Controlling Nipah Virus in Bangladesh and India

Alexander Kaye

August 2019

1 Abstract

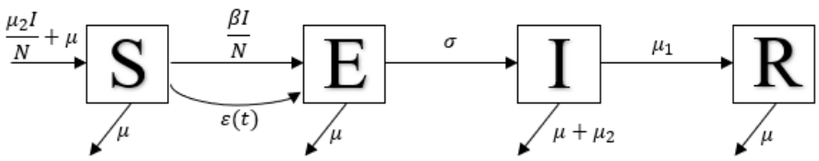
Nipah virus is an emerging tropical disease with annual outbreaks occurring in Asia, mostly in Bangladesh. The virus is carried by bats and then transmitted to humans as the bats eat raw date palms and transfer their urine and saliva to the sap which is then drank as a delicacy. There is also evidence of human to human transmission after the initial spill over event from the bat population. No vaccine currently exists but there is ongoing research into creating one. We aim to look at how efficacy and administration rate of such a vaccine would have limited the spread of the disease during the 2004 Faridpur, Bangladesh outbreak.

We begin by constructing a model that describes how the virus spreads throughout the human population on a village level. We then fit this model to reality by using outbreak data from Faridpur 2004. Using this model we are then able to explore how vaccination can reduce the expected size of an outbreak, we do this by looking at how many people per day we need to vaccinate and the required probability of success of such a vaccine in order to sufficiently limit the spread of the disease.

By going through this process we discover that we can make a sizeable reduction in the expected outbreak size even with a poor vaccine efficacy, provided we vaccinate enough people each day. However to reduce the expected infection size significantly a very effective vaccine is required.

2 The Model

We are using an SEIR model with demography, it is assumed that those who recover from the disease are not immune and so return to the susceptible class. When people die from the disease they enter the 'removed' class and it is assumed that when this happens, a birth also occurs to ensure the population (S+E+I+R) remains constant. A schematic of the can be found below.



ϵ is a seasonal forcing term representing transmission from bats to humans. It is assumed that ϵ is a step function that is non-zero from the start of the date palm collection season to the end of the season as this is the time in which all spillover events are observed and outbreaks start, we work out this period later. An alternative model is available which includes a hospitalised class but the results aren't as accurate as this model.

3 ABC Algorithm

An adaptive approximate Bayesian computation algorithm is used in order to fit the five free parameters in our model (β, ϵ, σ, the season start time and the season end time) to match the case data. We don't take μ_1 or μ_2 to be free as there isn't enough data to determine μ_2 and thus we fix it at $\frac{1}{16}$ as the mean duration of illness from onset of symptoms to death is around 16 days see (https://jcm.asm.org/content/56/6/e01875-17). We can also determine μ_1 from this as we know the fatality rate for each outbreak.

We assume each parameter is uniformly distributed over a sensible range and use these distributions as priors. We then sample from each of the priors and run the model (using a Gillespie algorithm) with these parameters. Supposing that i people are observed to be infected on the first day, the initial conditions for the model are $S = N - i$, $I = i$, $R = 0$ and E is selected from a uniform prior between 0 and the total number of infections inclusive (E is also treated as a free parameter and fitted using the algorithm). The summary statistics we use for comparison are cumulative number of infections over the outbreak. The error between case data and simulated data is then measured using the ℓ_2 metric.

For an outbreak occurring over n days we use the case data to come up with a cumulative infection vector; $\underline{i_c} = (i_{c_1}, i_{c_2}, ..., i_{c_n})$. Where $i_{c_i} - i_{c_{i-1}}$ is the number of deaths observed on day $i \in \{2, ...n\}$ and i_{c_1} is the number of infected people on the first day. After each simulation we generate a simulated cumulative infection vecto; $\underline{i_s} = (i_{s_1}, i_{s_2}, ..., i_{s_n})$. We then define the error, e between the case data and the simulated data as follows:

$$e := \| \underline{i_c} - \underline{i_s} \|_{\ell_2} = (\sum_{k=1}^{n} | i_{c_k} - i_{s_k} |^2)^{1/2} \tag{1}$$

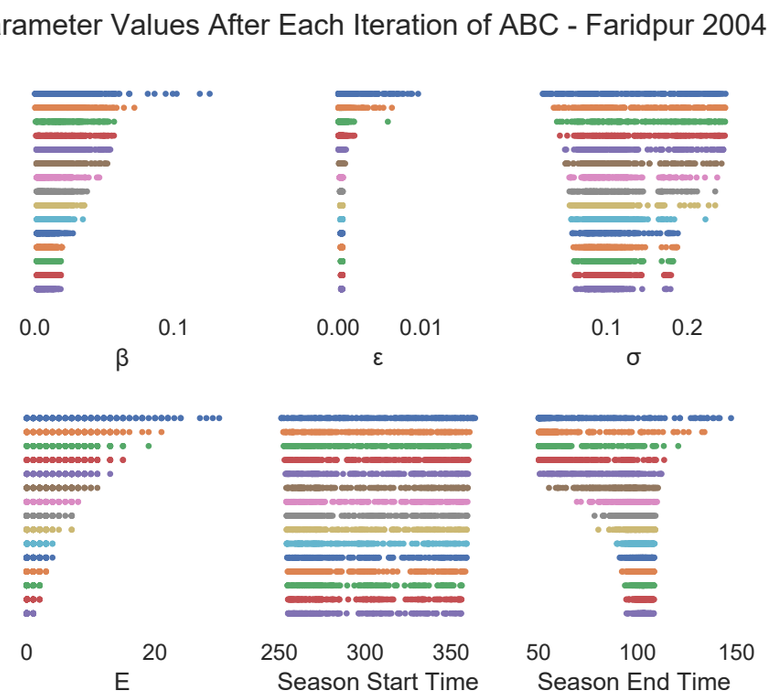
If the error is less than some initial tolerance τ we accept the parameters and store them, also storing the value of e for this simulation. We repeat this process, in this case 200 times, and plot a histogram for each set of parameters. We then repeat the algorithm again, this time the prior we use for each parameter is the normalised histogram we generated for that parameter. We also change the tolerance τ to be the median of all the values of e that were generated. This process is then repeated until the average error is below an acceptable amount.

4 ABC Results

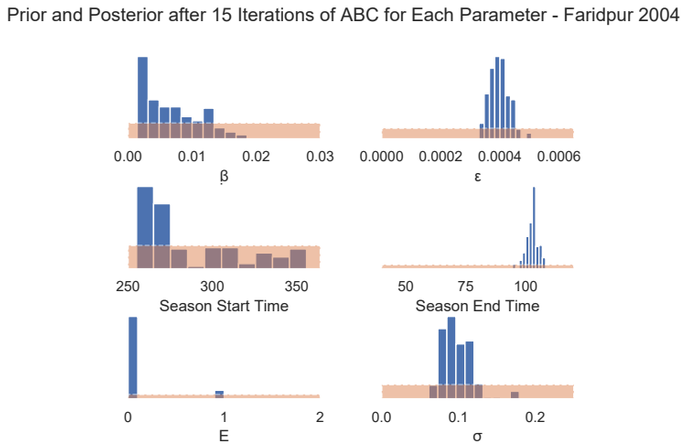
We perform ABC on the data set from the 2004 outbreak in Faridpur, Bangladesh using the following data set and a total population size of 1500. There isn't enough data to get an accurate distribution for the season start time as the Faridpur outbreak occurs near the beginning of the year, we need to look at more outbreaks to get a better distribution.

Date	Cumulative Number of Cases
19/02/04	1
04/03/04	2
06/03/04	3
09/03/04	4
16/03/04	5
19/03/04	6
24/03/04	7
25/03/04	9
26/03/04	11
29/03/04	12
30/03/04	14
31/03/04	17
01/04/04	26
02/04/04	27
04/04/04	28
12/04/04	31
13/04/04	33
14/04/04	34
15/04/04	35
16/04/04	36

There were 28 deaths in total. The following are the parameter values after each iteration of ABC:



We also include plots for the prior and posterior for each parameter



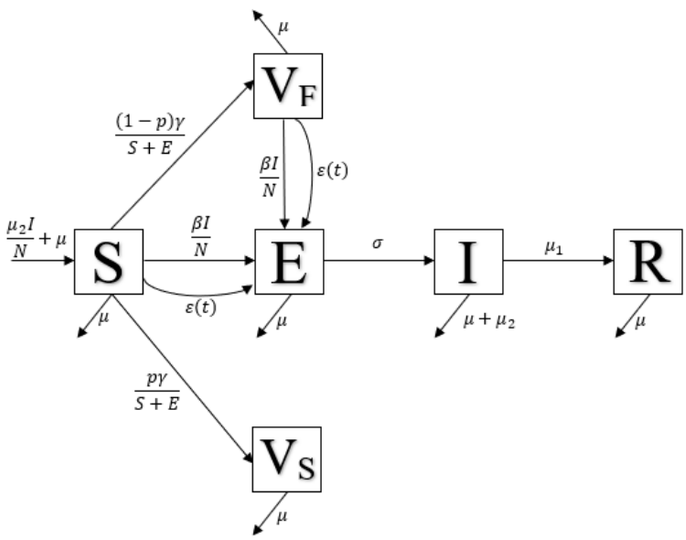
5 Forward Simulating

For each simulation we sample parameters from the posterior distributions generated by ABC and then calculate total number of infections from bats and from humans and total number of deaths. This is repeated many times (in this case 10,000) and an average is taken. We wish to see which model best replicates real life. We only accept simulations where the outbreak takes off, we define this as an outbreak where at least one new infection occurs. Average number of total infections is 35.72
Average number of total infections from bats is 31.93
Average number of total infections from humans is 3.79
Standard deviation of number of total infections is 7.94
This means roughly 10.61% of transmissions are human to human.

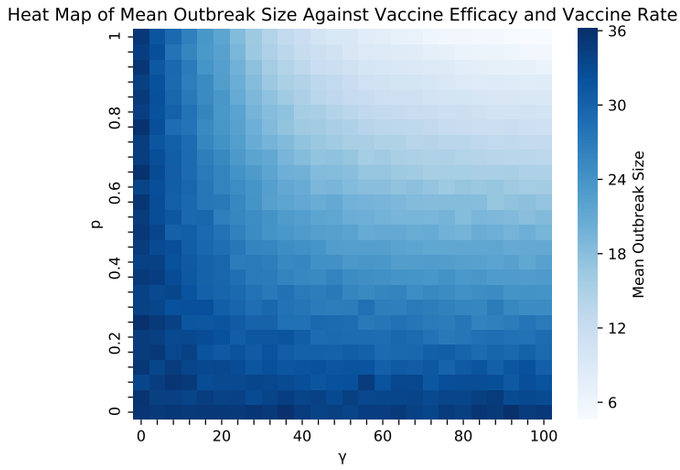
6 Control

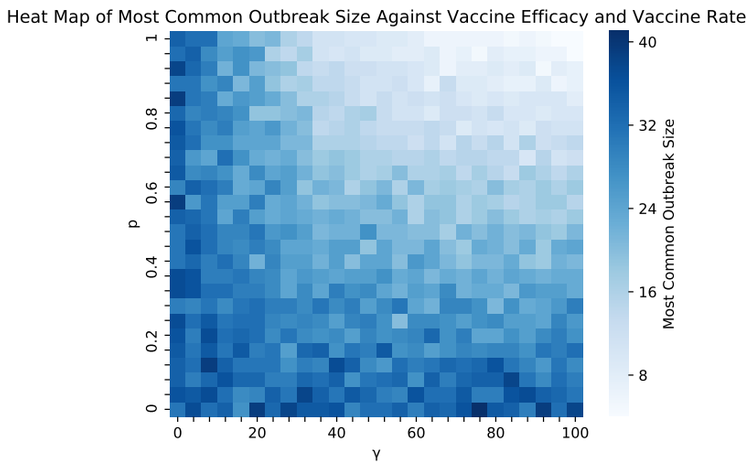
6.1 Vaccination on First Infection

Since our model is a start of outbreak model we can begin vaccination as soon as the simulation starts. Vaccination occurs at a rate of γ people per day, it is successful with probability p and vaccination is done randomly to the susceptible and exposed population. People who are successfully vaccinated move into the 'V_S' class and people who are unsuccessfully vaccinated move into the V_F class where they are still able to pick up the disease and transmit it. The schematic of this process is shown below (where $N = S + E + I + R + V_S + V_F$):



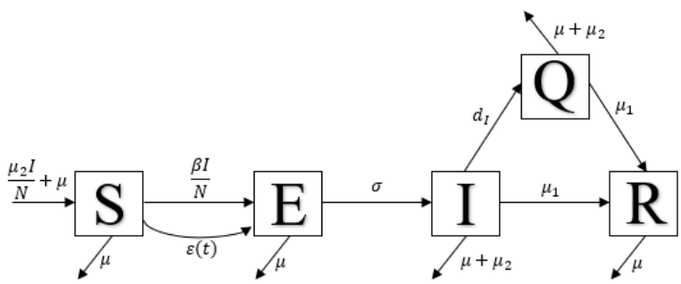
Below is a heat map of number of vaccinations per day against total number of infections using parameter data for the 2004 outbreak in Faridpur:





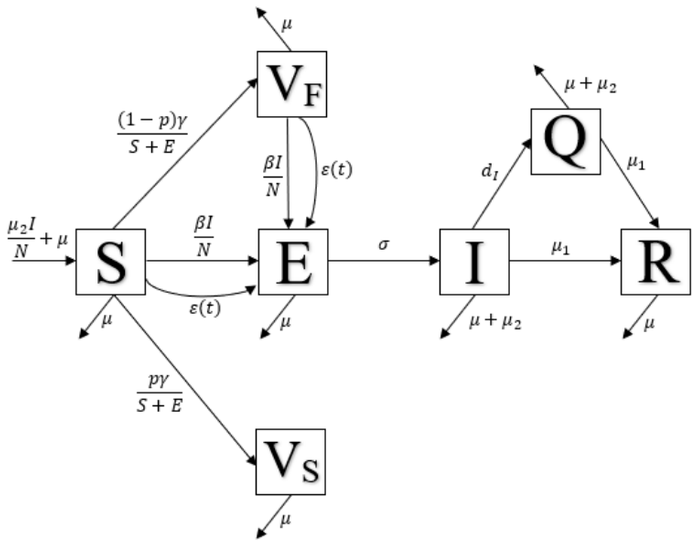
6.2 Quarantine

Here we consider moving infectious people into isolation. It is assumed that we detect infectious people and transfer them to the quarantine class Q at a rate d_I. Once in quarantine they are released upon recovery which occurs at a rate μ_1 as usual. The schematic of this process is shown below (where $N = S + E + I + R + Q$):



6.3 Vaccination on First Infection and Quarantine

We combine the two control strategies above, i.e. vaccinate the susceptible and exposed population randomly as soon as the first infection is observed and move infectious individuals into quarantine. The schematic of this process is shown below (where $N = S + E + I + R + V_S + V_F + Q$):



7 Parameter Fitting over all Outbreaks

There have been various outbreaks of Nipah virus, most of which occurring in Bangladesh. Our aim is to find a set of parameters which are able to describe all significant outbreaks of the virus. The period in which the seasonal forcing term is active coincides with the raw date palm collection period, but this varies place to place. To account for this we include season start time and season end time as free parameters in ABC and fit them to all outbreaks. We also need to know the population for each town in which the outbreaks occurs, however this data is very difficult to find so we take the population for each outbreak as 1500. To do this, we sample from our prior distributions and then simulate using these parameters separately for each outbreak. The error for each outbreak is calculated as detailed previously, and then we take the average error to work out the total error which will be the acceptance threshold for use in ABC. If the total number of infections for the nth outbreak is I_{n_T} then we choose the initial number of exposed individuals from a discrete uniform distribution between 0 and I_{n_T} inclusive and we fit the initial number of exposed to each outbreak. We also choose the initial number of infectious individuals to be the number of infectious individuals first observed for each outbreak.