

# Stochastic Invasion of a VoC and Onward Transmission

March 2021

We cross-check the results of our Gillespie simulation with analytic results for a continuous-time multi-type branching process model with immigration that models the stochastic invasion of a variant strain into a heterogeneous population.

## 0.1 Model Description

We consider the disease dynamics of an invading Variant of Concern (VoC) into a population that has had previous exposure to a wild-type strain and has undergone an incomplete vaccination program with 3 different vaccines (Pfizer (Pf), Astra-Zeneca (AZ) and a new, updated vaccine (N)). The novel strain is assumed to exhibit some resistance to both infection-acquired and vaccine-acquired immunity. We model the invasion and early transmission of this new strain using a multi-type branching process model [1].

We consider an epidemic in which the population is divided into 16 types, namely, 8 exposed types  $E_{d,v}$  and 8 infectious types  $I_{d,v}$  where  $d \in \{\text{sus}, \text{rec}\}$  denotes whether an individual is susceptible to the current wild-type strain of COVID or has been infected previously and has recovered, and where  $v \in \{\text{U}, \text{AZ}, \text{Pf}, \text{N}\}$  denotes whether an individual is unvaccinated or has been vaccinated with the Astra-Zeneca, Pfizer or new vaccine, respectively. We order the types in our model combinatorially so that types 1-8 correspond to exposed cases and types 9-16 correspond to infectious cases. Exposed cases become infectious at a constant rate  $\sigma$  and infectious cases have a constant recovery rate  $\gamma$ .

Each type of individual has a relative susceptibility to the VoC determined by their previous exposure to other variants and by their vaccination status. In our model this is interpreted as each infectious type having equal transmissibility  $\beta$  and exerting a force of infection on susceptible individuals of type  $j$  ( $1 \leq j \leq 8$ ) given by  $\beta_j = c_j \beta$ , where the constant  $c_j$  scales  $\beta$  by the relative proportions of type  $j$  in the population and by the relative susceptibility of each type to the VoC. At each infection event in our model an infectious individual is replaced by an exposed case of type  $j$  and an identical copy of itself. For convenience, we denote the lifetime of a particle of a case of type  $i$  by  $\omega_i = \sigma$  for  $1 \leq i \leq 8$  and by  $\omega_i = \sum_{j=0}^8 \beta_j + \gamma$  for  $9 \leq i \leq 16$ . The generating function for the offspring distribution of this process, the distribution of the number of cases of each type produced by a case of type  $i$  at a single event, is given by:

$$P_i(s) = \begin{cases} s_{i+8} & \text{for } 1 \leq i \leq 8 \\ \frac{\sum_{j=0}^8 \beta_j s_j s_i}{\omega_i} + \frac{\gamma}{\omega_i} & \text{for } 9 \leq i \leq 16 \end{cases} \quad (1)$$

where  $s = (s_i)_{i=1}^{16}$  is a vector of length 16 whose components correspond to each different type of infected individual. In our model, we initially consider a process that starts with a single case of type  $i$  at time  $t = 0$  and no other cases, and consider the total number of cases of all types at time  $t$ , which we denote by  $Y_i(t)$  for  $(1 \leq i \leq 16)$ . For each  $1 \leq i \leq 16$ , the random variable  $Y_i(t)$  has generating function  $Q_i(t, s) = \sum_{n=0}^{\infty} \Pr(Y_i(t) = n) s^n$ . Since  $s$  is a vector,  $s^n$  denotes the product  $s_1^{n_1} \cdots s_{16}^{n_{16}}$ , where  $n_i$  is the number of cases of type  $i$  and  $n = \sum_i n_i$ .

We obtain expressions for the  $Q_i(t)$  by solving:

$$\frac{\partial Q_i(t, s)}{\partial t} = -\omega_i [Q_i(t, s) - P_i(Q)], \quad Q(0, s) = s \quad (2)$$

where  $Q(t, s) = [Q_i(t, s)]_{i=1}^{16}$  is a vector. Setting  $s = 0$  gives the vector  $q(t) = Q(0, s)$  of extinction probabilities for a process starting with a single particle of type  $i$ .

We then consider a process that begins with no cases of any type, but that allows immigration of exposed cases of type  $i$  at a rate  $\eta_i$  and write  $\eta = \sum_i \eta_i$ . The total number of particles of all types at time  $t$  for this process is denoted by  $Z(t)$  and has the generating function  $R(t, s) = \sum_{n=0}^{\infty} \Pr(Z(t) = n) s^n$ , which we obtain by solving:

$$\frac{\partial R(t, s)}{\partial t} = -\eta R(t, s) + \sum_{i=0}^{16} \eta_i R(t, s) Q_i(t, s), \quad \text{subject to } R(0, s) = 1 \quad (3)$$

Setting  $s = 0$  as before, we find the probability of zero cases of any type at time  $r(t) = R(t, 0)$ . This should no longer be interpreted as a probability that a strain goes extinct, as immigration allows new cases to enter the population even when the total number of cases is zero. However, we can loosely interpret  $\lim_{t \rightarrow \infty} 1 - r(t)$  as the probability that a VoC is successfully established in the population and that the resulting epidemic grows exponentially.

## 0.2 Calculation of Means

The vector of the mean number of cases of each type at time  $t$ ,  $m(t) = \mathbb{E}[Z(t)] = (\mathbb{E}[Z_i(t)])_{i=1}^{16}$  is given by:

$$m(t) = m(0)e^{t\Omega} + \int_0^t \eta(\tau)e^{(t-\tau)\Omega} d\tau \quad (4)$$

where the matrix  $\Omega$  has  $(i, j)^{th}$  entry given by  $\Omega_{i,j} = \frac{\partial P_i}{\partial s_j}(\mathbf{1}) - \omega_i$ .

## 0.3 Calculation of Variances

In order to obtain the variance matrix for the random vector  $Z(t) = (Z_i(t))$ , we consider the variance matrix  $W_i(t)$  for an outbreak that begins with no cases and considers only immigration of cases of type  $i$ , which can be calculated via:

$$W_i(t) = \int_0^t \eta_i(t) \mathbb{E}[Y_T^* Y_T | T = \tau] d\tau \quad (5)$$

for  $1 \leq i \leq 16$ . If the matrix  $\Omega$  is diagonalisable, so that  $\Omega = ADA^{-1}$ , we have that:

$$W_i(t) = \eta_i H \Delta H^{-1} \text{Vec}(C) \quad (6)$$

where the matrix  $H = A \otimes (A^*)^{-1} \otimes (A^*)^{-1}$ ,  $\text{Vec}$  is the operator that vectorises the matrix  $C$  by stacking its columns.  $C$  is the  $r^2 \times r$  matrix consisting of block  $r \times r$  matrices  $C_i$ , given by:

$$C_i = \omega_i \left( \frac{\partial^2 P_i}{\partial s_j^2}(\mathbf{1}) + \text{diag}\left(\frac{\partial P_i}{\partial s_j}(\mathbf{1})\right) + e_i^* e_i - e_i^* \frac{\partial P_i}{\partial s_j}(\mathbf{1}) - \frac{\partial P_i}{\partial s_j}(\mathbf{1})^* e_i \right) \quad (7)$$

In 6,  $\Delta$  is the diagonal matrix whose  $(i, j, k)^{th}$  entry is given by:

$$\frac{1}{\delta_i - (\delta_j + \delta_k)} [e^{\delta_i t} - e^{(\delta_j + \delta_k)t}] \quad (8)$$

where  $\delta_i$  is the  $i^{th}$  diagonal entry of the matrix  $D$ .

Finally, the variance-covariance matrix of the random vector  $Z(t) = (Z_i(t))$  is given by

$$\text{Var}[Z(t)] = \sum_{i=1}^{16} W_i(t) \quad (9)$$

$$\Rightarrow \text{Var}\left[\sum_{i=1}^{16} Z_i(t)\right] = \sum_{j=1}^{16} \sum_{k=1}^{16} \text{Var}[Z(t)] \quad (10)$$

## 0.4 Consideration of Deterministic Limit

We finally consider the number of cases required for the ratio of the mean number of cases to the standard deviation to be constant over time. we interpret this quantity as the width of the confidence region around the mean number of cases over time, which after a certain limit should grow proportionally with the mean curve. For different scenarios, we plot the mean number of cases over time along with the standard deviation divided by the mean (see Figure 1), and note that the latter quantity reaches a steady state in all cases by the time 100 cases have been infected, on average. In our main analysis, we seed our deterministic models with 2,000 infected cases on the 17th May 2021, which we can be sure satisfies the necessary conditions for a deterministic approximation to be valid.

## References

- [1] Karin S Dorman, Janet S Sinsheimer, and Kenneth Lange. In the garden of branching processes. *SIAM review*, 46(2):202–229, 2004.

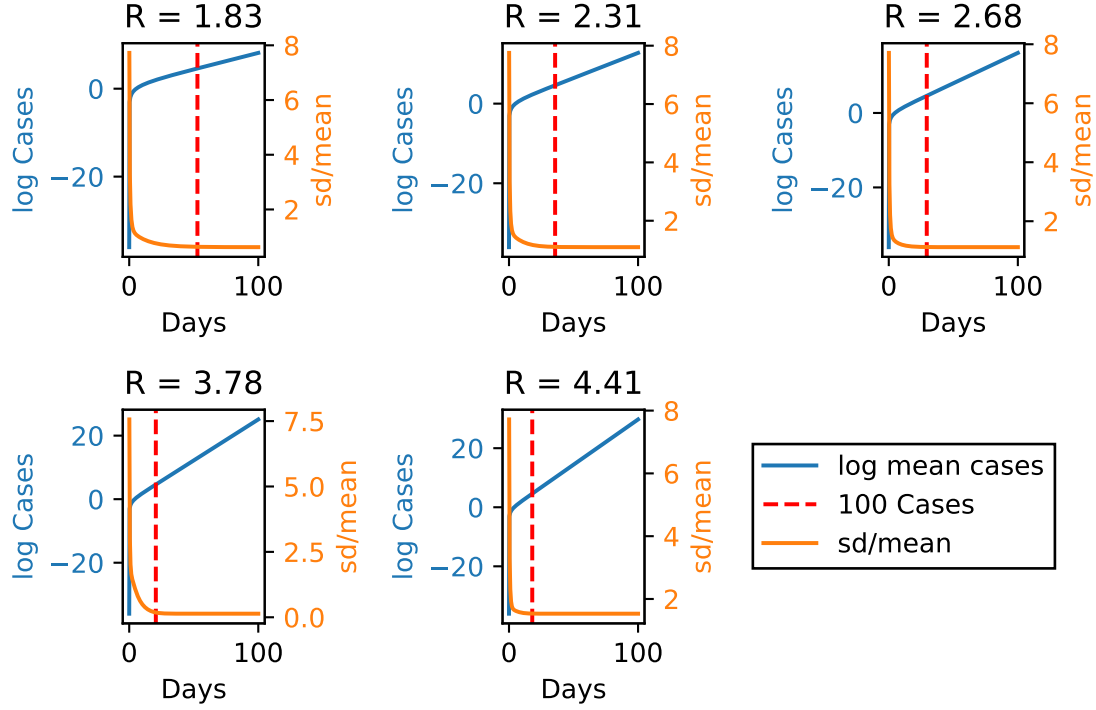


Figure 1: Plot of the log of the expected number of cases over time (blue curve) for different values of  $R$  for the invading VoC, as well as the standard deviation of the number of cases divided by the mean (orange curve). By the time the expected number of cases has reached 100 (dashed red line), the gradient of the standard deviation divided by the mean is approximately zero. Once this has happened, we assume that the number of cases is approximately normally distributed over time, centred on the mean, and hence a deterministic approximation of the epidemic is valid.

Parameter	Interpretation	Value
$\sigma$	Transition rate between latent exposed and infectious period (days <sup>-1</sup> )	0.3
$\gamma$	Recovery rate (days <sup>-1</sup> )	0.4
$R_{eff}$	Effective reproduction number of resident variant	[1.22, 2.94]
$c_{\text{var}}$	Relative transmissibility of VoC compared with resident variant	[0.8, 1.5]
$\beta$	Infectiousness of an infectious case with VoC	$c_{\text{var}} \times R_{eff} \times \gamma$
$\eta$	Importation rate (days <sup>-1</sup> )	[0.02, 0.4]

Table 1: Parameter choices for stochastic invasion model.