

# What Explains India's Second Wave of COVID-19 Infections?

Subhonmesh Bose<sup>1\*</sup>, Shuai Hao<sup>2</sup>, Sebastian Souyris<sup>2</sup>, Albert Charles England III<sup>3</sup>, Anton Ivanov<sup>2</sup>, Ujjal Kumar Mukherjee<sup>2</sup>, Sridhar Seshadri<sup>2</sup>

<sup>1</sup>Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign; Illinois, USA

<sup>2</sup>Gies College of Business, University of Illinois at Urbana-Champaign; Illinois, USA

Each affiliation should be a separate paragraph.

<sup>3</sup>OSF HealthCare Heart of Mary Medical Center; Urbana, Illinois, USA

\* Corresponding author. 4058 ECE Building, 306 N Wright St., Urbana, IL 61801, [bores@illinois.edu](mailto:bores@illinois.edu), +1-217-244-2101.

**DRAFT 07/11/2021**

## Summary

India experienced a rapid surge in COVID-19 infections during March-April 2021 that overwhelmed the healthcare system. This paper shows that the circulation of the Delta variant of the SARS-CoV-2 virus, amplified by a super-spreader event, likely caused India's second wave. We analyze publicly available district-wise data of COVID-19 infections in India over 2020 and 2021. We use a combination of time-series analysis, epidemiological modeling, and parameter estimation methods. The data shows that a super-spreader event (Kumbh Mela festival being the possible candidate) is probable responsible for circulating the highly transmissible B.1.617.2 (Delta) variant of the SARS-CoV-2 virus that caused India's second wave. Thus, the Indian experience serves as a cautionary tale that calls for increased genomic sequencing to identify

variants of concern and to regulate super-spreader events, while vaccination remains the only long-term solution.

## Introduction

COVID-19 in India was on the decline between December 2020 and February 2021. A rapid spike in infections appeared in late March 2021. The swiftness of the onslaught left millions suffering from significant shortages in emergency medical supplies and hospital capacities. The death toll from COVID-19 in India on May 25, 2021, became the highest for any nation in a single day. This surge in cases is termed as the “second wave” of COVID-19—the first wave in mid-2020 had largely dissipated by then. In this paper, we analyze data from the reported COVID-19 cases across India to find plausible causes for the second wave.

The second wave differs from the first wave in two ways. First, the increase in infections during the second wave coincides with the growth of a highly transmissible random mutation B.1.617.2 of the SARS-CoV-2 virus <sup>1</sup>, now referred to as the “Delta”<sup>1</sup> variant that was first identified in Maharashtra in late 2020. Second, the growth of new infections in the second wave is sharp and almost simultaneous over most parts of India. The first wave, on the other hand, resembles a spatio-temporal diffusion. We contrast the dissimilar growth patterns of the disease in the two waves to deduce plausible reasons behind the second wave.

We demonstrate that the infection propagation dynamics of the second wave cannot be explained by an epidemiological diffusion model. Our cross-covariance studies on COVID-19

---

<sup>1</sup> The B.1.617.2 (Delta) variant was identified as a variant of concern by the World Health Organization on May 11, 2021. It is 1.64 times more transmissible than the B.1.1.7 (Alpha) variant, or 2.3-3.2 times more transmissible than the wild-type SARS-CoV-2 virus. Hospitalization rates have been 1.5-1.8 times higher than that after infection with the Alpha variant and 2.3-2.8 times higher after infection with the wild-type virus <sup>15,21</sup>.

infection data identifies Maharashtra (where the B.1.617.2 variant was first identified) and the Ganges belt as the likely origins of the second wave. Therefore, we posit that the second wave is a result of rapid circulation of the Delta variant by a super-spreader event (the baths at Kumbh Mela in March being the likely candidate).

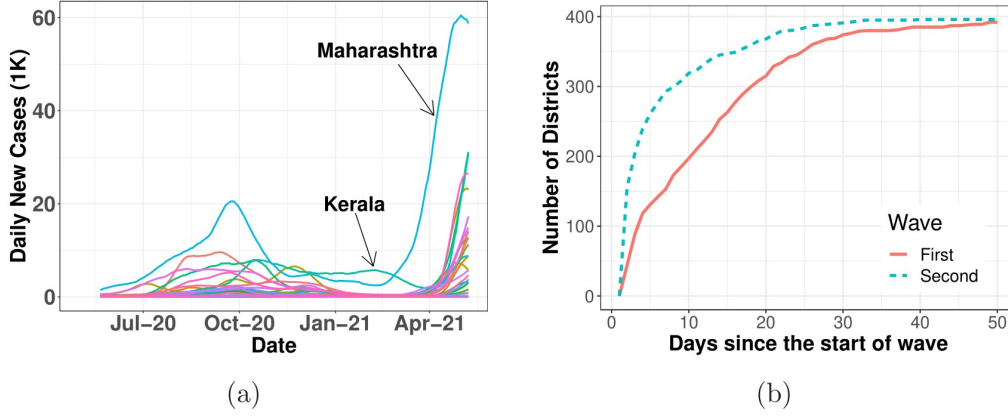


Figure 1. Plot (a) shows the number of daily new COVID-19 cases for all states in India, smoothed via a 21-day moving average. Plot (b) illustrates the number of districts with 5 or more cumulative new COVID-19 cases since the start dates of the two waves.

India witnessed declining levels of daily new infections between the two waves. This “dip” is puzzling, given that lockdown measures were largely relaxed during this period. Inadequate testing/reporting during this dip and possibly higher virulence of the Delta variant (leading to more hospitalizations) could have contributed to the surge. Without additional evidence, it remains challenging to quantify the roles played by the mutation, a super-spreader, and under-reporting.

The rest of the paper is organized as follows. We first describe the differences in the spatio-temporal signatures of the two waves. Then, we employ cross-covariance among time-series of COVID-19 infections in various districts to study possible sources of the infections in the two waves. Then, we fit a compartmental epidemiological model to the data to comment on the lack of fit during the second wave. We conclude with policy recommendations based on the Indian experience.

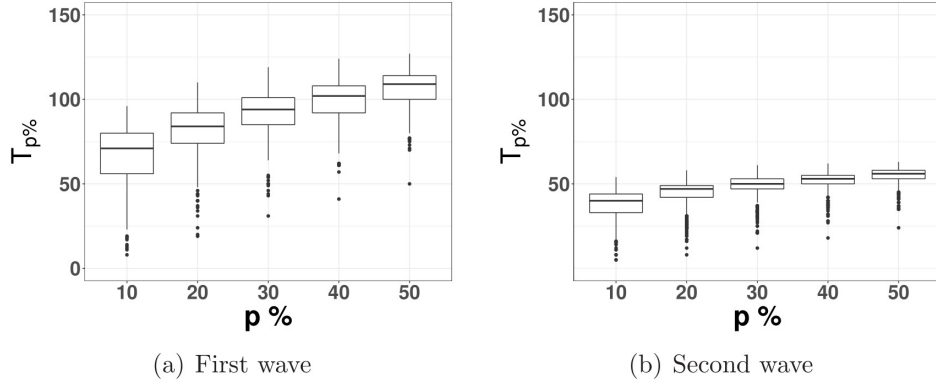


Figure 2. Box plot in (a) portrays  $T_{p\%}$  across 394 districts during the first wave. The box plot in (b) portrays the same for the second wave.

### The Simultaneity of Infections in the Second Wave

Consider the daily number of COVID-19 infections in various states in Figure 1(a), drawn using district-wise COVID-19 positivity results from <https://www.covid19india.org/>. We sub-select 394 districts among the 675 districts in India with complete data of daily new positive cases from 05/01/2020 to 05/08/2021; see our data repository. We aggregate the data from the districts to compute state-wise infection levels in Figure 1(a). Notice that the state of Maharashtra shows a steady growth of infections since January 2021, where the Delta variant was first detected in late 2020. In all other states, the second wave does not appear till April 2021, when the infections abruptly rise everywhere. We consider 05/01/2020 to 10/31/2020 and 03/01/2021 to 05/08/2021 as the date ranges for the first and the second waves, respectively. With these date ranges, consider the number of districts with  $\geq 5$  cumulative number of COVID-19 infections since the starts of the two waves in Figure 1(b). Evidently, the first wave exhibits a slower growth in the number of districts with new infections than the second wave.

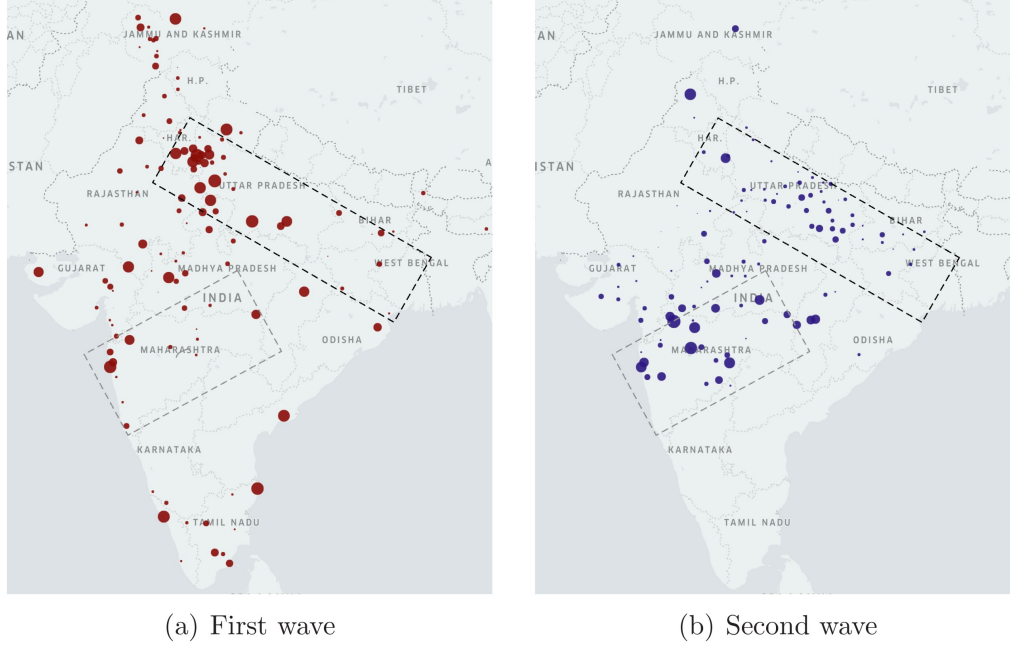


Figure 3. Bubble plots of  $\kappa_*$  for districts that have  $\kappa_* > 0$  for the first wave in (a) and the second wave in (b). The size of the bubble indicates the magnitude of  $\kappa_*$ . The dotted areas identify Maharashtra and the Ganges belt.

To further illustrate the simultaneity of the second wave, consider the time elapsed  $T_{p\%}$  since the start dates of the two waves to reach  $p\%$  of their levels on 09/15/2020 and 05/08/2021, respectively. The box plot of  $T_{p\%}$  for  $p = 10, 20, 30, 40, 50$  in Figure 2 clearly show a larger *spread* among  $T_{p\%}$ 's in the first wave (Figure 2(a)) compared to that in the second wave (Figure 2(b)).

## Origins of the COVID-19 Waves in India

Cross-covariance of time-series data can shed light on the causal effect of one time-series on the other. We use cross-covariance among the time series of COVID-19 infections across 394 districts to identify this causal effect, ignoring the impacts of possible inadequate testing and erroneous reporting.

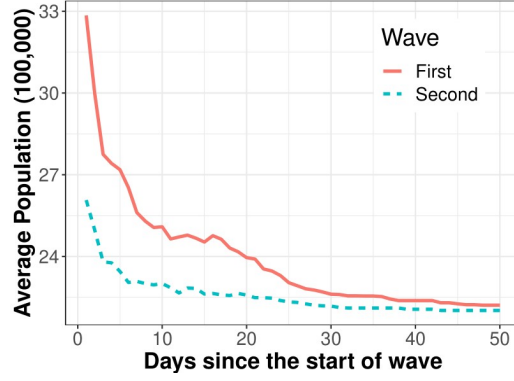


Figure 4. Variation of average population of districts with cumulative number of 5 or more infections since the start of the waves as in Figure 1. Start dates of the two waves are 05/01/2020 and 03/01/2021.

Consider the sequence of new daily COVID-19 infections in district  $i$  on day  $t$  as  $\rho_t^i$ , where  $t$  is measured as days since the start dates of each wave. Then, the cross-covariance function over  $T$  time-periods is described by

$$\mathcal{C}(i, j; \kappa) := \frac{1}{T} \sum_{t=1}^T (\rho_t^i - \langle \rho^i \rangle) (\rho_{t+\kappa}^j - \langle \rho^j \rangle) \quad (1)$$

where  $\kappa$  denotes the time-shift of one time-series with respect to the other in calculating the covariance. Here,  $\langle \rho^i \rangle$  and  $\langle \rho^j \rangle$  compute the empirical means of the  $T$ -length time series  $\rho^i$  and  $\rho^j$ , respectively. The value of  $\kappa_*(i, j)$  at which the covariance is maximized between the daily new infections in district  $i$  and that in district  $j$ , denotes the number of days by which the infection pattern in district  $j$  roughly lags the pattern in district  $i$ . We vary  $\kappa$  in  $[-30, 30]$  with  $T = 92$  days for the first wave and  $T = 69$  days for the second wave to compute  $\kappa_*(i, j)$ .

A time-lag between the patterns of district  $i$  and district  $j$  does not imply that infected people from district  $i$  came in direct contact with people in district  $j$  to drive the spread of COVID-19. However, consistent positive values of  $\kappa_*(i, j)$ 's for multiple  $j$ 's suggests that district  $i$  is an epicenter of the infection spread.

Figure 3 portrays a bubble plot of  $\sum_{j=1}^N \kappa_*(i, j)$  over those locations  $i$  for which this sum is positive. A larger bubble indicates higher likelihood of a location being a source of the

infection spread. Figure 3 reveals important differences between the likely epicenters of the two waves. The first wave developed around large cities (possibly from contact with infected people engaged in foreign travel). The second wave largely originates from Maharashtra and, to a smaller extent, the Ganges belt. Figure 4 affirms that the infection spread from higher to lower populated districts during the first wave. Populations of the districts affected through time is much “flatter” during the second wave, suggesting the role of super-spreader events in affecting the nation almost uniformly. Maharashtra as a likely epicenter for the second wave, corroborates our theory about the circulation of the Delta variant, aligned with the findings from genomic sequencing of COVID-positive patients. The remaining epicenters in the Ganges belt point towards the Kumbh Mela festival as the super-spreader event <sup>2-4</sup>.

### **Epidemiological Model to Explain India’s Infection Dynamics**

In this section, we provide additional evidence for our hypothesis that super-spreader events played a dominant role in India’s second wave. To do so, we consider an epidemiological diffusion model, estimate its parameters from the district-wise test results, and demonstrate lack of fit during the second wave. To this end, consider the susceptible-infected-removed (SIR) compartmental model, described by

$$\begin{aligned} S^i(t+1) &= S^i(t) - \beta_{\text{int}}^i(t)I^i(t)S^i(t) - \beta_{\text{ext}}^i(t)I_{\text{ext}}^i(t)S^i(t), \\ I^i(t+1) &= I^i(t) + \beta_{\text{int}}^i(t)I^i(t)S^i(t) + \beta_{\text{ext}}^i(t)I_{\text{ext}}^i(t)S^i(t) - \gamma I^i(t), \\ R^i(t+1) &= R^i(t) + \gamma I^i(t). \end{aligned} \quad (2)$$

Here,  $S^i(t)$  denotes the fraction of the population in district  $i$  that is susceptible to the infection on day  $t$ . Similarly,  $I^i(t)$  and  $R^i(t)$  are the fractions of the infected population and the removed (recovered or deceased) population in district  $i$  on day  $t$ , respectively. Again, we consider

$N = 394$  districts. Per <sup>5</sup>, we choose  $\gamma = (14 \text{ days})^{-1}$  as the combined constant rate of recovery and death from COVID-19. Parameter  $\beta_{\text{int}}^i(t)$  captures the rate at which the infected population within district  $i$  contributes to new infections within that district on day  $t$ . Similarly,  $\beta_{\text{ext}}^i(t)$  “lumps” the impact of infections outside of district  $i$  towards new infections in district  $i$  on day  $t$ . The population-weighted external infections seen from district  $i$  is computed as

$$I_{\text{ext}}^i(t) := \frac{\sum_{j:(j,i) \in \mathcal{G}} I^j(t) P^j}{\sum_{j:(j,i) \in \mathcal{G}} P^j}, \text{ where } P^j \text{ is the population of district } j, \text{ per the 2011 Census. We allow}$$

neighbors of district  $i$  according to a graph  $\mathcal{G}$  on  $N$  nodes as the only places whose infection levels directly impact district  $i$ 's infections in one day. We construct  $\mathcal{G}$  to encode the intuition that geographically distant districts do not directly contribute to each other's infection spread within one day, unless through long-distance travels such as those on airplanes. To construct  $\mathcal{G}$ , we first add an edge between districts  $i$  and  $j$ , if the distance between the centroids of these districts  $d^{i,j}$  is below 1594.5 km. We compute  $d^{i,j}$  using the great circle distance formula<sup>6</sup> applied to the latitudes and longitudes of the centroids from the Kaggle data hub<sup>7</sup>. Then, we add edges between any two districts that contain airports and have at least one direct flight between them. For the flight network, we use the origin-destination airport pairs from<sup>8</sup>. The resultant graph  $\mathcal{G}$  then has 4599 edges, much less than that in a fully-connected graph over 394 nodes.

We use two sets of data—the cumulative fraction of COVID-19 cases  $Q^i$  on  $t = 1$  in district  $i$  and the fraction of new COVID-positive cases  $\Delta^i(t)$  in district  $i$  on days  $t = 1, \dots, T$ . Specifically,  $Q^i$ 's yield

$$S^i(1) = 1 - Q^i, \quad I^i(1) = (1 - \gamma)Q^i, \quad R^i(1) = \gamma Q^i, \quad (3)$$

that are then propagated using  $\Delta$ 's via



$$\begin{aligned}
S^i(t+1) &= S^i(t) - \Delta^i(t), \\
I^i(t+1) &= I^i(t) + \Delta^i(t) - \gamma I^i(t), \\
R^i(t+1) &= R^i(t) + \gamma I^i(t).
\end{aligned} \tag{4}$$

We now describe our procedure to estimate the  $\beta$  parameters from the time-series of  $S^i, I^i, R^i$  over 05/01/2020 – 05/08/2021. Since we expect mobilities, and in turn the  $\beta$ 's, to change gradually, we assume that  $\beta_{\text{int}}^i(t), \beta_{\text{ext}}^i(t)$  remain constant over 2-week periods. Denote them as  $\beta[1], \dots, \beta[26]$ , where  $\beta[\tau]$  stands for  $\beta_{\text{int}}(t), \beta_{\text{ext}}(t)$  across all districts from  $t = 14\tau - 13$  to  $t = 14\tau - 1$  days. Then, we minimize  $\varphi$  to estimate  $\beta[1], \dots, \beta[26]$ , where

$$\begin{aligned}
\varphi(\beta[1], \dots, \beta[26]) &:= \sum_{i=1}^N \sum_{\tau=1}^{26} \sum_{t=14\tau-13}^{14\tau} \left( \Delta^i(t) - \beta_{\text{int}}^i[\tau] I^i(t) S^i(t) - \beta_{\text{ext}}^i[\tau] I_{\text{ext}}^i(t) S^i(t) \right)^2 \\
&+ \lambda \sum_{\tau=1}^{25} \sum_{i=1}^N \left[ \left( \beta_{\text{int}}^i[\tau+1] - \beta_{\text{int}}^i[\tau] \right)^2 + \left( \beta_{\text{ext}}^i[\tau+1] - \beta_{\text{ext}}^i[\tau] \right)^2 \right] + \rho \sum_{\tau=1}^{26} \sum_{i=1}^N \beta_{\text{ext}}^i[\tau]^2. \tag{5}
\end{aligned}$$

The first summand in equation (5) equals the regression error in explaining the emergence of daily new infections using  $\beta$ 's. The second and the third summands in equation (5) penalize deviations of  $\beta$ 's across consecutive 2-week periods. This penalty encodes the intuition that  $\beta$ 's should vary gradually over time. The last summand in the definition of equation (5) adds a regularization that reduces the tendency of the estimator to explain the emergence of new infections solely using external infections. For the estimation process, we use  $\lambda = \rho = 10^{-3}$ .

The quantile plot of estimated  $\beta_{\text{int}}$ 's in Figure 5(a) across the districts over time reveal the two waves of COVID-19 infections. The jump in  $\beta_{\text{ext}}$  during the second wave appears particularly pronounced, compared to that during the first wave. Despite an explicit penalty on large  $\beta_{\text{ext}}$ 's in our design objective, the estimation procedure favors  $\beta_{\text{ext}}$ 's to explain the emergence of new infections, without a similarly sharp increase in  $\beta_{\text{int}}$ 's. The quantiles of the regression errors over time in Figure 5(c) during the second wave are also much higher than those during the first wave. These observations indicate that the epidemiological model fits the

data from the first wave better than from the second wave, indicating the role of an unmodeled super-spreader event.

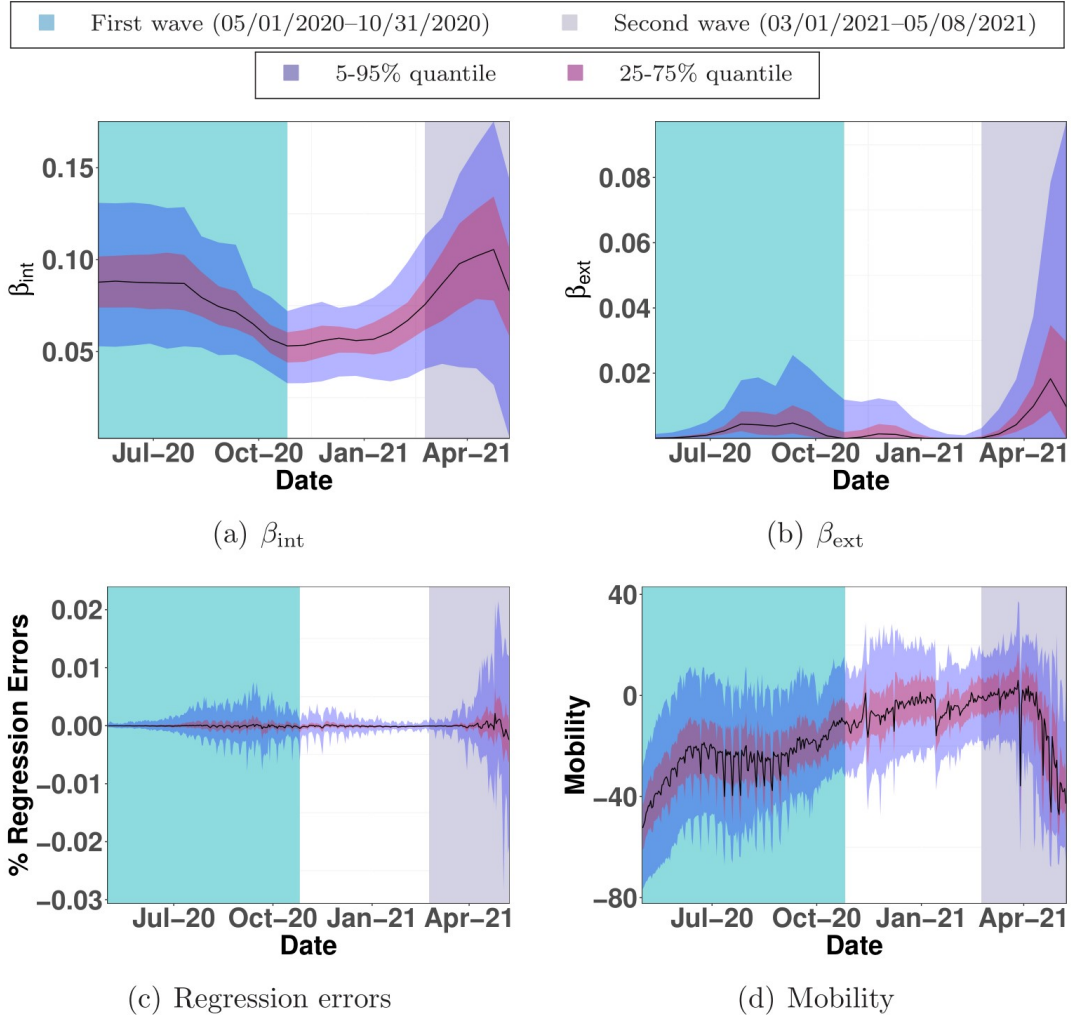


Figure 5. Plots (a) and (b) capture the quantiles of  $\beta_{int}$ 's and  $\beta_{ext}$ 's across districts over 2-week time windows from 05/01/2020 to 05/08/2021. The regression errors in explaining the emergence of daily new infections with the estimated  $\beta$ 's is given in (c). Plot (d) shows the quantiles of mobility variations across districts.

In Figure 5(d), we plot the quantiles of average mobility across districts over time. To gauge the district-wise mobility of the Indian population, we use data from the Google COVID-19 Community Mobility Reports<sup>9</sup> that catalogs variations in the population's activities such as visitations to parks, grocery and pharmacy, retail and recreation, transit stations and workplaces. Specifically, the data reports the level of these activities compared to the baselines computed as

the median values over 01/03/2020–02/06/2020, measured from anonymized cellphone data. We compute the mobility levels as the average of the above activity levels in the five reported categories for our plots. We expect  $\beta_{\text{int}}$  to correlate well with mobility of populations within districts. Lockdown measures and people's response to said measures impact mobility patterns. Notice that  $\beta_{\text{int}}$ 's during the two waves in Figure 5(a) align well with mobility in Figure 5(d) but does not surprisingly do so between the waves. During this time, India's caseload decreased while mobility increased. Some have argued in favor of inadequate asymptomatic testing and/or possible under-reporting of symptomatic cases between the waves, to explain India's curious trend in COVID-19 infections. Others have sought to explain it through herd immunity and possibly genetic factors <sup>10,11</sup>. Data from asymptomatic tests, perhaps only in a few locations, during such times are required to settle said debate <sup>12</sup>.

### **Lessons to be Learned from India's Second Wave**

Our analysis of India's second wave provides actionable policy insights to curb such an outbreak in the future. Whenever a region shows heightened growth in infections compared to neighboring regions, one must carefully consider the possibility of a new mutation. Maharashtra's anomalous infection growth in early 2021 could have served as an early warning sign, given the discovery of the Delta variant there in late 2020. Detection of easily transmissible and possibly more virulent mutations relies on adequate testing (both symptomatic and asymptomatic), followed by genomic sequencing <sup>4</sup>. India has only identified 13,000 genetic sequences (0.5/1000 cases) <sup>3,13</sup>, compared to the United States which has identified 400,000 sequences (12/1000 cases), and even the latter has been deemed sub-optimal <sup>14</sup>. Given the variation in the abilities to test and sequence across different countries, it is imperative that a global initiative for the same is undertaken to

preempt future outbreaks caused by variants. SARS-CoV-2 will likely continue to mutate. Without a global alliance to quickly detect such variants and proactively undertake risk mitigation strategies, such variants are bound to percolate to other regions (as of June 22 2021, the Delta variant has reached 80 countries, is the dominant variant in the United Kingdom and comprises ~20% of the cases in the United States <sup>15</sup>).

Especially when sequencing capabilities are limited, it is imperative to undertake stringent measures against super-spreader events, even if reported cases are largely declining. Otherwise, potent local mutations can quickly develop a large geographical footprint, overwhelming the healthcare system, causing rapid loss of life, as the Indian experience illustrates.

India's second wave started to slow down in the second half of May, thanks to the lockdown measures (see our data for a list of lockdown dates). Temporary and targeted lockdowns are unavoidable to arrest the immediate growth of infection. One can anticipate such lockdowns using the growing number of districts with  $\geq 5$  infections (as in Figure 1(b)) as an early warning signal.

On March 10 2021, 1.5% and 0.3% of India's population had received at least one dose and were fully vaccinated, respectively <sup>16</sup>, leaving most of India's population vulnerable to the highly transmissible Delta variant. Studies indicate that most COVID-19 vaccines are effective against circulating variants <sup>17,18</sup> (with the notable exception of the AstraZeneca vaccine against the Beta (B.1.351) variant <sup>19</sup>, which however, is 80% effective against the Delta variant infection and 94% effective in preventing associated hospitalization). Vaccination remains the only long-term solution to tackle future outbreaks due to mutant variants <sup>20</sup>. Myopic reliance on declining numbers of daily infections alone can be unreliable, as the second wave in India illustrates <sup>3</sup>.

## Conclusions

The second wave of COVID-19 infections has overwhelmed India's healthcare system. The data suggests the circulation of the Delta variant, aided by a super-spreader event, as the likely explanation. Under-reporting of COVID-19 cases prior to the emergence of the second wave may have magnified the size of the second wave. Moving forward, it remains crucial to identify potent variants, possibly through international collaboration for genomic sequencing, and implement planned lockdowns to arrest the immediate infection growth. Vaccination remains the only viable long-term solution.

## References

1. Kuppalli, K. *et al.* India's COVID-19 crisis: a call for international action. *Lancet* (2021).
2. Lancet, T. India's COVID-19 emergency. *Lancet (London, England)* **397**, 1683 (2021).
3. Bhuyan, A. Experts criticise India's complacency over COVID-19. *Lancet* **397**, 1611–1612 (2021).
4. Aiyar, Y. *et al.* India's resurgence of COVID-19: urgent actions needed. *Lancet* (2021).
5. U.S. Center for Disease Control and Prevention. When You Can be Around Others After You Had or Likely Had COVID-19. (2021).
6. Wolfram Mathworld. Great Circle. (2021).
7. Kaggle. Indian Census Data with Geospatial indexing. (2021).
8. Data.world. Airports, Airlines, and Routes. (2021).
9. Google COVID-19 Community Mobility Reports. No Title. (2021).
10. Chandrashekhar, V. Herd immunity? India still has a long way to go, scientists say.

- Science* (80-. ). **370**, 513 (2020).
11. Cao, Y. *et al.* Comparative genetic analysis of the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different populations. *Cell Discov.* **6**, 11 (2020).
  12. Subramanian, R., He, Q. & Pascual, M. Quantifying asymptomatic infection and transmission of COVID-19 in New York City using observed cases, serology, and testing capacity. *Proc. Natl. Acad. Sci.* **118**, (2021).
  13. Srivastava, S., Banu, S., Singh, P., Sowpati, D. T. & Mishra, R. K. SARS-CoV-2 genomics: An Indian perspective on sequencing viral variants. *J. Biosci.* **46**, 1–14 (2021).
  14. Maxmen, A. Why US coronavirus tracking can't keep up with concerning variants. *Nature* **592**, 336–337 (2021).
  15. World Health Organization. *Weekly epidemiological update on COVID-19 - 22 June 2021. Edition 45. Emergency Situational Updates*  
<https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---22-june-2021> (2021).
  16. Our World in Data. Coronavirus (COVID-19) Vaccinations.  
<https://ourworldindata.org/covid-vaccinations?country=IND>.
  17. Lopez Bernal, J. *et al.* Effectiveness of the Pfizer-BioNTech and Oxford-AstraZeneca vaccines on covid-19 related symptoms, hospital admissions, and mortality in older adults in England: test negative case-control study. *BMJ* **373**, n1088 (2021).
  18. Sheikh, A., McMenamin, J., Taylor, B. & Robertson, C. SARS-CoV-2 Delta VOC in Scotland: demographics, risk of hospital admission, and vaccine effectiveness. *Lancet* (2021) doi:10.1016/S0140-6736(21)01358-1.

19. Madhi, S. A. *et al.* Efficacy of the ChAdOx1 nCoV-19 Covid-19 Vaccine against the B.1.351 Variant. *N. Engl. J. Med.* **384**, 1885–1898 (2021).
20. World Health Organization. *COVID-19 Strategic Preparedness and Response Plan (SPRP 2021)*. <https://www.who.int/publications/i/item/WHO-WHE-2021.02> (2021).
21. Public Health England. *SARS-CoV-2 variants of concern and variants under investigation in England. Technical briefing 16*.  
[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/994839/Variants\\_of\\_Concern\\_VOC\\_Technical\\_Briefing\\_16.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/994839/Variants_of_Concern_VOC_Technical_Briefing_16.pdf) (2021).

**Acknowledgments:** We thank the following staff of the Learning Resource Centre at the Indian School of Business, Hyderabad (India) in assisting us with compiling the data:  
Balasubramanian S. P., Amrita Shah, Gurusrinivasan K., Sruthi Kambar, Sarita Bhoi, M. Jayalakshmi and Surender Chukka.

**Funding:** The work was partially supported by a grant from the C3.ai Digital Transformation Institute.

**Author contributions:**

Writing – original draft: SB.

Methodology: SB, SSo.

Funding acquisition: SB.

Investigation: SH, SSo, ACEIII.

Formal Analysis: SH, SSo.

Visualization: SH, AI.

Software: SH, SSo, AI.

Data curation: SSo, AI.

Conceptualization: UKM, SSe.

Writing – review & editing: UKM, SSe.

Validation: UKM, SSe.

Supervision: SSe.

**Competing interests:** Authors declare that they have no competing interests.

**Data and materials availability:** All datasets and the code for the analysis are available in

<https://github.com/heart-analytics/COVID19-India>