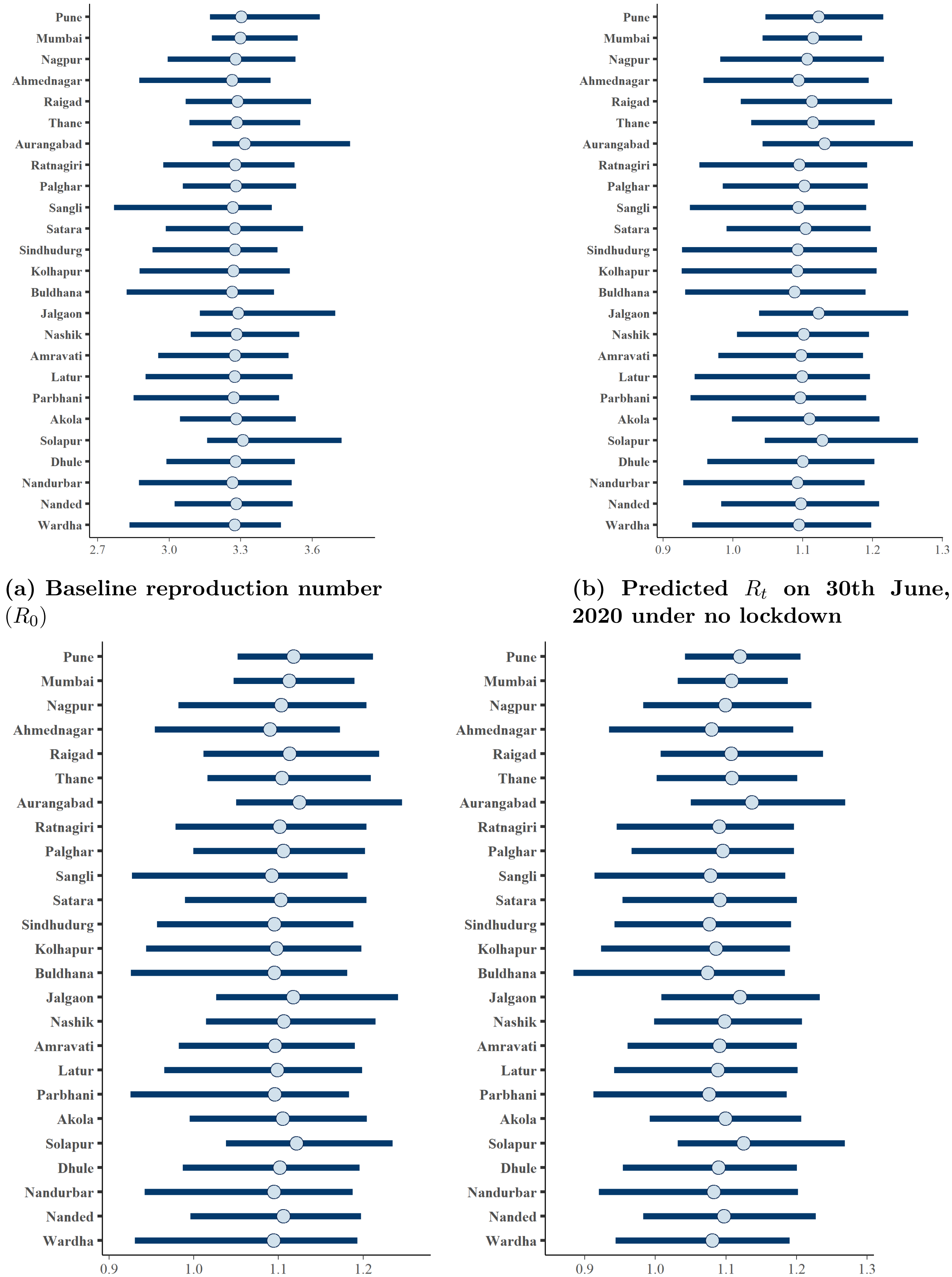
Another interesting and challenging job is to take care of asymptomatic cases which are very crucial in the spread of COVID-19. In recent times it has been observed that due to repeated mutations of this SARS-CoV-2 virus more infected people are becoming asymptomatic. This puts a lot of people at risk as the spreader does not know about his/her infection. More and more tests are desirable in this scenario. Pool testing may be a good option for the regions which are designated as containment zones. Pool testing will decrease the use of test kits and results may be available much sooner.

Getting good quality data is always a big challenge. Out of approximately 150000 individuals around 8% data do not contain district information. Same is true for data related deaths. This may result in biased estimates of the parameters involved and may hamper predictions.

Whatsoever mitigation policies the Government may take, it is essentially the duty of each and every citizen to abide by all the interventions with utmost sincerity. Regulations given by Ministry of Health and Family Welfare or Ministry of Home Affairs or local administrator may be followed religiously. Trace, test and quarantine (TTQ) is the need of the hour.

5. Data and Software

For this work we have used the data available in https://github.com/covid19india/api. All computations have been done using RStudio and Stan.



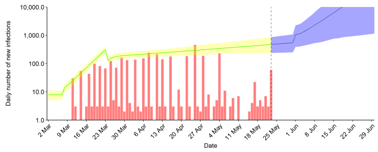
(a) Baseline reproduction number (R_0)

(b) Predicted R_t on 30th June, 2020 under no lockdown

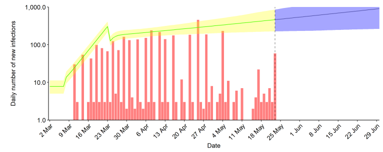
(c) Predicted R_t on 30th June, 2020 under partial lockdown

(d) Predicted R_t on 30th June, 2020 under complete lockdown

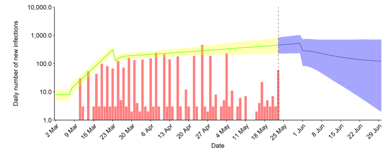
Figure 1: Average reproduction number for Maharashtra



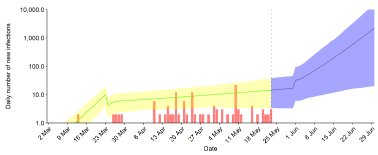
(a) Mumbai: under no further lockdown



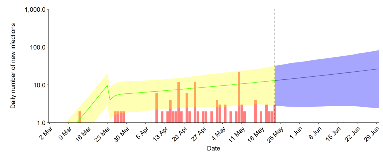
(b) Mumbai: under partial lockdown



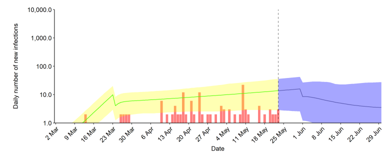
(c) Mumbai: under complete lockdown



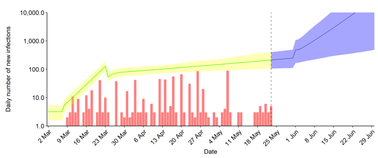
(d) Nagpur:Under no further lockdown



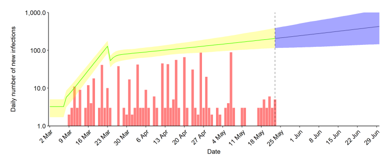
(e) Nagpur:Under partial lockdown



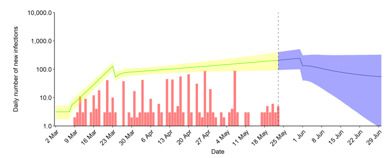
(f) Nagpur:Under complete lockdown



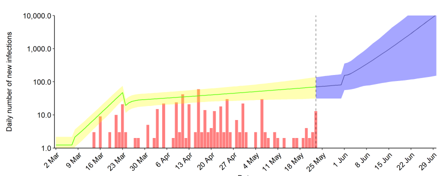
(g) Pune: under no further lockdown



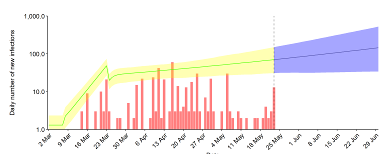
(h) Pune: under partial lockdown



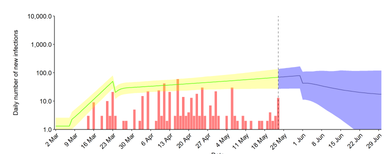
(i) Pune: under complete lockdown



(j) Thane: under no further lockdown

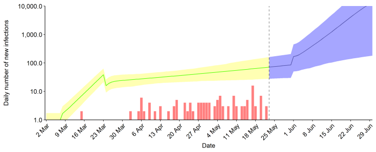


(k) Thane: under partial lockdown

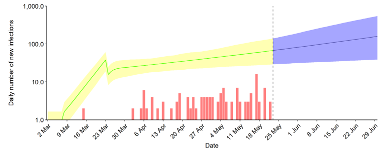


(l) Thane: under partial lockdown

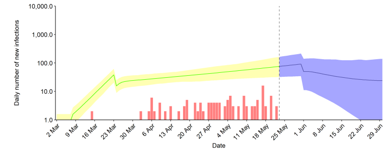
Figure 2: Prediction for different districts of Maharashtra



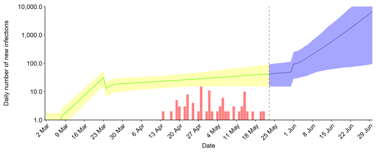
(a) Aurangabad: under no further lockdown



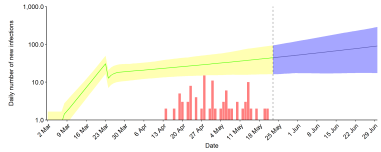
(b) Aurangabad: under partial lockdown



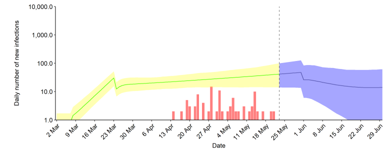
(c) Aurangabad: under complete lockdown



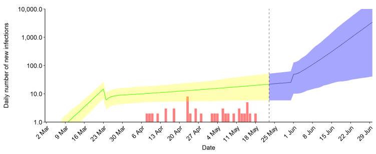
(d) Nashik:Under no further lockdown



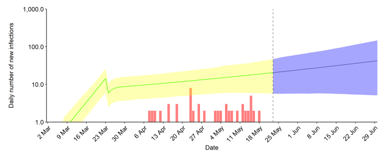
(e) Nashik:Under partial lockdown



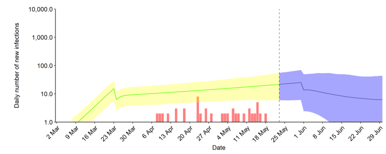
(f) Nashik:Under complete lockdown



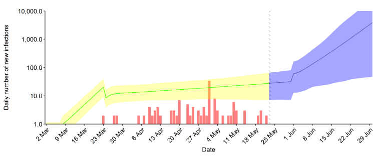
(g) Akola: under no further lockdown



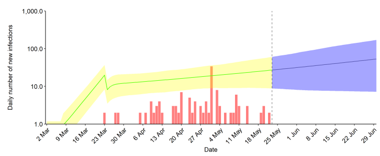
(h) Akola: under partial lockdown



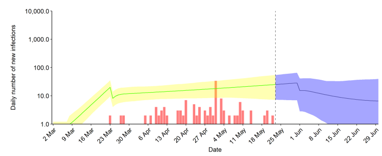
(i) Akola: under complete lockdown



(j) Palghar: under no further lockdown

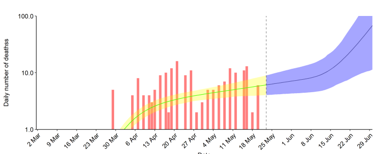


(k) Palghar: under partial lockdown

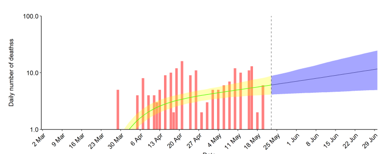


(l) Palghar: under partial lockdown

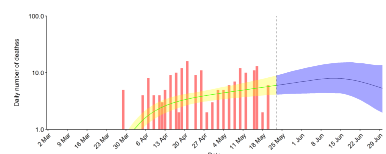
Figure 3: Prediction for different districts of Maharashtra



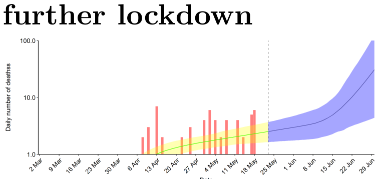
(a) Mumbai: under no further lockdown



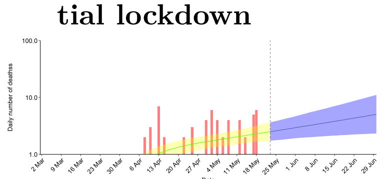
(b) Mumbai: under partial lockdown



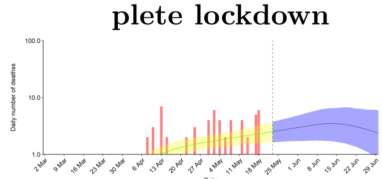
(c) Mumbai: under complete lockdown



(d) Pune: under no further lockdown

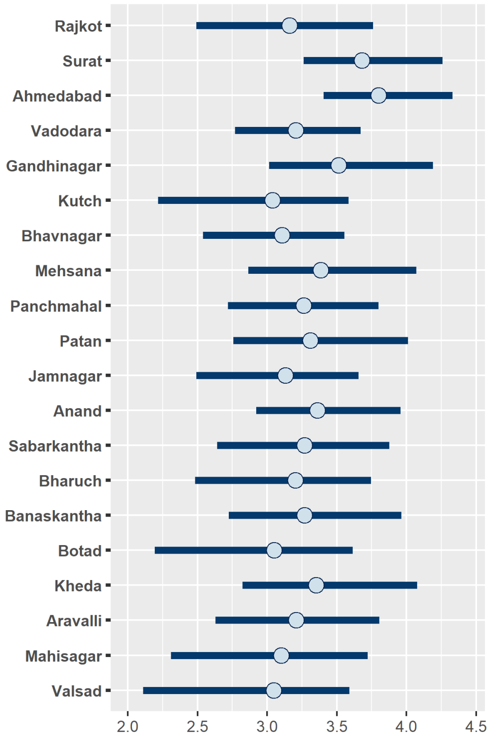


(e) Pune: under partial lockdown

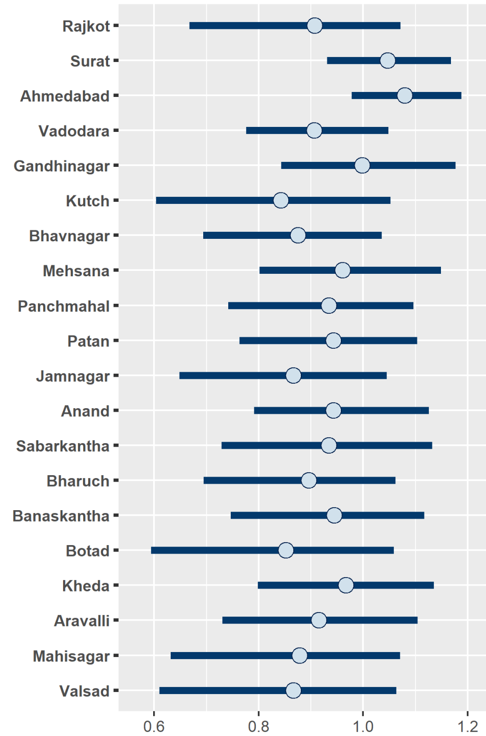


(f) Pune: under complete lockdown

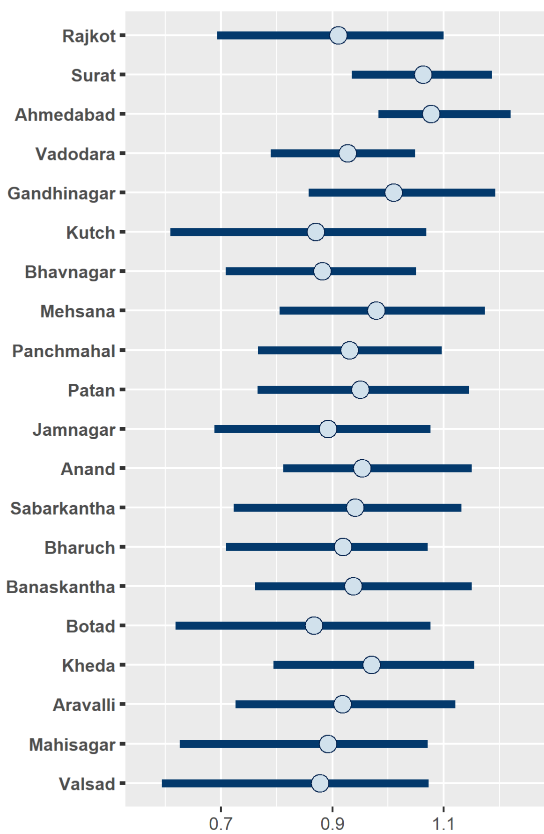
Figure 4: Daily death prediction for different districts of Maharashtra



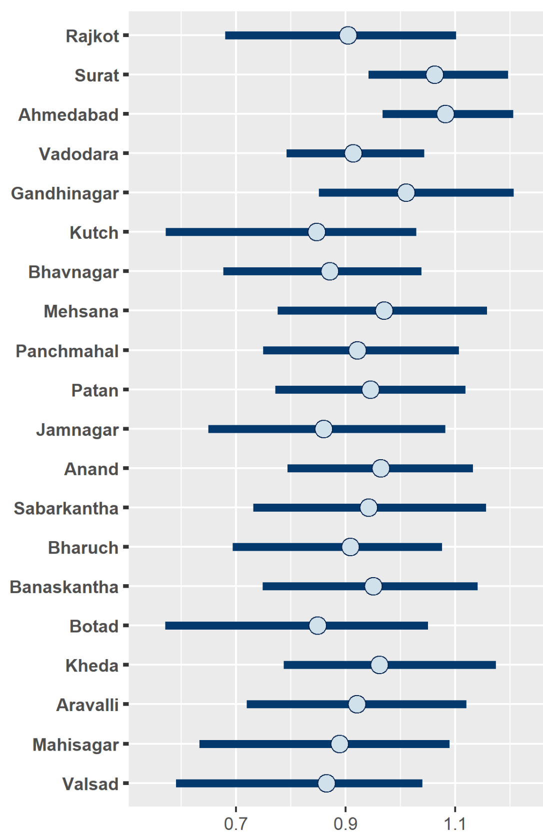
(a) Baseline reproduction number(R_0)



(b) Predicted R_t on 30th June, 2020 under no lockdown

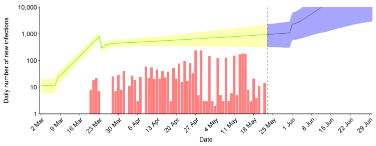


(c) Predicted R_t on 30th June, 2020 under partial lockdown

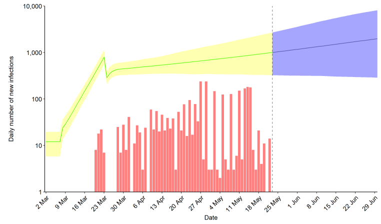


(d) Predicted R_t on 30th June, 2020 under complete lockdown

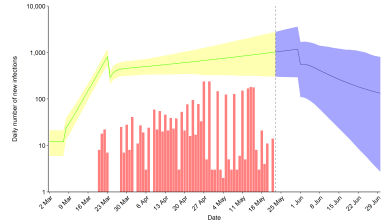
Figure 5: Average reproduction number for Maharashtra



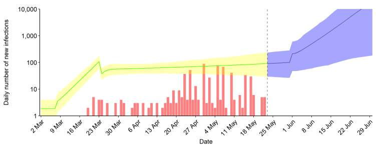
(a) Ahmedabad: under no further lockdown



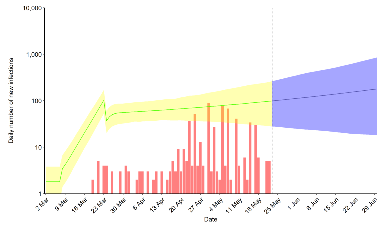
(b) Ahmedabad: under partial lockdown



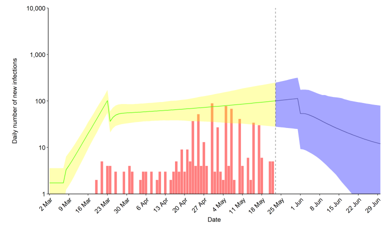
(c) Ahmedabad: under complete lockdown



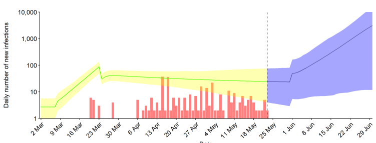
(d) Surat: under no further lockdown



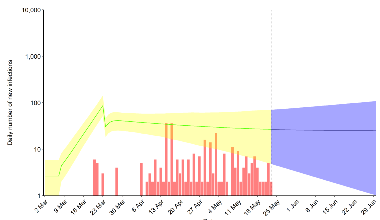
(e) Surat: under partial lockdown



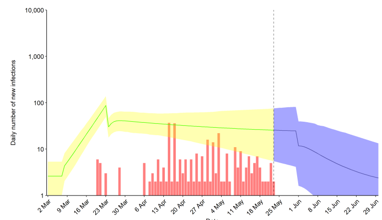
(f) Surat: under complete lockdown



(g) Vadodora: under no further lockdown

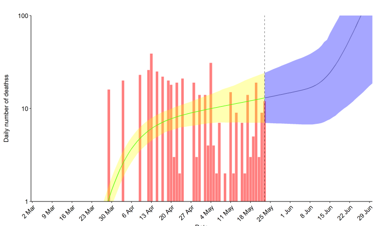


(h) Vadodora: under partial lockdown

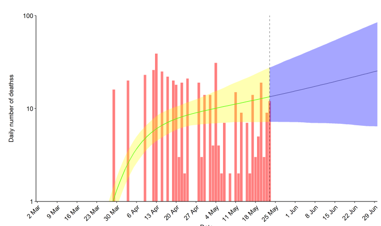


(i) Vadodora: under complete lockdown

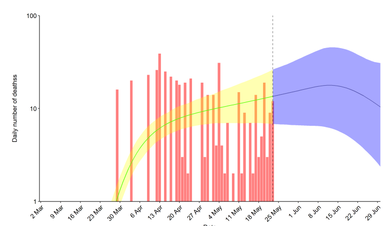
Figure 6: Daily infection prediction for different districts of Gujarat



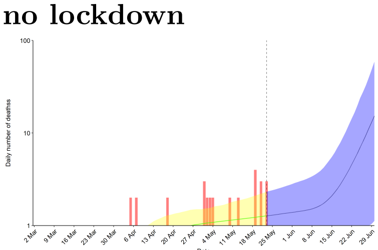
(a) Ahmedabad: under no lockdown



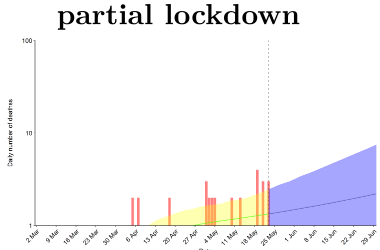
(b) Ahmedabad: under partial lockdown



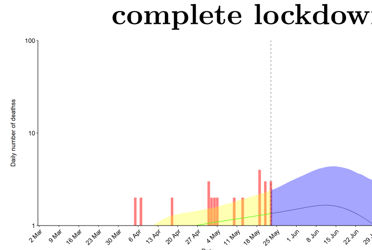
(c) Ahmedabad: under complete lockdown



(d) Surat: under no lockdown



(e) Surat: under partial lockdown



(f) Surat: under complete lockdown

Figure 7: Daily death prediction for different districts of Gujarat

Acknowledgments

The authors would like to thank the Editor and the referee for providing constructive suggestions leading to improvements in the presentation of this paper.

References

Cauchemez, S., Valleron, A. J., Boëlle, P. Y., Flahault, A. and Ferguson, N. M. (2008). Estimating the impact of school closure on influenza transmission from sentinel data. *Nature* **452**, 750–754.

Chatterjee, N. (2020). Transparency, Reproducibility, and Validity of COVID-19 Projection Models. *https://towardsdatascience.com/transparency-reproducibility-and-validity-of-covid-19-projection-models-78592e029f28*

Christopher, J. L. M. (IHME COVID-19 health service utilization forecasting team) (2020). Forecasting the impact of the first wave of the COVID-19 pandemic on hospital demand and deaths for the USA and European Economic Area countries. *https://doi.org/10.1101/2020.04.21.20074732.*

Cori, A., Ferguson, N. M., Fraser, C. and Cauchemez, S. (2013). A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology*, **178**, 1505–1512.

D'Souza, G. and Dowdy, D. (2020). What is Herd Immunity and How Can We Achieve It With COVID-19? *https://www.jhsph.edu/covid-19/articles/achieving-herd-immunity-with-covid19.html* **April 10**.

Fox, Justin (2020). How Bad Is the Coronavirus? Let's Run the Numbers. *Bloomberg Opinions* https://www.bloomberg.com/opinion/articles/2020-03-05/how-bad-is-the-coronavirus-let-s-compare-with-sars-ebola-flu.

Fraser, C. (2007). Estimating individual and household reproduction numbers in an emerging epidemic. *PLoS One*, **2(8)**, e758.

Korber, B., Fischer, W. F., Gnanakaran, S., Yoon, H. et al. Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2. *bioRxiv 2020.04.29.069054. https://doi.org/10.1101/2020.04.29.069054*

Lee, Y. S., Lei, B. and Mallick, B. (2020). Estimation of COVID-19 spread curves integrating global data and borrowing information. *medRxiv 2020.04.23.20077065*, *https://doi.org/10.1101/2020.04.23.20077065*

Luo, J. (2020). When Will COVID-19 End? Data-Driven Prediction *https://www.in.gr/wp-content/uploads/2020/05/covid19predictionpaper.pdf* https://ddi.sutd.edu.sg/

Mahase, Elisabeth (2020). Coronavirus: covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate*The British Medical Journal*, **368**. doi: https://doi.org/10.1136/bmj.m641

Nouvellet, P. et al. (2018). A simple approach to measure transmissibility and forecast incidence. *Epidemics,* **22**, 29–35, *doi:10.1016/j.epidem.2017.02.012.*

Seth, Flaxman, Mishra, S., Gandy, A. et al. (2020). Estimating the number of infections and the impact of nonpharmaceutical interventions on COVID-19 in 11 European countries. *Imperial College London*, *https://doi.org/10.25561/77731.*

Walker, P. G. T., Whittaker, C. et al. (2020). The Global Impact of COVID-19 and Strategies for Mitigation and Suppression. *Imperial College London*, *https://doi.org/10.25561/77735*.