*Proceedings of the 2022 Winter Simulation Conference*
*B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C.G. Corlu, L.H. Lee, E.P. Chew, T. Roeder, and P. Lendermann, eds.*

# MODELLING THE DELTA COVID-19 WAVE IN MUMBAI

Sandeep Juneja
Daksh Mittal

School of Technology and Computer Science
Tata Institute of Fundamental Research
Dr Homi Bhabha Road Navy Nagar Colaba
Mumbai, Maharashtra 400005, India

## ABSTRACT

Using agent based simulator (ABS), we attempt to explain the infectiousness of the delta variant through scenario analysis to best match the observed fatality data in Mumbai, where the variant initially spread. Our somewhat prescient conclusion, based on analysis conducted in March-April 2021 was that the new variant was 2-2.5 times more infectious than the original Wuhan variant. We also observed then that certain performance measures such as timings of peaks and troughs were quite robust to the variations in model parameters and hence can be reliably projected even in presence of model uncertainties. Furthermore, we introduce enhancements to help model variants, vaccinations, basic and effective reproduction number in ABS. Our analysis suggests an interesting observation - although slums have around half of Mumbai population and are much more dense and have higher prevalence, the effective reproduction number between slums and non-slums equalises early on and largely moves together thereafter.

## 1 INTRODUCTION

India was hit by a huge delta driven second wave of Covid-19 that started in February 2021. The wave first arose in the state of Maharashtra, and within it, Mumbai was amongst the first cities to see the increase, primarily because it is a crowded city that substantially relaxed its Covid restrictions on February 1.

Early on the epidemiologists and the modellers thought that the increase in the cases was due to the economy opening up. However, it soon became clear that this opening up may not explain the dramatic increase in cases and other health statistics. In this paper, we use and enhance the IISc-TIFR city simulator (Agrawal et al. 2020) to computationally understand the drivers of the delta wave. This study was largely conducted in March and April 2021 (Juneja and Mittal 2021). Given the inherent uncertainty in the role of the variants, and other important infection spread determining factors, we used our simulator to conduct an extensive scenario analysis - we played out many plausible scenarios through varying economic activity, reinfection levels, population compliance, infectiveness, prevalence and lethality of the possible variant strains, and infection spread via local trains (a key interaction space in Mumbai - over 8 million people use these crowded trains daily in normal times) to arrive at those that better explained the delta wave fatality numbers (scenario having least root mean square error (rmse) and mean absolute error (mae)). We focus on observed fatalities as opposed to observed cases because observed cases are much less predictable - they rely on administration's testing policy that can vary with time and location, and some segments of population (e.g., wealthy neighbourhoods) are much more likely to get tested as compared to some others (e.g., people living in slums).

A key feature of the delta wave in Mumbai was that the observed fatalities were low in February and March and saw a *phase change* or a steep increase in the growth rate after around late March (Figure 1). We conducted extensive experiments to replicate this observed sharp *convexity*. This was not an easy

phenomena to replicate, and we found that explanations such as increased laxity in the population, increased reinfections, increased intensity of infections in Mumbai transportation, increased lethality in the virus, or a combination of these, did a poor job of matching the observed pattern. Our agent-based city simulator (ABCS) has the granularity to model each individual in an estimated 12.8 million population of Mumbai. This means that each run of each scenario takes around 8 to 10 hours to run on a high performance computing system (link), making it computationally prohibitive to implement more sophisticated stochastic gradient methods. Our analysis thus is based running many plausible scenarios and identifying those that appear to match the data well.

We found that the most likely explanation was presence of small amount of highly infectious variant on February 1 that grew rapidly thereafter and became a dominant strain by Mid-March. The scenario where the variant was 2.5 times more infectious than the original Covid strain, and accounted for 1.5% to 2.5% of the infected population on Feb 1, appeared to match the data well. This variant increased to over 60% by mid March. Another scenario where the variant was 2 times more infectious and the compliance was less than the earlier scenario, also explained the data reasonably well. Subsequent virus studies appear to have validated these numerically driven conclusions (Public Health England 2021), (Katella, Kathy 2022). While realistically there might have been multiple variants operating in Mumbai, in our scenario analysis, to maintain simplicity, we assumed that there was a single variant.

In addition, we observed and highlighted that timings of peaks and troughs of delta wave occurred roughly at same time across all the scenarios considered (see, e.g., Figures 3, 5, 7, 11). This robustness suggested that they were likely to be accurate and can be used to make projections. This was borne out by observations later. Specifically:

**1.** We projected that the Mumbai delta wave of fatalities would likely peak around the first week of May (First reported by us on March 31 (Juneja, S. 2021b)). Again, this appeared largely invariant under many plausible scenarios. Since cases typically lead fatalities by two to three weeks, this in turn suggested that the cases would peak a few weeks before first of May, as indeed turned out to be the case.

**2.** In the scenarios that loosely matched the Mumbai's vaccination drive (age based vaccinations of the order of 15 to 20 lac new people a month; 75% efficacy), we projected that the fatalities reduce to Jan. and Feb. levels by June 1. This again suggested that the reported cases will substantially decrease in Mumbai by early to mid May (reported by us on April 15 (Juneja, S. 2021a)).

The remaining paper is structured as follows: We first summarise our ABCS briefly in Section 2 and the enhancements to the model in Section 3. We then analyse the delta driven Covid wave in Mumbai by playing out various scenarios in Section 4. In Section 5, we describe the methodology to calculate the basic reproduction number ($R_0$) and the effective reproduction number ($R_t$) in agent based simulators for epidemics. We make an interesting empirical observation that the $R_t$ between slums and non-slums equalises early on and thereafter largely moves together.

## 2 AGENT-BASED CITY SIMULATOR

We first informally spell out the dynamics of our infection spread model. A more detailed discussion can be seen in (Agrawal et al. 2020) and (Juneja and Mittal 2021). The model consists of individuals and various interaction spaces such as households, schools, workplaces, community spaces and transport (local trains in Mumbai). Further smaller interaction spaces (sub networks) within each workplace, school, and community, are considered with the interpretation that people associated to same sub network have high contact rate amongst them compared with others outside the sub network. In some more detail, we create 'project' networks at each workplace, a 'class' network in each school consisting of students of the same age, a random community network among people in a given ward to model interactions amongst friends and relatives, and a neighbourhood sub network among people living in close vicinity. Infected individuals interact with susceptible individuals in these interaction spaces. The number of individuals living in a household, their age, whether they go to school or work or neither, schools and workplaces size and composition all have distributions that may be set to match the available data. The model proceeds

in discrete time steps of constant width $\Delta t$ (six hours in our set-up). At a well chosen time zero, a small number of individuals can be set to either exposed, asymptomatic, or symptomatic states, to seed the infection. At each time $t$, an infection rate $\lambda_n(t)$ is computed for each susceptible individual $n$ based on its interactions with other infected individuals in different interaction spaces (households, schools, workplaces and community). In the next $\Delta t$ time, each susceptible individual moves to the exposed state with probability $1-\exp\{-\lambda_n(t)\cdot\Delta t\}$, independently of all other events. Further, disease may progress independently in the interval $\Delta t$ for the population already afflicted by the virus. The computation of $\lambda_n(t)$ and the probabilistic dynamics of disease progression are briefly summarized below (details available in (Agrawal et al. 2020), also see (Ferguson et al. 2020)). Simulation time is then incremented to $t+\Delta t$, and the state of each individual is updated to reflect the new exposures, changes to infectiousness, hospitalisations, recoveries, quarantines, etc., during the period $t$ to $t+\Delta t$. The overall process repeats incrementally until the end of the simulation time. The simulation dynamics are illustrated in Figure 2.

**Computing $\lambda_n(t)$ :** A susceptible individual $n$ at any time $t$ receives a total infection rate $\lambda_n(t)$ which is sum of the infection rates $\lambda_n^{home}(t)$, $\lambda_n^{school}(t)$, $\lambda_n^{class}(t)$, $\lambda_n^{workplace}(t)$, $\lambda_n^{project}(t)$, $\lambda_n^{transport}(t)$, $\lambda_n^{community}(t)$, $\lambda_n^{neighborhood}(t)$, $\lambda_n^{random-community}(t)$ coming in from infected individuals in respective interaction spaces of individual $n$. Briefly: The transmission rate ($\beta$) of virus by an infected individual in each interaction space is the expected number of eventful (infection spreading) contact opportunities with all the individuals in that interaction space. It accounts for the combined effect of frequency of meetings and the probability of infection spread during each meeting. An infected individual can transmit the virus in the infective (pre-symptomatic or asymptomatic stage) or in the symptomatic stage. Each individual has two other parameters: a severity variable (individual attendance in school and workplace depends on the severity of disease) and a relative infectiousness variable, virus transmission is related linearly to this. Both bring in heterogeneity to the model. $\lambda_n^{home}(t)$ is the sum of the average transmission rate coming in to individual $n$ from each infected individual in his home. $\lambda_n^{school}(t)$, $\lambda_n^{class}(t)$, $\lambda_n^{workplace}(t)$ and $\lambda_n^{project}(t)$ are also computed similarly based on infected individuals in respective interaction spaces of the individual $n$. $\lambda_n^{transport}(t)$ for an individual is the product of travel transmission rate ($\beta_T$), fraction of population travelling at time $t$ and the ratio of distance travelled by infected individuals to the total distance travelled by all the people at time $t$. Different wards in the city constitute different communities. Community infection rate of each community is sum of transmission rate from all the infected individuals of the community weighted inversely based on the distance between the individual and community centre. $\lambda_n^{community}(t)$ seen by an individual $n$ in community $c$ is sum of the community infection rates from different communities of the city weighted inversely based on the distance between the two communities and distance between individual $n$ and its community center. An age dependent factor determining the mobility of the individuals in the community is also accounted in $\lambda_n^{community}(t)$. $\lambda_n^{random-community}(t)$ and $\lambda_n^{neighborhood}(t)$ are also similarly calculated.

**Disease progression :** We use a model of COVID-19 progression based on descriptions in (Verity et al. 2020) and (Ferguson et al. 2020). An individual may have one of the following states: susceptible, exposed, infective (presymptomatic or asymptomatic), recovered, symptomatic, hospitalised, critical, or deceased. Briefly: An individual after getting exposed to the virus at some time observes an incubation period which is random with the Gamma distribution of shape 2 and scale 2.29; the mean incubation period is then 4.58 days. Individuals are infectious for an exponentially distributed period of mean duration 0.5 of a day. This covers both presymptomatic transmission and possible asymptomatic transmission. We assume that three fifth of the patients recover, these are the asymptomatic patients; the remaining develop symptoms. Individuals either recover or move to the hospital after a random duration that is exponentially distributed with a mean of 5 days. The probability that an individual recovers depends on the individual's age. While hospitalised individuals may continue to be infectious, they are assumed to be sufficiently isolated, and hence do not further contribute to the spread of the infection. Further progression of hospitalised individuals to critical care and further to fatality is also age dependent.

**Public health safety measures:** We introduce methodologies to model different PHSMs such as lockdowns, home quarantine, case isolation, social distancing of elderly population, containment zones, contact tracing

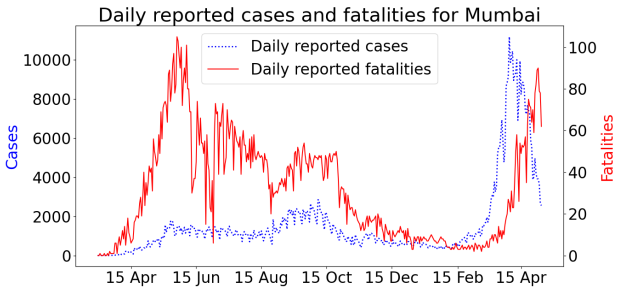


Figure 1: Daily reported cases and fatalities in Mumbai (Source : BMC Dashboard).

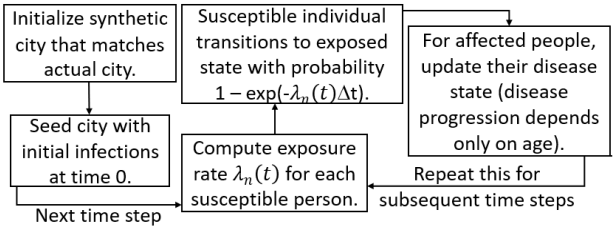


Figure 2: Simulation dynamics of the Agent Based Simulator.

Table 1: Basic Interventions as implemented in the simulator.

| Intervention | Description |
|---|---|
| No intervention | Business as usual |
| Lockdown | For compliant households, household rates are doubled, no workplace interactions except for 25% leakage (for essential services), community interactions reduce by 75%. For non-compliant households, workplace interactions only have a leakage of 25%, community interactions are unchanged, and household interactions increase by 50% |
| Case Isolation | Compliant symptomatic individuals stay at home for 7 days, reducing non-household contacts by 75%. Household contacts remain unchanged. |
| Home Quarantine | Following identification of a symptomatic case in a compliant household, all household members remain at home for 14 days. Household contact rates double, contacts in the community reduce by 75% |

and testing, mobility restrictions, masks etc. See (Agrawal et al. 2020) and (Harsha et al. 2020) for details on how these interventions are modelled in the simulator. Table 1 summarizes some of the key interventions implemented early on. The PHSM's mentioned above when implemented put some restrictions on the individual's mobility. However, it is often the case that when several restrictions are in place, only a fraction of the population comply with these restrictions. Therefore, we restrict the movement of only the individuals in the compliant fraction of households in the city.

**Modelling Mumbai:** A synthetic model of about 1.28 crore (12.8 million) residents of Mumbai that matches the city population ward-wise (24 wards, each further divided into slums and non-slums), and matches the numbers employed, numbers in schools, commute distances, etc is created. Mumbai slums are densely crowded and this leads to difficulties in maintaining social distancing and increased transmission between the infected and the uninfected. We account for this by selecting higher community transmission rates in the slums - 3 times the non slums. This factor ensures that model prevalence in July 2020 matches the high prevalence of 56% observed in Mumbai slums and lower prevalence of 16% in non slums in July 2020 (Malani et al. 2020). The contact rates are calibrated to match the observed growth of fatalities in the city till April 10, 2020, and to have roughly equal contribution of infections from the household, community and workplace networks (including the sub networks) in the "no-intervention" scenario as mentioned in (Agrawal et al. 2020). Compliance rates are again selected to match the observed fatality data early on (in April-June 2020) and later it is varied based on Google Mobility Report. We refer the reader to (Harsha et al. 2020) for the parameter values.

## 3 SOME MODEL ENHANCEMENTS

**Modelling multiple strains :** Recall that in the current methodology (Ferguson et al. 2020), a susceptible individual $n$ sees an incoming infection rate of $\lambda_n(t)$ from all the infected individuals in the city at time $t$, based on which it gets exposed with probability $1-\exp\{-\lambda_n(t)\cdot\Delta t\}$. However, with introduction of a new infectious strain, it is important to identify whether the person was exposed from an infected person with new strain or an infected person with original strain. To estimate this, we compute infection rates coming in from individuals infected with original strain ($\lambda_n^{original}(t)$) and those with new infectious strain ($\lambda_n^{infectious}(t)$) separately. To account for the increased infectiousness of the strain (let's say $k$ times more infectious then the original strain), if an individual is infected with the infectious strain, its infectiousness is increased by the factor $k$. Then, $\lambda_n(t)$ can be written as

$$\lambda_n(t) = \lambda_n^{original}(t) + \lambda_n^{infectious}(t)$$

In our model, initially a fixed percentage (2.5%) of individuals are chosen randomly from infected population (active) are assumed to be from a more infectious strain on a fixed day (February 1, 2021). After this initial seeding of infectious strain, it spreads in the following manner: every individual $n$ who gets exposed to the disease at time $t$ is deemed to be infected with the infectious strain with probability $\frac{\lambda_n^{infectious}(t)}{\lambda_n(t)}$ and with the original strain with probability $1-\frac{\lambda_n^{infectious}(t)}{\lambda_n(t)}$. The above methodology extends easily to more than two strains.

**Modelling vaccinations:** In our model, we assume that once a vaccine is administered to an individual, he/she instantly becomes immune from Covid if the vaccine is effective for that individual and he/she is susceptible. In reality, the vaccine starts to provide immunity around 2-4 weeks later. To account for this, we shift forward the vaccine administration date in the simulator by almost a month from the actual observed data. Immunity is modelled in the simulator by transferring the individual from susceptible state to the recovered state. Further, to capture vaccine's efficacy of around 75%, we assume that 75% of the people vaccinated have complete protection from Covid, while randomly chosen 25% of the vaccinated have no protection.

## 4 ANALYZING THE DELTA WAVE THROUGH SCENARIO ANALYSIS

In this section, we conduct an extensive scenario analysis using our ABS and identify the ones that better explain the Mumbai fatality data from February to April 2021, with its peculiar late escalation. To analyse the delta wave, we conduct extensive scenario analysis and came up with a scenario which appears to match the fatalities time series well. We call this scenario as the **'base case'**. We then present other scenarios as perturbations to the base case. We do not compare with the aggregate differential equations based model (e.g., the well known SIR, SEIR models) because it is not clear how to capture new variants, reinfections, mobility restrictions, partial lockdown and vaccinations accurately in those models.

### 4.1 Specifying the 'base case' that matches the observed fatality data well

**Mobility:** Mobility in our model is modelled through workplace attendance fraction. To determine workplace attendance schedule we relied on the Google Mobility Report and government orders. The complete details of workplace attendance schedule can be seen from Juneja and Mittal (2021), pg 13.

**Compliance:** Compliance levels are set at 60% in Non slums (NS) and 40% in slums (S) before December 2020 and outside festivals (during festival periods compliance is 40% NS and 20% S). These change to (50%, 30%) in Dec 2020, (40%, 20%) in Jan 2021, (20%, 10%) in 1-18 February 2021 and (40%, 20%) from Feb 19 to April 14, 2021. During the lockdown 15 April - 15 May 2021, it is (60%, 40%).

**Reinfection level:** While some cases of reinfection were reported in the literature, the phenomena appeared to be limited till April 2021. Further, if the reinfections are mostly mild, they have little affect on the

fatality numbers. We kept the reinfection level to zero in the base case. Other levels did not improve the fit to data. Later, we show results for the extensive perturbations of reinfection levels to assess their impact.
**Variants:** We assume that there existed a single variant that accounted for 2.5% of all the infected population on Feb. 1 in our model. These were randomly chosen amongst all the infected on Feb 1. Further we assumed that this variant was 2.5 times more infectious compared to the original strain.
**Vaccination:** The detailed vaccination schedule that we implemented in our model was based on the actual vaccination doses as administered in Mumbai (National Health Authority 2021). The complete details of vaccination schedule can be seen from Juneja and Mittal (2021), pg 12.
**Impact of trains:** The train transmission rate ($\beta_T$) is kept same as in the earlier report (Harsha et al. 2020) that best explains the data ($\beta_T = 40\%$ of home contact rate) .
**Variant virulence:** In the base case, we kept the virulence at the level of original variant.

See Figures 3 and 4 for the results associated with the base case. Recall that the base scenario included both the benefits of the vaccination effort and the ongoing lockdown (15 April - 15 May, 2021). In the graphs, the base scenario is compared with the following 4 scenarios : Scenario without the infectious strain and without the lockdown (Violet curve). This is the contra-factual setting where the mild wave is caused in Mumbai only due to the opening of the economy on February 1, and increased laxity in population before that. Scenario without vaccination and without lockdown (Orange curve). This helps us estimate the combined benefits of the lockdown and the vaccination effort. Scenario without lockdown and with vaccination (Red Curve). This helps isolate the benefits of the lockdown. Scenario with lockdown and without vaccination (Green curve). This helps isolate the benefits of the vaccination effort.
**Some observations suggested from the base case:** Figure 3 suggests that the lockdown substantially speeds up the return to normalcy. Under the blue curve one sees that daily fatalities return to Jan-Feb levels by June 1 due to the lockdown and vaccinations. The red curve corresponds to no lockdown and vaccinations and in this scenario the normalcy is delayed by a month. Importantly, vaccines make this normalcy lasting. Else, as the green curve shows, the fatality numbers would have started to increase from mid July onwards. Figure 4 illustrates the speed with which a highly infectious variant can come to dominate once the economy opened in the base scenario (blue curve in Figure 3).

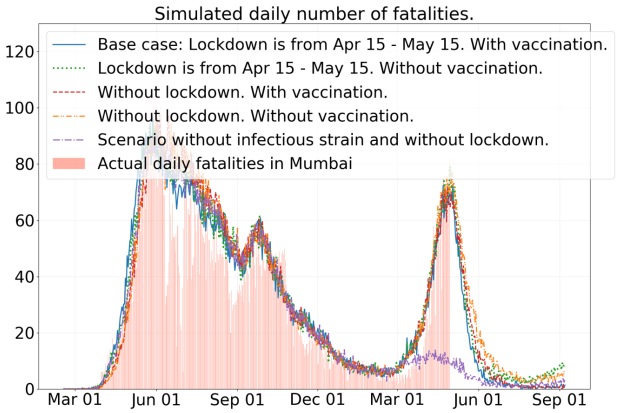


Figure 3: Base Case - 2.5% of infected people infected with infectious strain on Feb 1 (2.5 times more infectious), with lockdown from Apr 15 - May 15 and with vaccination.

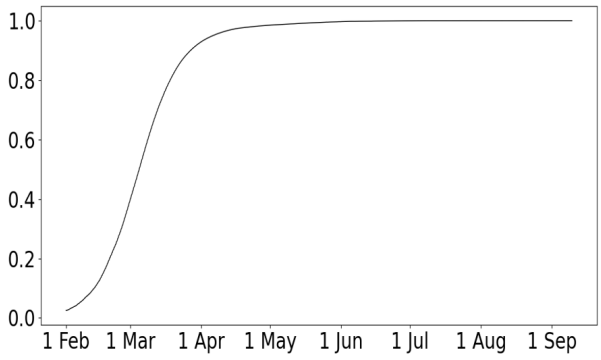


Figure 4: Fraction amongst the infected with the highly infectious strain in the base case. Scenario with 2.5% of infected people infected with infectious strain on Feb 1 (2.5 times more infectious).

### 4.2 Perturbed scenarios around the base case

The scenarios that we consider are summarized in Table 2. For comparing the simulated fatalities in different scenarios to the observed fatality data in Mumbai during delta wave through period 15 March -

30 April, 2021 (47 days), we considered following two types of errors: mean absolute error (mae) and root mean square error (rmse) (See Equations 1 and 2). This period (15 March - 30 April, 2021) was selected as this is when most of fatalities were reported. The results are summarised in Table 3. From Table 3 it can be seen that the low mae and mrse are observed in (i) Base case, (ii) Scenario with lower compliance and 2x infectiousness and (iii) Scenario with 1.5% fraction of people infected with new infectious strain on Feb 1.

$$mae = \frac{1}{47}\sum_{x=1}^{47}|\text{Simulated fatality on day x - Observed fatality on day x}| \tag{1}$$

$$rmse = \Big(\frac{1}{47}\sum_{x=1}^{47}(\text{Simulated fatality on day x - Observed fatality on day x})^2\Big)^{\frac{1}{2}} \tag{2}$$

Table 2: Different Scenarios considered around the base case.

| **Scenario tested** | **Cases considered** |
|---|---|
| Economy opening on Feb 1 | 65%, 75% |
| Compliance | weaker compliance (2.5x & 2x infectious strain) |
| Reinfection | 0%, 5%, 10%, 2.5% each month (Feb-Jul) |
| Reinfection without infectious strain | 20%, 5% each month (Feb-Jul) |
| Initial fract. of infected people with new strain | 1.5%, 2.5%, 5%, 10% |
| Infectiousness of new strain | 1.5x, 2x, 2.5x, 3x |
| Trains scale | 0.20, 0.40, 0.80 |
| Vaccine effectiveness | 75%, 55%, No vaccination |
| Virulence of New Strain | Same as original strain, 1.3 times the original strain |

Workplace opening schedules and compliance levels are same as the base case in all the figures hereafter unless specified.

**1. Economy opening up:** Figure 5 shows the fatality curve when the city opened up to level 75% on February 1 (red curve) instead of the base case (blue curve) where it opened at a level 65%. This leads to an increase in fatality levels from around March 1 onwards and somewhat earlier peak time for the fatalities. This illustrates the broader phenomena that increasing or reducing economic activity would lead to a shift in the fatality curve that does not help align the base curve with the steepness of the observed curve. A variable shift, where we first reduce the economic activity and later increase it may help in a better curve fit, but that does not match with our experience of the city situation.

**2. Compliance (or laxity):** In Figure 6, we show a lower compliance scenario (red curve) and compare it to the base scenario (blue curve). Again, the red curve leads to an increase from March 1 onwards in the fatality curve. Its not clear how to reduce or increase compliance in a realistic manner that would achieve the steepness of the observed fatality data.

In Figure 7 we consider a more promising scenario where the compliance is lowered as before, however the variant virus infectiousness is reduced to 2 times from 2.5 times. This matches the data equally well except at the peak. Further lowering compliance levels however may be an unrealistic match to reality. Nonetheless, this example illustrates that infectiousness of the new variant of order 2 is also consistent with the data. Matching any curve too closely leads to over fitting and is not desirable. Further, peak fatality data around mid to end April may be high due to avoidable fatalities as Mumbai medical infrastructure was quite stretched around mid-April, so we do not look for a close match there.

**3. Reinfection:** In Figure 8 we consider the cases where reinfection is 5% (green curve) on Feb. 1. Technically, this means that we convert randomly chosen 5% of the recovered population on Feb 1, and treat it thereafter as susceptible. Red curve captures the case where the reinfection is set to 10% on Feb 1. As Figure 8 shows, increase in reinfection as specified simply leads to higher fatality numbers that

Table 3: mean absolute error and root mean squared error for different scenarios considered as compared to the actual observed fatalities. (*Scenario does not match the fatality data in 2020).

| Scenario | mae | rmse |
|---|---|---|
| Base case | 8.3 | 10.1 |
| Economy opening 75% | 22.2 | 24.5 |
| Weaker comp. (2.5x infe.) | 26.2 | 29.2 |
| Weaker comp. (2x infe.) | 9.8 | 11.7 |
| Reinfection 5% | 32.7 | 35.9 |
| Reinfection 10% | 52.2 | 58.0 |
| Reinfection 2.5% (Feb-Jul) | 27.3 | 30.3 |
| Reinf. (no infe.) 20% | 35.4 | 39.6 |
| Reinf. (no infe.) 5% (Feb-Jul) | 15.5 | 19.9 |
| Virulence of New Strain (1.3x) | 17.4 | 19.6 |

| Scenario | mae | rmse |
|---|---|---|
| Fract. of people (new strain) 1.5% | 7.7 | 9.6 |
| Fract. of people (new strain) 5% | 18.7 | 21.3 |
| Fract. of people (new strain) 10% | 28.9 | 32.1 |
| Infe. of new strain 1.5x | 23.5 | 32.0 |
| Infe. of new strain 2x | 17.1 | 23.3 |
| Infe. of new strain 3x | 30.8 | 34.9 |
| Trains scale 0.20 | 17.9 | 19.7 |
| Trains scale 0.80 | NA* | NA* |
| Vaccine effectiveness 55% | 10.7 | 12.8 |
| No vaccination | 9.3 | 11.2 |

more-or-less increase linearly from March 1 with a high slope. In the orange curve we consider a case where reinfections are introduced gradually at 2.5% each month from Feb 1 to July 1. The curve again increases very steeply. While these are some ad-hoc numbers, its clear that reinfections would need to increase in a very specific manner and combine with appropriate compliance and variant evolution to result in a curve that matches the observed fatality curve.

In Figure 9 we assume that the variant does not exist and play only with the reinfection scenarios. The green curve corresponds to 20% reinfection amount on Feb. 1. The red curve considers the case where reinfections are introduced gradually at 5% each month from Feb 1 to July 1. These curves belabour the point that the fatality data can be explained using only reinfections in a very specific manner, involving large numbers closer to late March. A pattern of that sort does not appear to suggest itself from the anecdotal reports of reinfected cases, and appears unlikely.

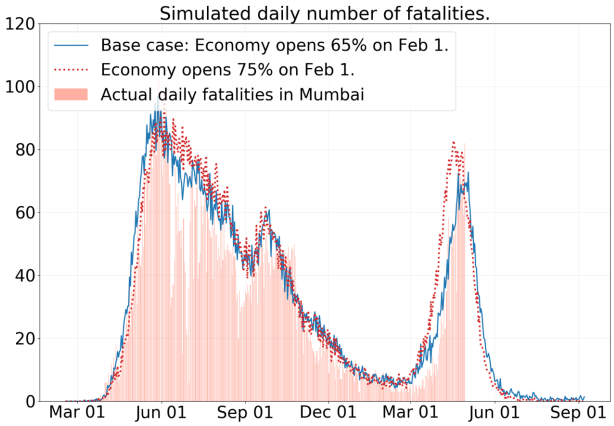


Figure 5: Scenario: Opening up of the city.

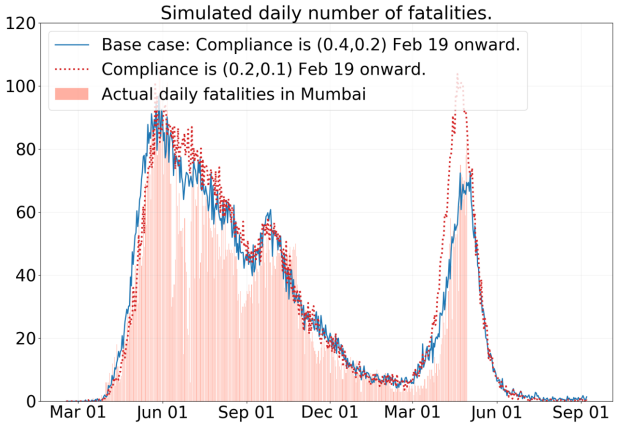


Figure 6: Scenario: Variable compliance.

**4. Fraction of Infected people infected with the new strain on Feb 1 :** In Figure 10, the red curve corresponds to the case where the new strain is assigned 10% of the infected on Feb 1. The orange curve corresponds to 5% , the green curve corresponds to 1.5%, and blue curve with 2.5% denotes the base case. These curves suggest that values close to but smaller than 2.5% on Feb. 1 perhaps may match the observed data a little better than the base case. However, our broad conclusions are unchanged by this.

**5. Infectiousness of new strain :** In the base scenario we considered that the new strain was 2.5 times more infectious than the original strain. In Figure 11, we compare this base scenario with the scenarios where new strain is 1.5 times (green curve), 2 times (orange curve) and 3 times (red curve) more infectious

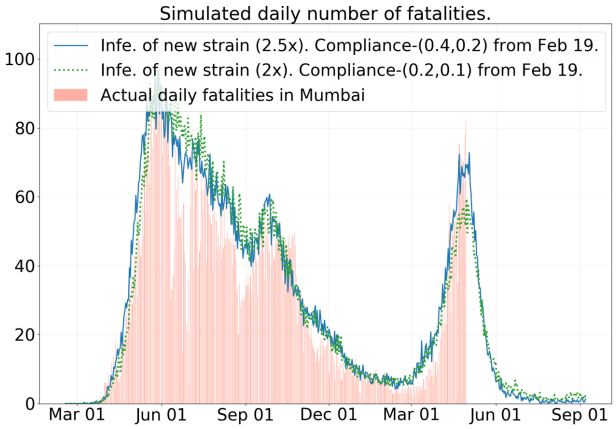


Figure 7: Scenario: Lower compliance and (2x) infectious strain.

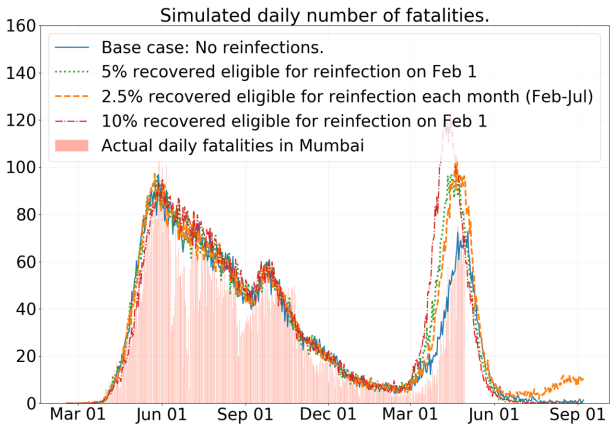


Figure 8: Scenario: Reinfections - fraction of recovered people eligible for reinfection.

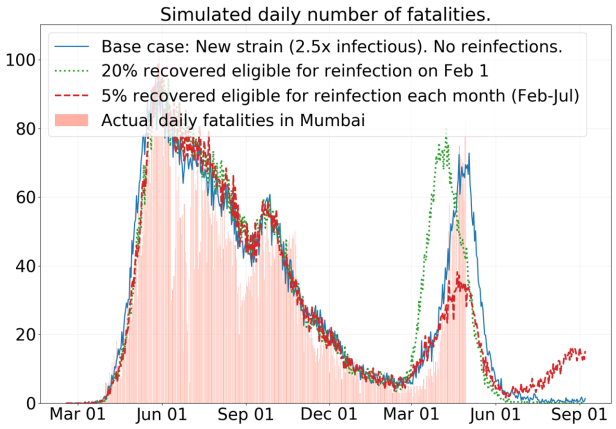


Figure 9: Scenario: Reinfections. The more infectious strain is absent.

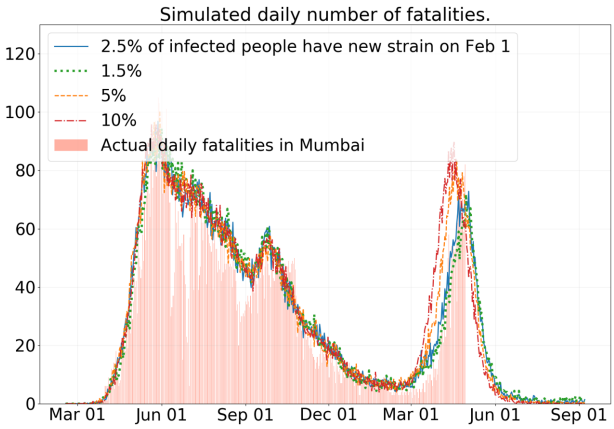


Figure 10: Scenario: Fraction of infected people infected with new strain on Feb 1.

than the original strain. The curves reaffirm that 2.5 times infectiousness (blue curve) is a good fit compared to the neighbouring values.

**6. Trains:** To capture the varying levels in infection spread through trains, in Figure 12, we consider the scenarios of low ($\beta_T$ is half the base case - green curve) and high ($\beta_T$ is twice the base case - red curve). Both the curves provide a similar fit compared to the base case. The high infection rate red curve matches the steepness of observed fatalities better, however, it results in much higher fatalities in July and August 2020 than were actually observed. These curves generally support the idea that an infectious variant grew quickly in February and March, and that trains may have played a role in its growth pattern.

**7. Vaccination:** To capture the scenario that vaccines may have lower effectiveness, as well to indicate the effect of the vaccination drive not being as extensive as in the base case, in Figure 13, we consider the case where vaccine efficacy is 55% (green curve) and where it is zero (red curve). Figure 13 suggests that even a moderately successful vaccination drive will help keep the fatalities low in August and September. Indeed, the fatalities were observed to be low from August to December before the Omicron wave hit Mumbai.

**8. Variant virulence:** Figure 14 shows the scenario where the new strain virulence or fatality rate is set to 1.3 times that of the original strain (red curve). This appears to overshoot the observed fatality data around the peak values in late April (when the effect of the new strain is more pronounced as it becomes dominant around late March).

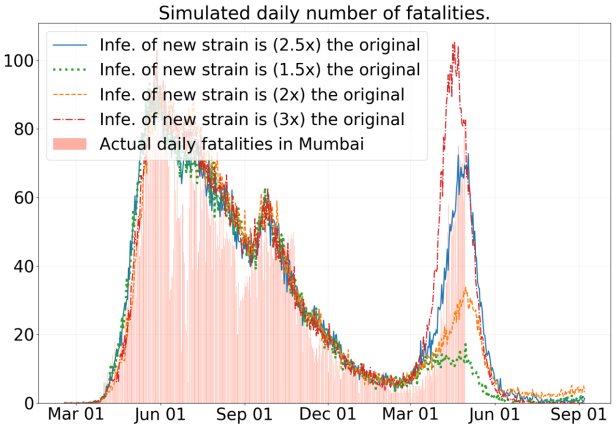


Figure 11: Scenario: Infectiousness of new strain.

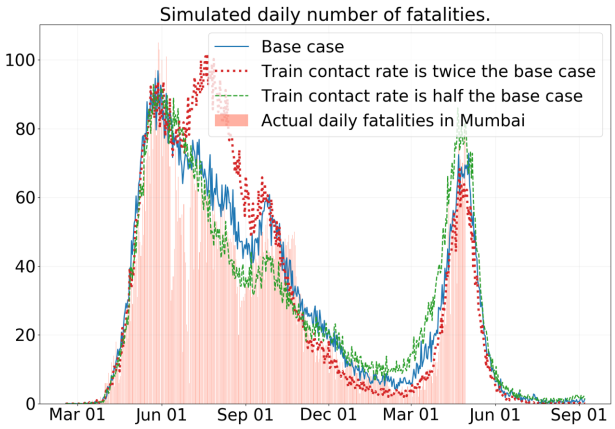


Figure 12: Scenario: Train infectivity.

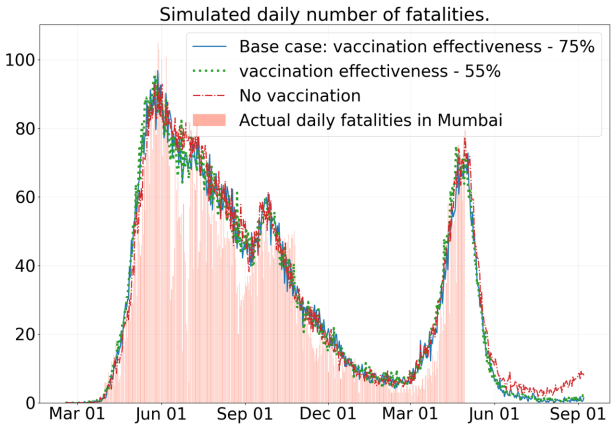


Figure 13: Scenario: Vaccination effectiveness.

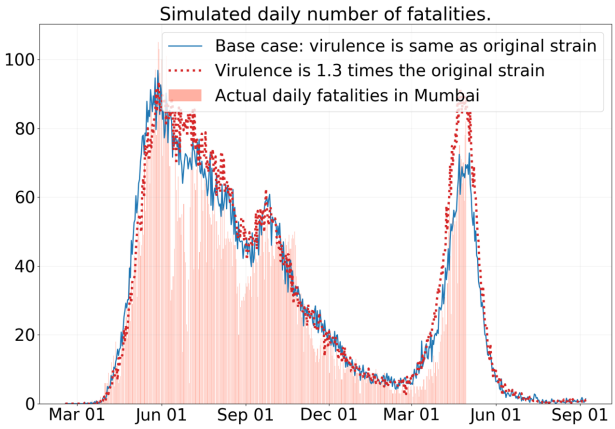


Figure 14: Scenario: Variant virulence.

Technically, due to lack of medical data, the increased virulence was implemented in our model by increasing the probability of an individual transitioning from symptomatic state to hospitalised state, from hospitalised to critical state, from critical state to fatality, each by a factor of cube root of 1.3.

## 5 THE BASIC ($R_0$) AND EFFECTIVE REPRODUCTION NUMBER ($R_t$)

**Modelling $R_0$ :** Recall that $R_0$ denotes the expected number of individuals a single randomly selected exposed person infects in a city where everyone else is susceptible. In Table 4, we report the $R_0$ for our model when it is fitted to the fatality data from March-April 2020. See (Agrawal et al. 2020) for details of how our model was fit to data. We also report the $R_0$ from a variant that is two times or two and half times more infectious than the original strain as per our model. The $R_0$ depends on whether the individual is in a less dense non-slum area, or in a high dense slum area. The $R_0$ for overall city picks a person at random over the total population of the city. $R_0$ in our model is estimated in the following manner:

A single individual is seeded (exposed to the disease) in the city on Day 0, while all the other people in the city are susceptible. The seeded individual transitions from one disease state to another and infect other susceptible individuals when he is in infective or symptomatic state. We count the total number of individuals infected directly by the seeded individual until he recovers or is dead. Any susceptible individual $n$ in the city sees the incoming infection rate $\lambda_n(t)$ at time $t$. Let $\lambda_n^{seeded}(t)$ be the contribution of the seeded individual to the infection rate that the individual $n$ sees. The susceptible individual $n$ gets exposed with probability (1-exp(-$\lambda_n(t)\Delta t$)) at time $t$.

Table 4: $R_0$ values for non slum areas, overall city and slum areas.

| Relative infe. w.r.t. base case | $R_0$ for non slums | $R_0$ for overall city | $R_0$ (inferred) for slums |
|---|---|---|---|
| 1 | $2.17 \pm 0.108$ | $4.02 \pm 0.27$ | 5.66 |
| 2 | $3.135 \pm 0.15$ | $5.56 \pm 0.28$ | 7.71 |
| 2.5 | $3.55 \pm 0.17$ | $6.56 \pm 0.31$ | 9.24 |

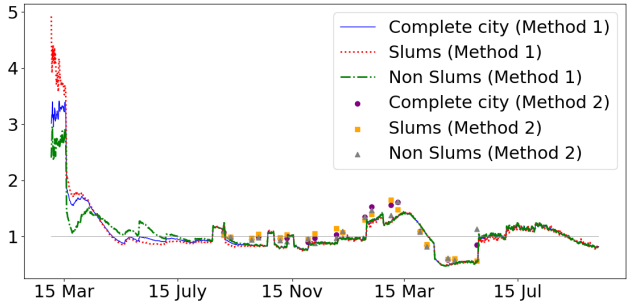


Figure 15: $R_t$ for complete city, slums and non-slums. Method 1: ratio of avg. no. exposed b/w time $t$ & $t+1$ divided by no. who stop being infective. Method 2: avg. no. of infections caused by a person exposed at time $t$.

*How to decide whether an Individual is infected from seeded individual?* Every Individual who gets exposed, is assigned to be infected from the seeded individual with probability $\frac{\lambda_n^{seeded}(t)}{\lambda_n(t)}$. This way we sum over all the people infected by the seeded individual and average it over multiple runs to arrive at $R_0$.

**Modelling $R_t$ :** Recall that $R_t$ is designed to measure the growth rate of the disease at time $t$. $R_t > 1$ signifies increasing infection in the system and $R_t < 1$ signifies declining infection. We define $R_t$ in two different ways. In the first method, it equals the ratio of the average number exposed between time $t$ and $t+1$ divided by the average number of people who stop being infective (in our model this equals the sum of the people who recover from the infectious symptomatic state and those who stop being infectious to others because they are transferred to hospitals) in this time period. Interestingly, as can be seen in Figure 15, this number becomes roughly equal for slums and non-slums in Mumbai in August 2020, and more or less stays equal thereafter. The rationale for this may be that the interactions between the two populations in communities as well as in local trains, tends to move the two groups to equalize the differences.

Another way to define $R_t$ is to track the average of the number of infections caused by an individual exposed at time $t$. These are measured at a few times and plotted as dots, squares and triangles in Figure 15. Interestingly, this definition gives very similar numbers to the earlier one, and again these numbers are quite close for both slums and non-slums. Observe that this is true even though prevalence between slums and non-slums is substantially different. The fact that $R_t$ equalized across different population is interesting and suggests this can be theoretically established. From practical viewpoint, one advantage is one can expect the local $R_t$ to well represent $R_t$ for the overall city, after the disease has become sufficiently prevalent.

## 6 CONCLUSION

The main point of the paper was to highlight that computational techniques may also have a role to play in estimating infectiousness of the variants. In particular, we tried a variety of scenarios and identified those that gave a good fit to observed fatality profile in Mumbai during the delta wave in the city. Our observation, made in late April 2021, that the new variant may be 2-2.5 times more infectious than the original one, is ball park the accepted number amongst researchers, and we were amongst the first to make

this observation. We further observed in the paper that while there are many uncertainties in modelling covid, e.g., prevalence amongst population, infection rates of the variants present, reinfection rates, etc., certain performance measures including timing of peaks and the subsequent trough in Mumbai appeared robust to reasonable variations in the underlying parameters. We further made the surprising observation that $R_t$ becomes more or less equal among slums and non slums after the disease has become sufficiently prevalent although the two populations are very different.

## ACKNOWLEDGMENTS

We acknowledge the support of A.T.E. Chandra Foundation. We also acknowledge the support of the Department of Atomic Energy, India to TIFR under project no. 12-R&D-TFR-5.01-0500. We thank World Health Organization, Regional office for South-East Asia, WHO-SEARO, team for our discussions and the teams feedback.

## REFERENCES

Agrawal, S., S. Bhandari, A. Bhattacharjee, A. Deo, N. Dixit, P. Harsha, S. Juneja, P. Kesarwani, A. Swamy, P. Patil, N. Rathod, R. Saptharishi, S. Shriram, P. Srivastava, R. Sundaresan, N. K. Vaidhiyan, and S. Yasodharan. 2020. "City-Scale Agent-Based Simulators for the Study of Non-Pharmaceutical Interventions in the Context of the COVID-19 Epidemic". *Journal of the Indian Institute of Science* 100(4):809–847.

Ferguson, N., D. Laydon, G. Nedjati Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunuba Perez, G. Cuomo-Dannenburg, A. Dighe, I. Dorigatti, H. Fu, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, L. C. Okell, S. V. Elsland, H. Thompson, R. Verity, E. Volz, H. Wang, Y. Wang, P. G. Walker, C. Walters, P. Winskill, C. Whittaker, C. A. Donnelly, S. Riley, and A. C. Ghani. 2020. "Impact of Non-pharmaceutical Interventions (NPIs) to Reduce COVID19 Mortality and Healthcare Demand". Technical Report 9, MRC Centre for Global Infectious Disease Analysis, Imperial College London, U.K.

Harsha, P., S. Juneja, D. Mittal, and R. Saptharishi. 2020. "COVID-19 Epidemic in Mumbai: Projections, Full Economic Opening, and Containment Zones Versus Contact Tracing and Testing: An Update". Technical report, TIFR Mumbai, India.

Juneja, S. 2021a. "Projections for Cases in Second Covid-19 Wave for Mumbai". https://twitter.com/sandeepjuneja66/status/1382688189287182355, accessed 15th April 2021.

Juneja, S. 2021b. "Projections for Fatalities in Second Covid-19 Wave for Mumbai". https://twitter.com/sandeepjuneja66/status/1377228683614711809, accessed 31st March 2021.

Juneja, S., and D. Mittal. 2021. "Modelling the Second Covid-19 Wave in Mumbai". Technical report, TIFR Mumbai, India.

Katella, Kathy 2022. "5 Things To Know About the Delta Variant (Yale Medicine)". https://www.yalemedicine.org/news/5-things-to-know-delta-variant-covid, accessed 5th March 2022.

Malani, A., D. Shah, G. Kang, G. Lobo, J. Shastri, M. Mohanan, R. Jain, S. Agrawal, S. Juneja, S. Imad, and U. Kolthur-Seetharam. 2020. "Seroprevalence of SARS-CoV-2 in Slums Versus Non-slums in Mumbai, India". *The Lancet Global Health* 9(2):E110–E111.

National Health Authority 2021. "National Co-Win Statistics". https://dashboard.cowin.gov.in/, accessed 15th April.

Public Health England 2021. "SARS-CoV-2 Variants of Concern and Variants Under Investigation in England". Technical Report 15, Public Health England, Wellington House, London.

Verity, R., L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, G. Cuomo-Dannenburg, H. Thompson, P. G. Walker, H. Fu, A. Dighe, J. T. Griffin, M. Baguelin, S. Bhatia, A. Boonyasiri, A. Cori, Z. Cucunuba, R. FitzJohn, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, D. Laydon, G. Nedjati-Gilani, S. Riley, S. V. Elsland, E. Volz, H. Wang, Y. Wang, X. Xi, C. A. Donnelly, A. C. Ghani, and N. M. Ferguson. 2020. "Estimates of the Severity of Coronavirus Disease 2019: A Model-based Analysis". *The Lancet Infectious Diseases* 20(6):669–677.

## AUTHOR BIOGRAPHIES

**SANDEEP JUNEJA** is a Senior Professor at School of Technology and Computer Science, Tata Institute of Fundamental Research. His research interests lie in applied probability including in sequential learning, mathematical finance, Monte Carlo methods, and game theoretic analysis of queues. Lately, he has been involved in modelling Covid-19 spread in Mumbai, and in mathematics of certain epidemiological models. His email address is juneja@tifr.res.in.

**DAKSH MITTAL** is a Research Associate at School of Technology and Computer Science, Tata Institute of Fundamental Research. His email address is dakshmittal30@gmail.com.