

---

**SUPPLEMENTARY MATERIALS: EXTENDING THE  
SUSCEPTIBLE-EXPOSED-INFECTED-REMOVED (SEIR) MODEL  
TO HANDLE THE HIGH *false negative* RATE AND  
SYMPTOM-BASED ADMINISTRATION OF COVID-19 DIAGNOSTIC  
TESTS: *SEIR-fansy***

---

A PREPRINT

**Ritwik Bhaduri** \*  
Indian Statistical Institute  
Kolkata, West Bengal  
[ritwik.bhaduri@gmail.com](mailto:ritwik.bhaduri@gmail.com)

**Ritoban Kundu** †  
Indian Statistical Institute  
Kolkata, West Bengal  
[ritoban.kundu@gmail.com](mailto:ritoban.kundu@gmail.com)

**Soumik Purkayastha**  
Department of Biostatistics  
University of Michigan  
[soumikp@umich.edu](mailto:soumikp@umich.edu)

**Michael Kleinsasser**  
Department of Biostatistics  
University of Michigan, Ann Arbor, USA.  
[mkleinsa@umich.edu](mailto:mkleinsa@umich.edu)

**Lauren J. Beesley**  
Department of Biostatistics  
University of Michigan  
[lbeesley@umich.edu](mailto:lbeesley@umich.edu)

**Bhramar Mukherjee**  
Department of Biostatistics  
University of Michigan  
[bhramar@umich.edu](mailto:bhramar@umich.edu)

September 22, 2020

---

\*Co-First Author

†Co-First Author

## 1 Differential Equations for Non-Instantaneous Testing

$$\begin{aligned}
 \frac{\partial S}{\partial t} &= -\beta \frac{S(t)}{N} (\alpha_P P(t) + \alpha_U U(t) + F(t) + T(t)) + \lambda - \mu S(t) \\
 \frac{\partial E}{\partial t} &= \beta \frac{S(t)}{N} (\alpha_P P(t) + \alpha_U U(t) + F(t) + T(t)) - \frac{E(t)}{D_e} - \mu E(t) \\
 \frac{\partial U}{\partial t} &= \frac{(1-r)E(t)}{D_e} - \frac{U(t)}{\beta_1 D_r} - \delta_1 \mu_c U(t) - \mu U(t) \\
 \frac{\partial T}{\partial t} &= \frac{r E(t)}{D_e} - \frac{T(t)}{D_t} - \mu T(t) \\
 \frac{\partial P}{\partial t} &= \frac{(1-f) T(t)}{D_t} - \frac{P(t)}{D_r} - \mu_c P(t) - \mu P(t) \\
 \frac{\partial F}{\partial t} &= \frac{f T(t)}{D_t} - \frac{\beta_2 F(t)}{D_r} - \frac{\mu_c F(t)}{\delta_2} - \mu F(t) \\
 \frac{\partial RU}{\partial t} &= \frac{U(t)}{\beta_1 D_r} + \frac{\beta_2 F(t)}{D_r} - \mu RU(t) \\
 \frac{\partial RR}{\partial t} &= \frac{P(t)}{D_r} - \mu RR(t) \\
 \frac{\partial DU}{\partial t} &= \delta_1 \mu_c U(t) + \frac{\mu_c F(t)}{\delta_2} \\
 \frac{\partial DR}{\partial t} &= \mu_c P(t)
 \end{aligned}$$

## 2 Calculation of $R_0$ for Misclassification Model

We calculate the basic reproduction number  $R_0$  using the **The Next Generation Matrix Method** as described by van den Driessche [3]. Suppose the whole population is divided into  $n$  compartments in which there are  $m < n$  infected compartments. Let  $x_i$ ,  $i = 1, 2, \dots, m$  be the number of infected individuals in the  $i^{\text{th}}$  infected compartment at time  $t$ . Now, the epidemic model is:

$$\frac{\partial x_i}{\partial t} = F_i(x) - V_i(x)$$

Here,  $V_i(x) = [V_i^-(x) - V_i^+(x)]$ , where  $V_i^+(x)$  represents the rate of transfer of individuals into compartment  $i$  from all other components containing individuals infected with the disease (here  $E$ ,  $U$ ,  $P$  and  $F$ ) and where  $V_i^-(x)$  represents the rate of transfer of individuals out of compartment  $i$ . Here,  $F_i(x)$  represents the rate of appearance of new infections in compartment  $i$ . Let  $x_0$  denote the disease free equilibrium. Now  $\mathcal{F}$  and  $\mathcal{V}$  are  $m \times m$  matrices such that :

$$\mathcal{F}_{ij} = \frac{\partial F_i}{\partial x_j}(x_0) \quad \mathcal{V}_{ij} = \frac{\partial V_i}{\partial x_j}(x_0)$$

Now,  $\mathcal{F}\mathcal{V}^{-1}$  is called the **Next Generation Matrix**. The basic reproduction number  $R_0$  is calculated by the spectral radius or the largest eigenvalue of  $\mathcal{F}\mathcal{V}^{-1}$ . For our case,

$$\mathcal{F} = \begin{pmatrix} \beta S(\alpha_P P + \alpha_U U + F) \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \mathcal{V} = \begin{pmatrix} \frac{E}{D_e} + \mu E \\ U \left( \frac{1}{\beta_1 D_r} + \delta_1 \mu_c + \mu \right) - \left( \frac{1-r}{D_e} \right) E \\ P \left( \frac{1}{D_r} + \mu_c + \mu \right) - \left( \frac{r(1-f)}{D_e} \right) E \\ F \left( \frac{\beta_2}{D_r} + \frac{\mu_c}{\delta_2} + \mu \right) - \left( \frac{rf}{D_e} \right) E \end{pmatrix}$$

Now, we calculate the jacobian of  $\mathcal{F}$  and  $\mathcal{V}$  at the Disease Free Equilibrium (DFE).

$$\dot{\mathcal{F}} = \frac{\partial \mathcal{F}}{\partial X} = \begin{bmatrix} 0 & \beta\alpha_u S_0 & \beta\alpha_p S_0 & \beta S_0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\dot{\mathcal{V}} = \frac{\partial \mathcal{V}}{\partial X} = \begin{bmatrix} \left(\frac{1}{D_e} + \mu\right) & 0 & 0 & 0 \\ -\left(\frac{1-r}{D_e}\right) & \left(\frac{1}{\beta_1 D_R} + \delta_1 \mu_c + \mu\right) & 0 & 0 \\ -\left(\frac{r(1-f)}{D_e}\right) & 0 & \left(\frac{1}{D_R} + \mu_c + \mu\right) & 0 \\ -\left(\frac{rf}{D_e}\right) & 0 & 0 & \left(\frac{\beta_2}{D_R} + \frac{\mu_c}{\delta_2} + \mu\right) \end{bmatrix}$$

Now, we need to find the inverse of  $\dot{\mathcal{V}}$ . Since it is a lower triangular matrix, it is easy to find the inverse.

$$\dot{\mathcal{V}}^{-1} = \begin{bmatrix} \frac{1}{\left(\frac{1}{D_e} + \mu\right)} & 0 & 0 & 0 \\ \left(\frac{1-r}{D_e}\right) \frac{1}{\frac{1}{\beta_1 D_R} + \delta_1 \mu_c + \mu} \cdot \frac{1}{\left(\frac{1}{D_e} + \mu\right)} & \frac{1}{\beta_1 D_R + \delta_1 \mu_c + \mu} & 0 & 0 \\ \frac{r(1-f)}{D_e} \cdot \frac{1}{\frac{1}{D_R} + \mu_c + \mu} \cdot \frac{1}{\left(\frac{1}{D_e} + \mu\right)} & 0 & \frac{1}{\frac{1}{D_R} + \mu_c + \mu} & 0 \\ \left(\frac{rf}{D_e}\right) \frac{1}{\frac{\beta_1}{D_R} + \frac{\mu_c}{\delta_2} + \mu} \cdot \frac{1}{\left(\frac{1}{D_e} + \mu\right)} & 0 & 0 & \frac{1}{\frac{\beta_1}{D_R} + \frac{\mu_c}{\delta_2} + \mu} \end{bmatrix}$$

Now, we multiply  $\dot{\mathcal{F}}$  and  $\dot{\mathcal{V}}^{-1}$ . The spectral radius of  $\dot{\mathcal{F}}\dot{\mathcal{V}}^{-1}$  gives the basic reproduction number. Note that the matrix  $\dot{\mathcal{F}}\dot{\mathcal{V}}^{-1}$  has only one non-zero row, which is the first one. All other rows of  $\dot{\mathcal{F}}\dot{\mathcal{V}}^{-1}$  are 0. Hence, the spectral radius is given by  $\left(\dot{\mathcal{F}}\dot{\mathcal{V}}^{-1}\right)_{11}$  (i.e., the  $(1,1)^{th}$  element of  $\dot{\mathcal{F}}\dot{\mathcal{V}}^{-1}$ ). So,

$$R_0 = \left(\dot{\mathcal{F}}\dot{\mathcal{V}}^{-1}\right)_{11} = \frac{\beta \cdot S_0}{\mu D_e + 1} \left( \frac{\alpha_u(1-r)}{\frac{1}{\beta_1 D_r} + \delta_1 \mu_c + \mu} + \frac{\alpha_p r(1-f)}{\frac{1}{D_r} + \mu_c + \mu} + \frac{rf}{\frac{\beta_2}{D_r} + \frac{\mu_c}{\delta_2} + \mu} \right)$$

### 3 Motivation Behind the 3-parameter Multinomial Model

We can observe from figure (S1) that mCFR varies widely across countries and also across time. Now note that while countries like Belgium, USA, Italy, Spain have very high mCFR, India and Russia have comparatively much lower mCFR. Also, we observe that initially most countries experience high mCFR and it gradually settles to a comparatively lower value in most countries as the case counts and recoveries rise. This supports modeling mCFR as a time-varying quantity.

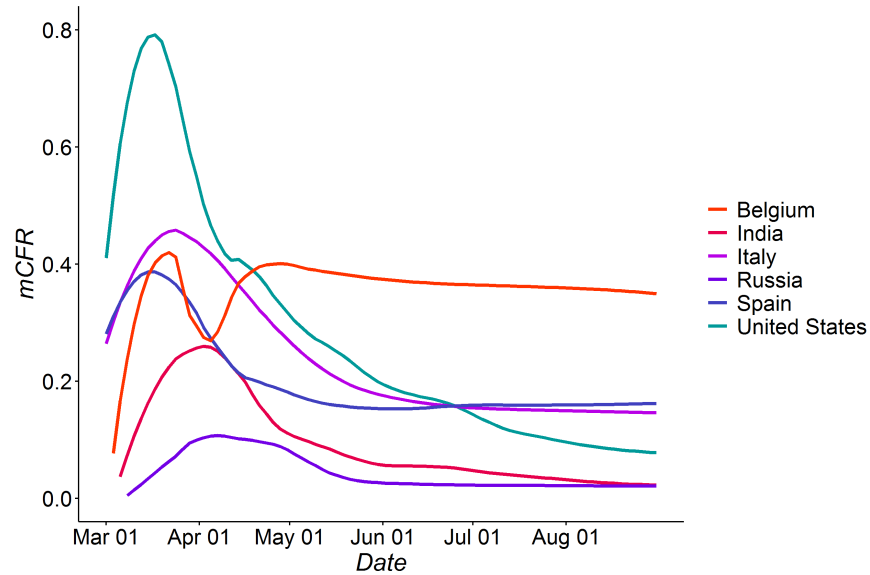


Figure S1: Variation of mCFR with time

#### 4 Transmission Dynamics Diagram for Symptoms Model

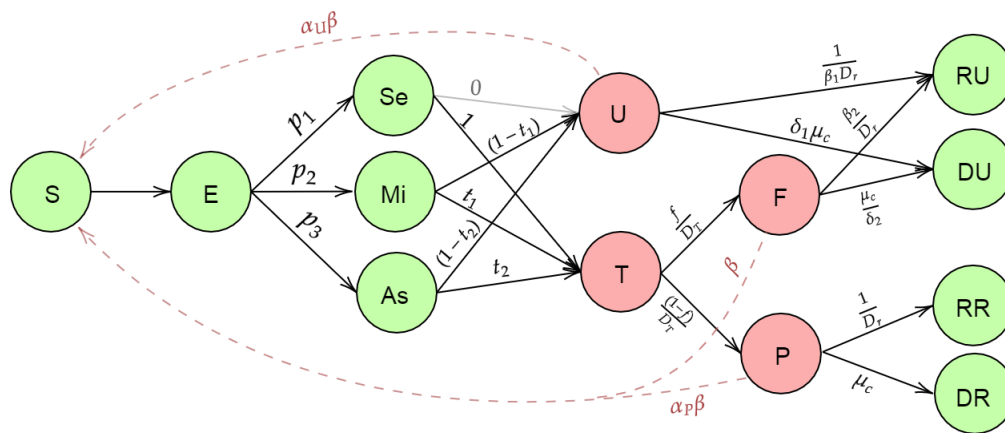


Figure S2: Misclassification Model with Symptoms

The only difference in this model from the Multinomial 2-parameter model is that, from the Exposed Node, an infected person can enter into one of three possible nodes: Severe Symptomatic Infectious (Se), Mild Symptomatic Infectious (Mi) and Asymptomatic Infectious (As) with probabilities  $p_1$ ,  $p_2$ , and  $p_3$  respectively. Now, we assume people with severe symptoms (people in  $Se$ ) are tested with probability 1. While the  $Mi$  people and the  $As$  people are tested with probabilities  $t_1$  and  $t_2$ .

#### 5 Misclassification model - complete distributional assumptions

In the main paper, we have given the distribution of observed nodes given the other nodes and parameters. Here, we describe the distribution of the latent nodes also. After getting the estimates of the parameters using MCMC, we want to obtain model-based forecasts. In order to predict the future counts, we use the following multinomial random sampling

strategy:

$$\begin{aligned}
 \zeta_{S \rightarrow E}, \zeta_{S \rightarrow O}, \zeta_{S \rightarrow S} &\sim \text{Multinomial}(S(t-1), p_{S \rightarrow E}, \mu, 1 - p_{S \rightarrow E} - \mu) \\
 \zeta_{E \rightarrow U}, \zeta_{E \rightarrow P}, \zeta_{E \rightarrow F}, \zeta_{E \rightarrow O}, \zeta_{E \rightarrow E} &\sim \text{Multinomial}(E(t-1), \frac{(1-r)}{D_e}, \frac{r(1-f)}{D_e}, \frac{rf}{D_e}, \mu, \\
 &\quad 1 - p_{E \rightarrow U} - p_{E \rightarrow P} - p_{E \rightarrow F} - \mu) \\
 \zeta_{U \rightarrow RU}, \zeta_{U \rightarrow DU}, \zeta_{U \rightarrow O}, \zeta_{U \rightarrow U} &\sim \text{Multinomial}(U(t-1), \beta_1^{-1} D_r^{-1}, \delta_1 \mu_c, \mu, 1 - \beta_1^{-1} D_r^{-1} - \delta_1 \mu_c - \mu) \\
 \zeta_{P \rightarrow RR}, \zeta_{P \rightarrow DR}, \zeta_{P \rightarrow O}, \zeta_{P \rightarrow P} &\sim \text{Multinomial}(P(t-1), D_r^{-1}, \mu_c, \mu, 1 - D_r^{-1} - \mu_c - \mu) \\
 \zeta_{F \rightarrow RU}, \zeta_{F \rightarrow DU}, \zeta_{F \rightarrow O}, \zeta_{F \rightarrow F} &\sim \text{Multinomial}(F(t-1), \beta_2 D_r^{-1}, \delta_2^{-1} \mu_c, \mu, 1 - \beta_2 D_r^{-1} - \delta_2^{-1} \mu_c - \mu) \\
 \zeta_{RU \rightarrow O}, \zeta_{RU \rightarrow RU} &\sim \text{Multinomial}(RU(t-1), \mu, 1 - \mu) \\
 \zeta_{RR \rightarrow O}, \zeta_{RR \rightarrow RR} &\sim \text{Multinomial}(RR(t-1), \mu, 1 - \mu)
 \end{aligned}$$

where  $\zeta_{X \rightarrow Y}$  denotes the number of individuals moving from compartment  $X$  to compartment  $Y$  at time  $t$ .  $\zeta_{X \rightarrow 0}$  denotes the number of individuals in compartment  $X$  that die at time  $t$ . The counts in each compartment at time  $t$  are given by,

$$\begin{aligned}
 S(t) &= \zeta_{S \rightarrow S} \\
 E(t) &= \zeta_{E \rightarrow E} + \zeta_{S \rightarrow E} \\
 U(t) &= \zeta_{U \rightarrow U} + \zeta_{E \rightarrow U} \\
 P(t) &= \zeta_{P \rightarrow P} + \zeta_{E \rightarrow P} \\
 F(t) &= \zeta_{F \rightarrow F} + \zeta_{E \rightarrow F} \\
 RU(t) &= \zeta_{RU \rightarrow RU} + \zeta_{U \rightarrow RU} + \zeta_{F \rightarrow RU} \\
 RR(t) &= \zeta_{RR \rightarrow RR} + \zeta_{P \rightarrow RR} \\
 DU(t) &= \zeta_{DU \rightarrow DU} + \zeta_{U \rightarrow DU} + \zeta_{F \rightarrow DU} \\
 DR(t) &= \zeta_{DR \rightarrow DR} + \zeta_{P \rightarrow DR}
 \end{aligned}$$

Given the parameters and the counts at time  $(t-1)$ , we obtain the predicted counts for time  $t$ . Using this approach, we obtain the posterior means of the future predicted counts at each of the 9 compartments using the MCMC estimated parameters. For the purpose of future prediction beyond the training period, we use the parameter estimates from the last time period.

## 6 Differential Equations for the Selection Model

$$\begin{aligned}
 \frac{\partial S}{\partial t} &= \frac{\partial S e^0}{\partial t} + \frac{\partial M i^0}{\partial t} + \frac{\partial A s^0}{\partial t} \\
 \frac{\partial S e^0}{\partial t} &= \frac{-\beta S e^0 (\alpha_p P(t) + \alpha_{ui} UI(t) + F(t))}{N} - \mu S e^0(t) \\
 \frac{\partial M i^0}{\partial t} &= \frac{-\beta M i^0 (\alpha_p P(t) + \alpha_{ui} UI(t) + F(t))}{N} - \mu M i^0(t) \\
 \frac{\partial A s^0}{\partial t} &= \frac{-\beta A s^0 (\alpha_p P(t) + \alpha_{ui} UI(t) + F(t))}{N} - \mu A s^0(t) \\
 \frac{\partial E}{\partial t} &= \frac{-\beta S (\alpha_p P(t) + \alpha_{ui} UI(t) + F(t))}{N} - \frac{p_{11} E(t)}{D_e} - \frac{p_{12} E(t)}{D_e} - \frac{p_{13} E(t)}{D_e} \\
 S e^1(t) &= \frac{p_{11} E(t)}{D_e} \quad M i^1(t) = \frac{p_{12} E(t)}{D_e} \quad A s^1(t) = \frac{p_{13} E(t)}{D_e}
 \end{aligned} \tag{1}$$

$$\begin{aligned}
 \frac{\partial UI}{\partial t} &= Mi^1(t) - t_1(T - Se^1(t) - (Se^0(t) + \frac{\partial Se^0}{\partial t})) \left( \frac{Mi^1(t)}{Mi^1(t) + (Mi^0(t) + \frac{\partial Mi^0}{\partial t})} \right) \\
 &\quad + As^1(t) - t_2(T - Se^1(t) - (Se^0(t) + \frac{\partial Se^0}{\partial t})) \left( \frac{As^1(t)}{As^1(t) + (As^0(t) + \frac{\partial As^0}{\partial t})} \right) \\
 &\quad - \mu UI(t) - \delta_1 \mu_c UI(t) - \frac{UI(t)}{\beta_1 D_r} \\
 \frac{\partial P}{\partial t} &= (1-f)(t_1(T - Se^1(t) - (Se^0(t) + \frac{\partial Se^0}{\partial t})) \left( \frac{Mi^1(t)}{Mi^1(t) + (Mi^0(t) + \frac{\partial Mi^0}{\partial t})} \right) \\
 &\quad + (1-f)(t_2(T - Se^1(t) - (Se^0(t) + \frac{\partial Se^0}{\partial t})) \left( \frac{As^1(t)}{As^1(t) + (As^0(t) + \frac{\partial As^0}{\partial t})} \right) \\
 &\quad + (1-f)Se^1(t) - \mu P(t) - \mu_c P(t) - \frac{P(t)}{D_r} \\
 \frac{\partial F}{\partial t} &= f(t_1(T - Se^1(t) - (Se^0(t) + \frac{\partial Se^0}{\partial t})) \left( \frac{Mi^1(t)}{Mi^1(t) + (Mi^0(t) + \frac{\partial Mi^0}{\partial t})} \right) \\
 &\quad + f(t_2(T - Se^1(t) - (Se^0(t) + \frac{\partial Se^0}{\partial t})) \left( \frac{As^1(t)}{As^1(t) + (As^0(t) + \frac{\partial As^0}{\partial t})} \right) \\
 &\quad + fSe^1(t) - \mu F(t) - \frac{\mu_c F(t)}{\delta_2} - \frac{\beta_2 F(t)}{D_r} \\
 \frac{\partial RU}{\partial t} &= \frac{UI(t)}{\beta_1 D_R} + \frac{\beta_2 F(t)}{D_R} - \mu x_{R_U} \\
 \frac{\partial RR}{\partial t} &= \frac{P(t)}{D_R} - \mu R_R \\
 \frac{\partial DU}{\partial t} &= \delta_1 \mu_c UI(t) + \frac{\mu_c F(t)}{\delta_2} \\
 \frac{\partial DR}{\partial t} &= \mu_c P(t)
 \end{aligned}$$

## 7 Selection Model : Complete Distributional Assumptions

To generate data using the test model, we perform the following steps.

$$\zeta_{S \rightarrow E}, \zeta_{S \rightarrow O}, \zeta_{S \rightarrow S} \sim \text{Multinomial}(S(t-1), p_{S \rightarrow E}, \mu, 1 - p_{S \rightarrow E} - \mu)$$

Now, we assume the probability of an individual being severely symptomatic, mildly symptomatic or asymptomatic given he/she is susceptible is given by the probability vector  $\mathbf{p}_0 = (p_{01}, p_{02}, p_{03})$ . The probability for an infected individual is given by  $\mathbf{p}_1 = (p_{11}, p_{12}, p_{13})$ . To obtain the number of individuals in the groups  $Se^0$ ,  $Mi^0$ , and  $As^0$ , we assume that the outgoing individuals from the susceptible group follow the distribution given by  $\mathbf{p}_0$ .

$$Se_{new}^0(t), Mi_{new}^0(t), As_{new}^0(t) \sim \text{Multinomial}(\zeta_{S \rightarrow S}, \mathbf{p}_0)$$

Now, from our assumption that the individuals in  $E$  follow the distribution given by  $\mathbf{p}_1$ , we can write,

$$\zeta_{E \rightarrow Se^1}, \zeta_{E \rightarrow Mi^1}, \zeta_{E \rightarrow As^1}, \zeta_{E \rightarrow O}, \zeta_{E \rightarrow E} \sim \text{Multinomial} \left( E(t-1), \left( \frac{\mathbf{p}_1}{D_e}, \mu, 1 - \frac{1}{D_e} - \mu \right) \right)$$

Recall, we assume that all individuals with severe symptoms are tested provided adequate tests are available. This implies

$$Se_{tested}^0 = Se^0(t) \quad Se_{tested}^1 = Se^1(t) \quad Se_{tested} = Se_{tested}^0 + Se_{tested}^1$$

In the case when number of test  $T(t)$  is less than that of severe individuals, we assume that the number of tested  $Se^1$  and  $Se^0$  individuals is proportional to their respective counts.

$$Se_{tested}^0, Se_{tested}^1 \sim \text{Multinomial} \left( Se(t), \left( \frac{Se^0(t)}{Se^0(t) + Se^1(t)}, \frac{Se^1(t)}{Se^0(t) + Se^1(t)} \right) \right)$$

If the total number of remaining tests is greater than or equal to the number of mild and asymptomatic individuals, then all of them are tested i.e :

$$Mi_{tested}^0 = Mi^0(t), Mi_{tested}^1 = Mi^1(t), As_{tested}^0 = As^0(t), As_{tested}^1 = As^1(t)$$

If number of tests are not adequate for all the mild symptomatic and asymptomatic people to be tested, then the remaining tests (after testing the severe symptomatic people) are distributed among the mildly symptomatic and asymptomatic individuals in the ratio  $t_1 : t_2$ .

$$Mi_{tested}, As_{tested} \sim \text{Binomial} (T - Se_{tested}, (t_1, t_2))$$

As we did in the case of severely symptomatic, we allocate the tests among infected and uninfected mildly symptomatic (and also asymptomatic) individuals randomly.

$$Mi_{tested}^0, Mi_{tested}^1 \sim \text{Binomial} \left( Mi_{tested}, \left( \frac{Mi^0(t)}{Mi^0(t) + Mi^1(t)}, \frac{Mi^1(t)}{Mi^0(t) + Mi^1(t)} \right) \right)$$

$$As_{tested}^0, As_{tested}^1 \sim \text{Binomial} \left( As_{tested}, \left( \frac{As^0(t)}{As^0(t) + As^1(t)}, \frac{As^1(t)}{As^0(t) + As^1(t)} \right) \right)$$

$$\zeta_{UI \rightarrow RU}, \zeta_{UI \rightarrow DU}, \zeta_{UI \rightarrow O}, \zeta_{UI \rightarrow UI} \sim \text{Multinomial}(UI(t-1), \beta_1^{-1} D_r^{-1}, \delta_1 \mu_c, \mu, 1 - \beta_1^{-1} D_r^{-1} - \delta_1 \mu_c - \mu)$$

$$\zeta_{P \rightarrow RR}, \zeta_{P \rightarrow DR}, \zeta_{P \rightarrow O}, \zeta_{P \rightarrow P} \sim \text{Multinomial}(P(t-1), D_r^{-1}, \mu_c, \mu, 1 - D_r^{-1} - \mu_c - \mu)$$

$$\zeta_{F \rightarrow RU}, \zeta_{F \rightarrow DU}, \zeta_{F \rightarrow O}, \zeta_{F \rightarrow F} \sim \text{Multinomial}(F(t-1), \beta_2 D_r^{-1}, \delta_2^{-1} \mu_c, \mu, 1 - \beta_2 D_r^{-1} - \delta_2^{-1} \mu_c - \mu)$$

$$\zeta_{RU \rightarrow O}, \zeta_{RU \rightarrow RU} \sim \text{Multinomial}(RU(t-1), \mu, 1 - \mu)$$

$$\zeta_{RR \rightarrow O}, \zeta_{RR \rightarrow RR} \sim \text{Multinomial}(RR(t-1), \mu, 1 - \mu)$$

We also assume the false negative probability =  $f$ . The numbers of new individuals to P and F states are given by :

$$P_{new}, F_{new} \sim \text{Multinomial} \left( Se_{tested}^1 + Mi_{tested}^1 + As_{tested}^1, (1 - f, f) \right)$$

Finally we write the number of people in each state at time  $t$  as follows :

$$UI(t) = \zeta_{UI \rightarrow UI} + Se_{untested}^1 + Mi_{untested}^1 + As_{untested}^1$$

$$P(t) = \zeta_{P \rightarrow P} + P_{new}$$

$$F(t) = \zeta_{F \rightarrow F} + F_{new}$$

$$RU(t) = \zeta_{RU \rightarrow RU} + \zeta_{UI \rightarrow RU} + \zeta_{F \rightarrow RU}$$

$$RR(t) = \zeta_{RR \rightarrow RR} + \zeta_{P \rightarrow RR}$$

$$DU(t) = \zeta_{DU \rightarrow DU} + \zeta_{UI \rightarrow DU} + \zeta_{F \rightarrow DU}$$

$$DR(t) = \zeta_{DR \rightarrow DR} + \zeta_{P \rightarrow DR}$$

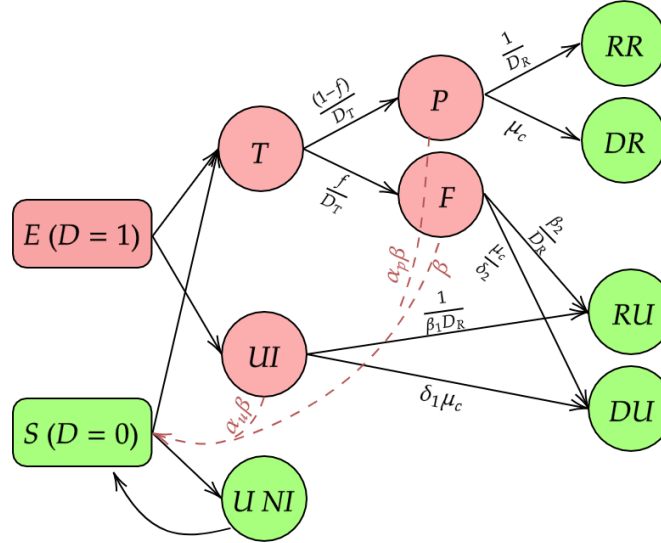


Figure S3: Special case of Selection Model : Uniform testing

## 8 Special case of Selection Model : Uniform testing

To understand the effect of selection bias on  $R_0$ , we consider a special case of the Selection model where we assume uniform testing. Here, uniform testing means tests are offered independently of symptoms. The model is represented diagrammatically in Figure (S3).

The transmission dynamics of this model are very similar to the Selection model. We provide the differential equations describing the dynamics of the nodes  $S, E, UI, P$  and  $F$ . The rest of the nodes ( $RU, RR, DU$  and  $DR$ ) have differential equations exactly same as in Selection Model.

$$\begin{aligned} \frac{\partial S}{\partial t} &= -\beta \frac{S(t)}{N} (\alpha_P P(t) + \alpha_U U(t) + F(t)) + \lambda - \mu S(t) \\ \frac{\partial E}{\partial t} &= \beta \frac{S(t)}{N} (\alpha_P P(t) + \alpha_U U(t) + F(t)) - \frac{E(t)}{D_e} - \mu E(t) \\ \frac{\partial UI}{\partial t} &= \frac{E(t)}{D_e} - T \frac{E(t)}{S(t) + E(t)} - \frac{UI(t)}{\beta_1 D_r} - \delta_1 \mu_c UI(t) - \mu UI(t) \\ \frac{\partial P}{\partial t} &= (1-f) T \frac{E(t)}{S(t) + E(t)} - \frac{P(t)}{D_r} - \mu_c P(t) - \mu P(t) \\ \frac{\partial F}{\partial t} &= f \cdot T \frac{E(t)}{S(t) + E(t)} - \frac{\beta_2 F(t)}{D_r} - \frac{\mu_c F(t)}{\delta_2} - \mu F(t) \end{aligned}$$

Now that we have all the differential equations governing the dynamics of this model, we calculate the basic reproduction number using Next Generation Matrix method [3]. Using calculations similar to what we did for the Misclassification model, we arrive at the following expression of  $R_0$  for Uniform testing model.

$$R_0 = \frac{\beta}{\mu + \frac{1}{D_e}} \left[ \frac{\alpha_u \left( \frac{1}{D_e} - T \right)}{\frac{1}{\beta_1 D_r} - \delta_1 \mu_c - \mu} + \frac{\alpha_p (1-f) T}{\frac{1}{D_r} - \mu_c - \mu} + \frac{f T}{\frac{\beta_2}{D_r} - \frac{\mu_c}{\delta_2} - \mu} \right] \quad (2)$$



## 9 Real Data Analysis for India

### 9.1 Table for Initial Values of the Different Compartments for India

Variable	Value	Justification
S(0)	1340933446	$N - (E(0) + U(0) + P(0) + F(0) + RU(0) + RR(0) + DU(0) + DR(0))$ ( $N = 1341$ million)
E(0)	48780	Thrice the number of current infected
U(0)	13821	$\frac{1-r}{r} (P(0) + F(0))$
P(0)	1829	Reported current infected on 1 <sup>st</sup> April
F(0)	610	$\frac{f}{1-f} P(0)$
RU(0)	958	$\left(\frac{1-r}{r} + \frac{f}{1-f}\right) RR(0)$
RR(0)	169	Reported recovered on 1 <sup>st</sup> April
DU(0)	329	$\left(\frac{1-r}{r} + \frac{f}{1-f}\right) DR(0)$
DR(0)	58	Reported deceased on 1 <sup>st</sup> April

Table S1: Initial Values of the Different Compartments

Now for India, we have fitted data from 1<sup>st</sup> April to 30<sup>th</sup> June. So for our prediction, we need the counts of the different compartments on the initial date, that is on 1<sup>st</sup> April. So the table S1 presents the counts of the compartments for India on 1<sup>st</sup> April.

### 9.2 Additional Plots for India

We have done our estimation using the MCMC Metropolis Method and predicted the counts for the different compartments by using the posterior means conditional on the estimated parameters. So the large number of iterations of MCMC provide a 95% credible interval for the parameters as well as for the predictions of the compartments. So the following figure shows the credible regions for the Reported Active, False negative active and Untested Active cases.

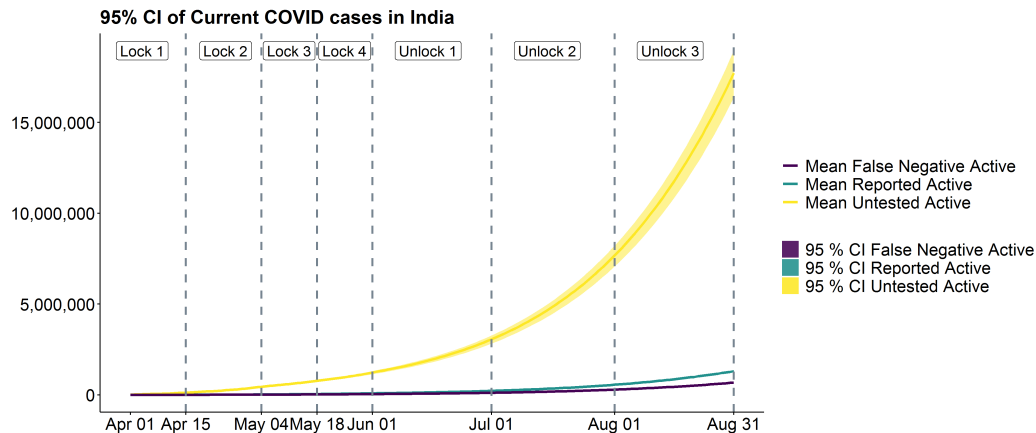


Figure S4: 95% Credible Intervals of estimates of Current Active Cases in India

Figure S4, shows the 95% CI's of the estimates of Current Active cases in India from 1<sup>st</sup> April to 31<sup>st</sup> August. Here, we have fit the model using the data from 1<sup>st</sup> April to 30<sup>th</sup> June dividing the training period into 5 parts as described

earlier. The estimates of  $\beta$  and  $r$  for the last period (1<sup>st</sup> June to 30<sup>th</sup> June) was used to predict the cases from 1<sup>st</sup> July to 31<sup>st</sup> August. As it is easily visible from Figure (S4), the CI's for estimates of untested cases is the highest. This is expected due to the much higher estimated number of untested cases than tested positives or false negatives.

## 10 Additional Plots for Maharashtra and Delhi

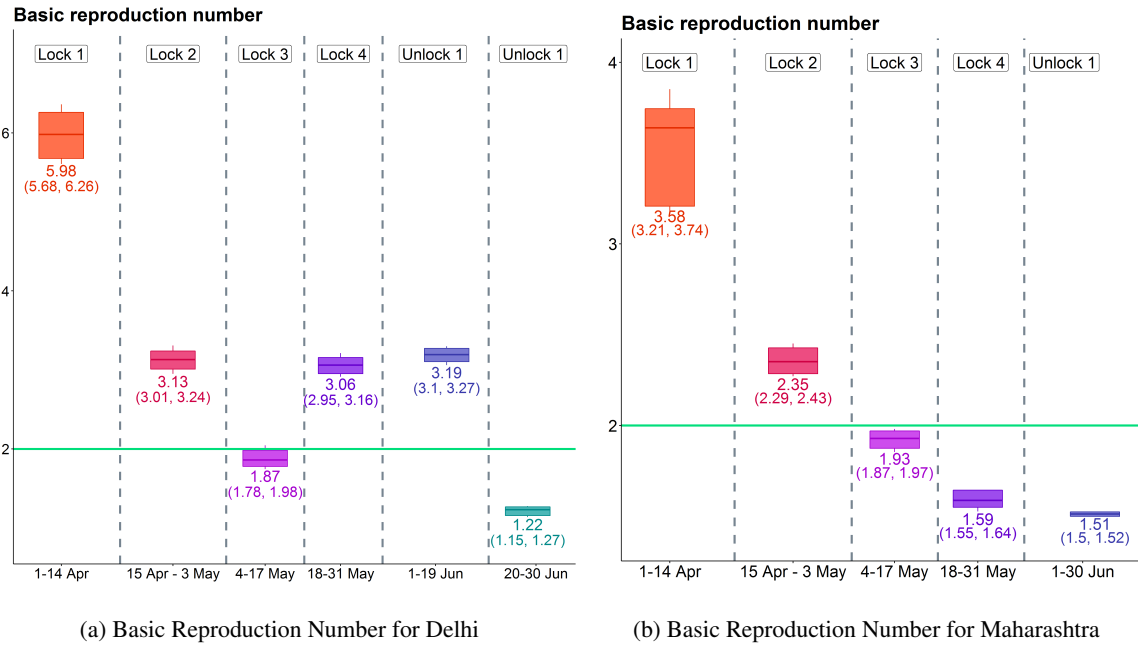


Figure S5: Basic reproduction number of Delhi and Maharashtra

From Figure (S5b), we observe that the estimates of basic reproduction number for Maharashtra have been strictly decreasing throughout the training period. The value of  $R_0$  started at 3.58 in the first 2 weeks of April and eventually dropped to 1.51 in June. The value of  $R_0$  dropped below 2 first time in lockdown 3 which was from 4<sup>th</sup> to 17<sup>th</sup> May. From figure (S5a), we note that the basic reproduction number was quite high ( $> 3$ ) in 1<sup>st</sup>, 2<sup>nd</sup>, 4<sup>th</sup> and 5<sup>th</sup> periods and it decreased to 1.22 in the 6<sup>th</sup> period.

## 11 Results for Simulations - Effect of Misclassification

In the main paper we have shown the effect of misclassification on number of total active cases. We concluded that the effect of misclassification on total active cases was substantial, but it was negligible on reported active cases. Here, we provide the mean estimates of  $R_0$  obtained by the 3 different models with 3 different false negative rates  $f = 0, 0.15$  and  $0.3$ .

	Basic Reproduction Number					MRE		
	$R_{01}$	$R_{02}$	$R_{03}$	$R_{04}$	$R_{05}$	Lower C.I	Mean	Upper C.I
Actual	3.99	3.65	2.12	1.59	1.69	-	-	-
Predicted Using $f = 0$	3.64	3.51	1.97	1.48	1.65	0.0036	0.0041	0.0045
Predicted Using $f = 0.15$	3.52	3.64	2.01	1.51	1.69	0.0035	0.004	0.0044
Predicted Using $f = 0.3$	3.83	3.73	2.04	1.53	1.71	0.0009	0.0012	0.0015

Table S2: Effect on Basic Reproduction Number

It is quite evident from the table (S2) that the  $R_0$  is quite robust with the change of the value of false negative rate ( $f$ ). Under all the false negative rates, the estimation of  $R_0$  is quite accurate which is evident from the MRE provided in the table (S2).

## 12 Sensitivity Analysis

Since we have not estimated the values of quite a few parameters, a sensitivity analysis is necessary. Now, as doing sensitivity analysis of all the parameters and initial values is impractical, we will do sensitivity analysis for the following parameters only.

1.  $E_0$  : The initial value of Exposed had been chosen as 3 times the sum of initial values of Untested, Confirmed and False Negative cases. Such a choice might seem arbitrary. Hence, we try 4 different values of  $E_0$  and check how the estimates of  $R_0$  and Current Active cases vary across different values of  $E_0$ .
2.  $\alpha_U$  : The value of  $\alpha_U$  had been taken as 0.5 in the main analysis. We also assumed  $\alpha_P = 0.5$ . So, we effectively assumed that the rate of transmission of disease by untested and tested positive individuals was same. Some things to consider when choosing the value of  $\alpha_U$  and  $\alpha_P$  were that individuals who were tested positive are quarantined and/or hospitalized reducing their rate of transmitting the disease. And untested cases are predominantly asymptomatic cases whose rate of spreading the virus is much less than symptomatic cases. So, we have  $\alpha_U < 1$  and  $\alpha_P < 1$ . However, we do not know if  $\alpha_U > \alpha_P$  or  $\alpha_U < \alpha_P$ . So we try 4 different values of  $\alpha_U$  here which are  $\alpha_U = 0.3, 0.5, 0.7$  and  $1$ .
3.  $D_e$  : We stated in the beginning of this paper that we have assumed the Incubation period equals the Latency period ( $= D_e$ ). We have taken  $D_e = 5.2$  days following the results by Lauer et al. [2]. However research by other groups suggest different values of incubation period like 6.4 days by Becker et al. [1] etc. So we consider 3 values of  $D_e$  for sensitivity analysis. They are  $D_e = 6.4, 5.2$  and  $4.1$  (lower limit of 95% CI of estimates of incubation period by Lauer et al. [2])
4.  $k$  : For Multinomial Symptoms model, one important parameters is  $k$  which is the ratio of probability of a mildly symptomatic person being tested to that of an asymptomatic person being tested. Since the probability of testing is higher for a mildly symptomatic person than an asymptomatic person, so  $k > 1$ . In our main analysis, we assumed  $k = 4$ . The choice of  $k$  was not supported by any data. So, we try 4 different values of  $k$  :  $k = 3, 4, 5$  and  $6$  and look at the different estimates.

### 12.1 Effect of initial value of Exposed

We start with the initial value of Exposed individuals. Throughout our analysis we have assumed that the number of exposed individuals on the starting day i.e. 1<sup>st</sup> April was thrice the number of total expected infected up to that day. So we check how much our estimates vary if we vary the starting value of Exposed ( $E_0$ ). So, we use 4 starting values for  $E_0$ :

1.  $E_0 = U_0 + P_0 + F_0$
2.  $E_0 = 2(U_0 + P_0 + F_0)$
3.  $E_0 = 3(U_0 + P_0 + F_0)$
4.  $E_0 = 4(U_0 + P_0 + F_0)$

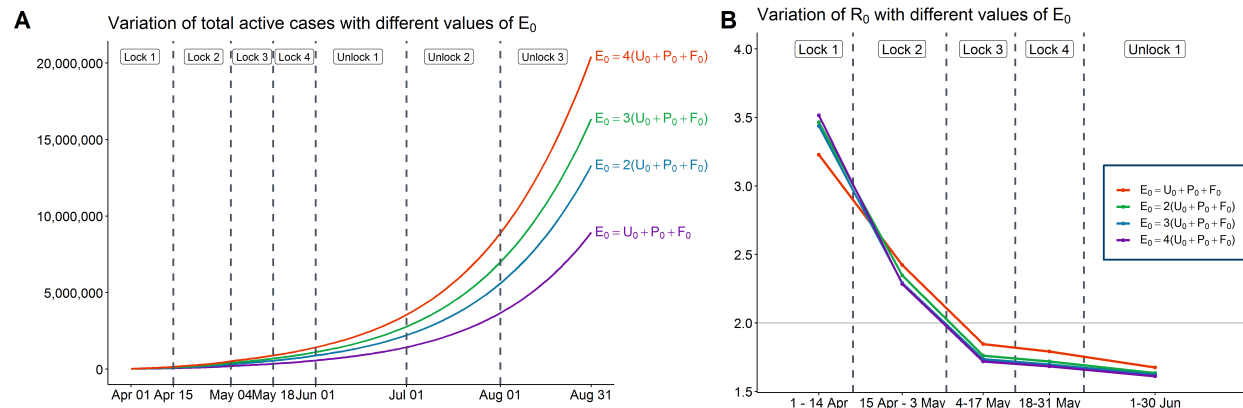


Figure S6: Variation of estimates of  $R_0$  with different values of  $E_0$

We can observe from subfigure B of Figure (S6) that our estimates of  $R_0$  are relatively robust with respect to choice of initial values of exposed. The only substantial variation is observed in the first time period - 1<sup>st</sup> - 14<sup>th</sup> April. Now, let us look at how the estimates of number of active cases change with different initial values.

We can observe from subfigure A of Figure (S6) that all the estimates for total active cases increases with increasing values of  $E_0$ . The estimate of total active cases on 30<sup>th</sup> June for  $E_0 = 4(U_0 + P_0 + F_0)$  was more than 2 times that for  $E_0 = (U_0 + P_0 + F_0)$ . Hence we observe that though the estimates of total active cases vary substantially with different initial number of Exposed people, the estimates of Basic Reproduction Number are much more robust to such variations. Now we look at the effect of  $\alpha_U$  on our estimates.

### 12.1.1 Effect of $\alpha_u$

In our main analysis we assumed  $\alpha_U = 0.5$ . Here, we try 4 different values of  $\alpha_U$ ,  $\alpha = 0.3, 0.5, 0.7$  and 1. First, we

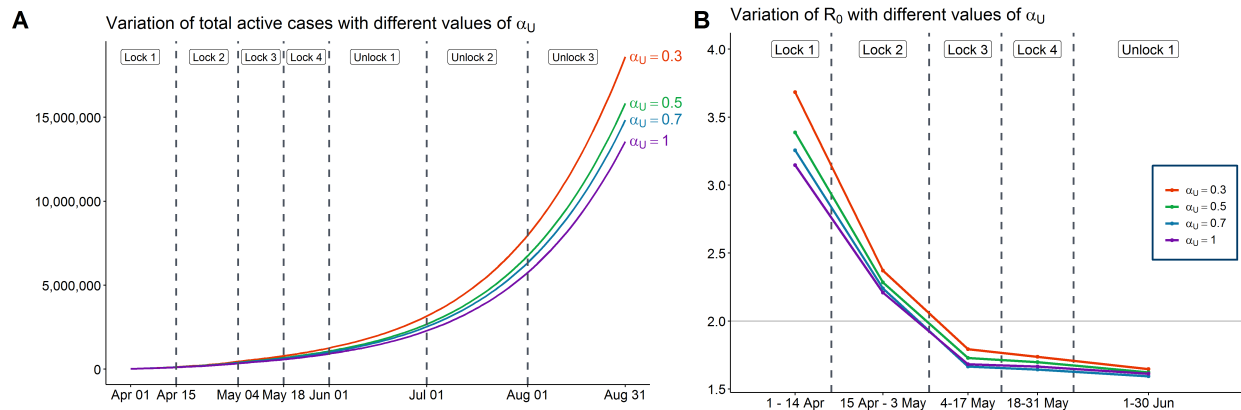


Figure S7: Variation of estimates of  $R_0$  with different values of  $\alpha_U$

look at the estimates of  $R_0$ . Similar to the previous section, from subfigure B of (S7), we observe that the estimates of  $R_0$  are more or less similar for different values of  $\alpha_U$ . Once again, the only  $R_0$  that substantially varies with different values of  $\alpha_U$  is the first one i.e.  $R_{01}$ . Now, we look at the estimates of total active cases.

Subfigure A of (S7) shows that the estimated value of total active cases decreases with increasing value of  $\alpha_U$ . The reason behind this is if the value of  $\alpha_U$  is higher, then a smaller number of untested cases will spread the same amount of infection as a larger number of cases would have if the value of  $\alpha_U$  had been lower.

So, once again, we observe that while the estimates of the number of active cases are influenced heavily by  $\alpha_U$ , the estimates of  $R_0$  remain relatively unaffected by the change.

### 12.2 Effect of $D_e$

For our main analysis, we had assumed  $D_e = 5.2$ . Here, we try 2 more values of  $D_e$  and check how our estimates vary with different values of  $D_e$ .

Again, we observe from subfigure B of (S8), that estimates of  $R_0$  are very robust with respect to different values of  $D_e$ . From subfigure A of Figure (S8), we note that the predicted number of active cases vary with the different values of  $D_e$ . However, unlike the previous cases, we do not substantial variation with different variation of  $D_e$ . Now, we move on to our last sensitivity analysis which is for the value of  $k$  in multinomial symptoms.

### 12.3 Effect of $k$

In Multinomial symptoms model, we defined  $k$  as the ratio of the probability of a mildly symptomatic individual getting tested to the same for an asymptomatic individual. We chose the value  $k = 4$  for our main analysis. We had argued why the value of  $k$  should be greater than 1 but could not provide any justification for choosing that particular value. So, we try 4 different values of  $k$ :  $k = 3, 4, 5$  and 6. We will start with the estimates of  $R_0$ .

Figure subfigure B (S9) shows that similar to previous cases, the estimates of  $R_0$  do not vary much with different values of  $k$ . We now look at the estimates of total active cases. From subfigure A of Figure (S9), we note that the estimates of

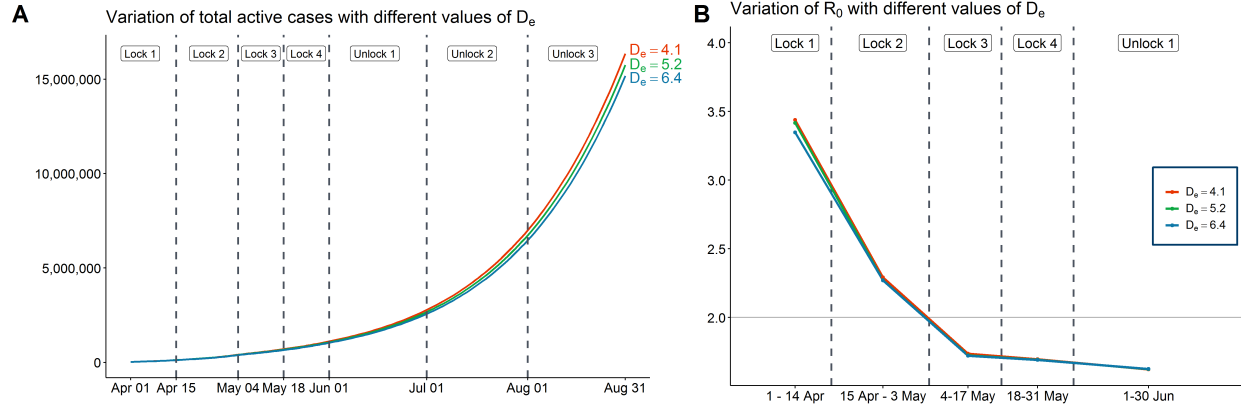


Figure S8: Variation of estimates of  $R_0$  with different values of  $D_e$

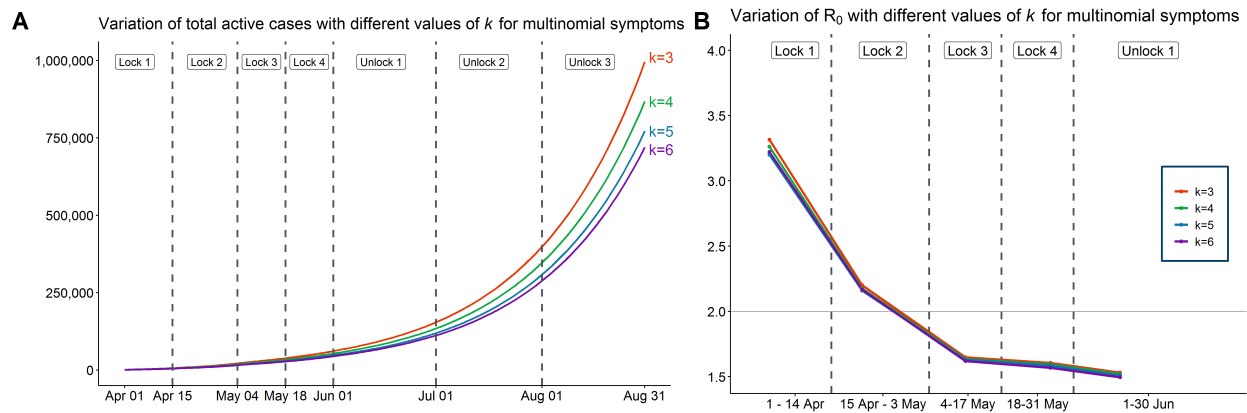


Figure S9: Variation of estimates of  $R_0$  with different values of  $k$  for multinomial symptoms model

total active cases vary with different values of  $k$  and with higher values of  $k$  we have lower predictions of number of total active cases.

To summarize, we observe that the estimates of the Basic Reproduction number are not substantially influenced by these parameters with an exception of the first Reproduction number. We also note that the estimated number of active cases varies with different values of parameters in most of the cases. It is clearly visible that the sensitivity of the total active case predictions vary across parameters. While it does not vary much with  $D_e$ , there is substantial variation with different values of  $E_0$ .

## References

- [1] B. Ja. Klinkenberg d, wallinga j. incubation period of 2019 novel coronavirus (2019-ncov) infections among travellers from wuhan, china, 20-28 january 2020. *Euro Surveill*, 25:1–6, 2020.
- [2] S. A. Lauer and K. H. Grantz. Qifang bi, forrest k jones, qulu zheng, hannah r meredith, andrew s azman, nicholas g reich, and justin lessler. 2020. “the incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: estimation and application.”. *Annals of internal medicine*, 14, 2020.
- [3] P. van den Driessche. Reproduction numbers of infectious disease models. *Infectious Disease Modelling*, 2(3): 288–303, 2017.