# ABSynthe description – EVD for Sierra Leone

This agent based model simulates an EVD outbreak in Sierra Leone, with three explicit contact levels. It also simulates a coalescent phylogeny for the outbreak.

## Contact structure

An underlying contact structure is defined and is fixed between simulation runs. Each individual is put into a household, a chiefdom and a district.

Each district has a different number of individuals, households and average household sizes. Average household size was determined separately for each district by dividing the district population by the number of households. This number was then rounded so that a complete number of people were in each household.

By using chiefdom data, there were 2.5% fewer people than using the total census population ie 2.5% were unaccounted for in the 2014 census. The total population was taken as the total in chiefdoms.

[More detail about exactly how many people, as well as average Hh size]

## Infection model

There are two dictionaries which describe the infection algorithm:

1. Day dictionary: day of outbreak as keys, with list of cases which become infected on that day [NB not symptomatic].
2. Case dictionary: links case objects to individual objects

The outbreak begins with a random individual in Kissi Teng chiefdom in Kailahun, as the real outbreak began here. To mimic the funeral which kickstarted the EVD outbreak in Sierra Leone, fourteen secondary cases are infected. According to Wauquier et al 2015, two of these are at the household level, three are within chiefdom and two were found Kenema, a different district. For the seven unaccounted for cases, three were put in a different chiefdom in Kailahun (there were three by the 10th June), and the remaining four were placed within chiefdom.

Therefore the epidemic is seeded with one individual who infects 2 within their household, 7 in the same chiefdom, 3 in the same district, and 2 in a different district.

This event is assumed to have happened on the 5th of May. The exact date of the funeral is unclear, but it appears that it was in the first week of May 2014.

Each case is initialised as a member of the **individual** class. When this initialisation happens, all infection characteristics are set for that individual.

Then it is determined whether any secondary cases are infected, by using the binomial probabilities for each contact level described above. For each case, we initialise them as a member of the **case** class, assign them an arbitrary case ID and store what contact level they are relative to their parent. We then ask **when** they are infected based on their parent’s infection trajectory, and add them to the day dictionary on the appropriate day. If the day is after the parent has recovered or died, then the case is not initialised.

The loop through the day dictionary begins. For each new day, if there is a new case, then they must be assigned to a real person within the contact network. This is achieved by calling a function on the parent individual which determines **who** within the appropriate contact level is infected. This is a random choice of individuals at the appropriate level. If the contact level is the country level (i.e. in a different district), the district is selected based on distance to the parent’s district, with closer districts more likely to be selected than those further away in terms of population centroids. For other levels, it is purely random.

If a person is selected and they are still susceptible, then they are initialised as a member of the individual class, and the infection process continues. If they are already infected, then that case object has not been assigned to a real person and is flagged for removal at the end of the simulation run.

Therefore there are **case** objects and **individual** classes. The cases may or may not become real people in the outbreak, depending on whether or not there are susceptible individuals to infect. They only have an arbitrary ID and the contact level they hypothetically occupy. Individuals relate to a real person in the contact network, and have a defined course of infection.

## First inputs

* Population size (based on pre-determined social structure)
* Epidemic length (default: 2000 days)
* Case fatality rate (default: 0.7)
* Arrays containing the location in terms of household, community and district for every person in Sierra Leone.

## Infection parameters

* Incubation time:
  + Gamma distribution
  + Mean = 8.5 days
  + SD = 7.6 days
* Death time:
  + Gamma distribution
  + Mean = 8.6 days
  + SD = 6.9 days
* Recovery time:
  + Gamma distribution
  + Mean = 15.2 days
  + SD = 6.2 days
* Time to infection:
  + Gamma distribution
  + Mean = 3.1 days
  + SD = 2.5 days

All of the above are taken from (WHO Ebola Response team, 2014).

Recovery time is the time from symptom onset to hospital discharge, the only recovery time measure available. Furthermore, the WHO paper gives 17.2 day average, but an individual must receive two negative Ebola tests in order to be discharged, and so the true recovery time is two days earlier (following (Kucharski et al., 2015)).

The time to infection is taken as the mean serial interval (11.6 days) minus the mean incubation period (8.5 days). This gives us the mean, and depending on whether an individual survives or dies, the dimensions of the distribution are different.

Transmission

The probability of infecting someone at each level is a negative binomial process with different lambdas for each level.

They are all based on **β,** the within household lambda. Therefore we have **aβ, bβ,** and **cβ** for the chiefdom transmission, district transmission and country transmission respectively.

Within each level, the probability of infecting anyone else is constant. I.e. once the secondary case has been assigned a level, there is an equal chance of anyone at that level being infected. When someone in a different district is infected, the district they are in is chosen based on distance to the focal district.

Before fitting, **β = 0.08.** This is based on the result from Glynn et al. (2018), which said that 32.7% of people infected at least one person in their household, and so p(0) = 0.673. Working this backward through the poisson equation gives a lambda of 0.396. This lambda is used as a the mean for a gamma distribution, to give the negative binomial.

## Outputs

* For each logged run (every 100th run):
  + Information file (csv) with person, parent, household, district, day infected and date of symptom onset
  + Transmission tree
  + Dates of sampling
  + Coalescent phylogeny
  + Skyline plot
* For each run:
  + Average R0
  + Epidemic size

## Assumptions

* Only infectious after symptoms begin
* Sampling is 100% and occurs on first day of symptom onset
* Individuals stay infectious for seven days after death (Prescott et al., 2015).
* No interventions – i.e. no safe burials and no isolation of infective individuals.