Block 4:

The applications of modelling

Practical notes

Introduction to Infectious Disease Modelling and its Applications – 2018

Session 26: Calculating the relative contribution of recent transmission to tuberculosis disease incidence

Practical

Overview and Objectives

As you saw in the lecture, the natural history of tuberculosis is complicated by fact that disease can occur either soon after recent infection, or many years thereafter either through reactivation or following reinfection. This practical is designed to illustrate how you would set up a model of the transmission dynamics of *M tuberculosis,* assuming that disease can occur as a result of these three mechanisms. The practical is structured in two parts.

In part one, you will set up a model of the transmission dynamics of *M tuberculosis* in a population in which the annual risk of infection has not changed much over time. In part two, you will explore how some of the assumptions influence estimates of the proportion of disease attributable to recent transmission.

By the end of this practical, you should understand:
1. How a simple model describing the transmission dynamics of *M tuberculosis* might be set up using Excel
2. The relationship between the key input parameters in the model
3. How the relative sizes of the annual risk of infection and the different risks of developing disease influence the overall proportion of disease attributable to recent transmission.

PART I: Setting up a model describing the transmission dynamics of *M tuberculosis*

1. Open up the spreadsheet TBaricon.xls. This model is designed to explore the effect of different assumptions about the risks of developing disease following recent (primary) infection, reactivation and following reinfection in a setting in which the annual risk of infection with *M tuberculosis* has changed little over time, at a level similar to that in several developing countries today.

You should see some:
a) **Blue** cells which contain the current value for the annual risk of infection and the total population size. The annual risk of infection is assumed to have remained constant over time and is set to be 1%, similar to that estimated for a rural setting in Malawi during the 1980s (Fine et al (1998)).
b) **Pink** cells in which we will set up expressions (by following the steps below) for the prevalence of infection and the number of individuals who are newly infected, already infected and newly reinfected.

We first use the annual risk of infection to calculate the proportion of individuals assumed to be infected in this setting.

2. Using the fact that for a given annual risk of infection λ, the proportion of individuals who are infected by a given age a is given by

$$1-(1-\lambda)^a$$

set up a suitable expression for the proportion of individuals aged 0 years who are infected and copy this expression down until the oldest age group.

Q1.1 What proportion of individuals are already infected by age 10, 20, 40 and 60 years in this population?

3. Using the fact that the number of individuals who will be newly infected during the coming year is given by:

proportion who are not infected $\times$ risk of infection $\times$ population size

set up a suitable expression in cell C21 for the number of individuals aged 0 years in this population who will be newly infected during the coming year.

4. Similarly, in cells D21 and E21 set up suitable expressions for the number of individuals of age 0 years in this population who

i) are already infected
ii) will be reinfected during the coming year.

assuming that anyone who has been infected for at least one year can be reinfected. Copy your expressions down until the oldest age group.

Q1.2 How many individuals aged 0 and 10 years will be newly infected or reinfected during the coming year?

We will assume that the risk of developing disease depends on the current infection status of individuals, i.e. whether they have been:

i) recently infected
ii) infected for at least one year
iii) recently reinfected

These assumptions will be incorporated by following the steps below.

5. Select rows 7 and 14, click with right mouse button and choose the unhide option. You should now see some blue cells containing the risk of developing disease soon after infection, through reactivation or following reinfection for people aged <10 and over 20 years.

The current values for these disease risks have been derived by fitting a model describing the transmission dynamics of *M tuberculosis* to data from the UK (see main lecture notes).

6. Select columns E and K, click with the right mouse button and choose the unhide option.

You should now see some

a) **Blue** cells containing the proportion of cases at different ages who are infectious (defined as those who are sputum smear or culture positive). At present, 10%, 65% and 90% of cases aged 10, 20 and 100 years are assumed to be infectious.
b) **Yellow** cells containing equations for how the age-specific proportion of disease which is infectious and the risks of disease changes between age groups; the

contents of these cells are plotted in Figures 1 and 2. (Don't worry about these equations for now).

Q1.3 Using the contents of cells C21-E21, how might you calculate the number of people who will experience infectious disease as a result of recent infection, through reactivation or following reinfection during the coming year?

7. Use your answer to set up appropriate expressions for the number of new cases per 100,000 population in cells L21-O21 during the coming year among individuals aged 0 years. Copy your expressions down until the oldest age group.

Q1.4 How many individuals aged 10, 20, 40 and 60 years are predicted to experience these different forms of disease in this setting?

We will now explore how the proportion of disease attributable to recent transmission changes with age according to this model.

PART II Analysing the proportion of disease attributable to recent transmission

Q2.1 According to Figure 4 in the spreadsheet, how does the proportion of disease attributable to recent infection, reactivation and reinfection change with increasing age? Is this what you would expect? Why?

1. Select columns Q and W, click with the right mouse button and choose the unhide option. In cells R21-U21 set up suitable expressions for the proportion of disease among individuals aged 0 years which is attributable to recent infection, reactivation, reinfection, and recent transmission (i.e. recent infection and reinfection combined). Copy your expressions down until the oldest age group.

Q2.2 According to Figure 5, how does the proportion of disease attributable to recent transmission change with increasing age? How do you think this age pattern would change if we were to assume that the annual risk of infection remained unchanged at:

a) 0.1% pa
b) 3% pa?
c) 10% pa?

Why? Test your hypotheses by amending the current value for the annual risk of infection in the model.

Q2.3 Given the relative contribution of the disease attributable to recent transmission, should it be easy or difficult to control disease in the population by interrupting transmission eg through introducing effective treatment?

Select columns AF and AN, click with the right mouse button and choose the unhide option. You should now see that Figure 4 now contrasts model predictions against incidence of bacteriologically-confirmed ("infectious") TB from a rural setting in Malawi (contained in columns AG-AL.

Q2.4 What do you notice about the fit of the model to the data? What might you change to improve the fit?

In your spare time, you may like to think about how you would need to change the model if you wanted to assume that the annual risk of infection changed over time, and to incorporate contact between individuals…

If you have time, try the supplementary questions (see the supplementary questions folder on Moodle or in the network folder containing the model files).

References:

Fine PEM, Bruce J, Ponnighaus JM, Nkhosa P, Harawa A, Vynnycky E Tuberculin sensitivity: conversions and reversions in a rural African population **Int J Tuberc Lung Dis** 3:962-75

Introduction to Infectious Disease Modelling and its Applications – 2018

Session 27: Application of models to veterinary epidemiology: spatial models

Practical

Overview and Objectives

In this practical you are going to develop your own model for an outbreak of highly pathogenic avian influenza (HPAI) in the commercial poultry sector in Germlandia. The first part of the practical will be "pen and paper" in which you need to consider the available information and work out which parts are needed to set-up and parameterise your model. You can then proceed to use the model which has been set up. In the final section you are asked to explore some interventions.

By the end of the practical you should

- Be able to extract parameters from real data for veterinary models
- Understand the structure of models for veterinary outbreaks
- Be familiar with setting up a simple meta-population model
- Understand how to incorporate common interventions

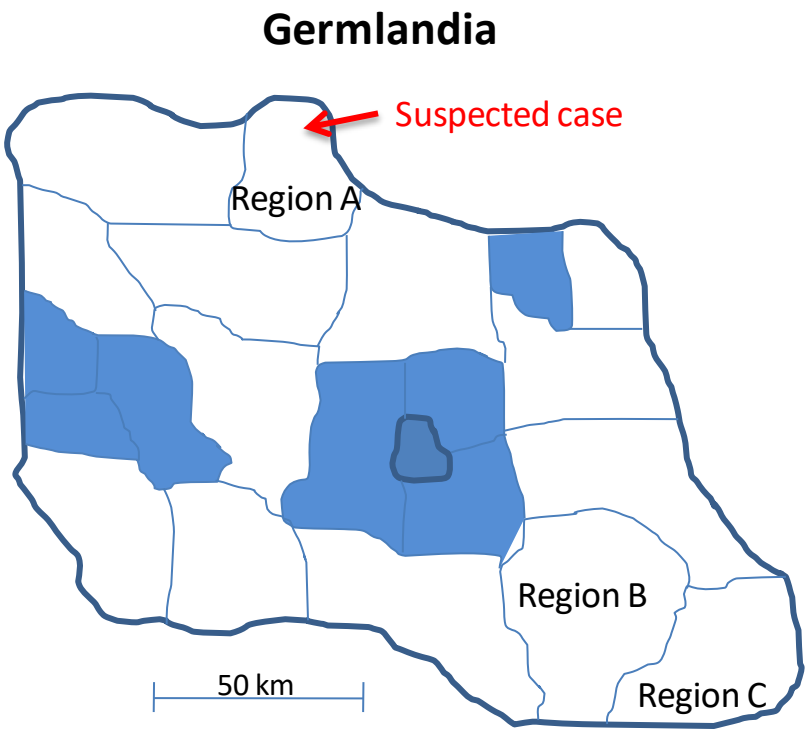
Part I: Model Development

Birds are a natural host for influenza viruses. In this host, a range of different subtypes of influenza viruses circulate. Such viruses are generally classified through the molecular subtype which is defined in terms of the haemagglutinin (H) and neuraminidase (N) outer membrane proteins. There are currently 16 subtypes of H (labelled H1…H16) and 9 subtypes of N (similarly labelled N1…N9). All H and N subtypes have been identified in birds and some are likely to be endemic in other parts of the world. Avian influenza is generally also further classified into low and high pathogenicity viruses and this classification overlaps with the HN classification. Low pathogenicity influenza viruses may not cause symptoms in birds but can result, for example, in reduced egg laying. High pathogenicity influenza viruses usually kill the host within 24 to 48 hours of clinical signs becoming apparent.

Germlandia is a Central European country bordering Poultrova on its North, Hensylvania to the West and the Chick Republic to the South and East. It has 22 administrative regions (including the capital city). At the current time Germlandia has just reported the first suspected case of HPAI avian influenza in poultry. This case has occurred on a farm in the far north of the country, in Region A (see map). Over the last 6 days 90% of the chickens in the affected flock have died. The case was notified today, and samples have been sent to the Virus Reference Laboratory for testing. As yet, no results are available, but HPAI is strongly suspected. The remaining chickens in the affected flock are being culled. The Ministry of Agriculture is considering policies for containing the outbreak, and they require your input on which policies may be effective. No control policies have been implemented yet, and it is unlikely that any will be before laboratory confirmation (expected within the next 48 hours). Possible policies under consideration include

culling of affected flocks, a transport ban of poultry and poultry products, the institution of hygienic measures, and culling of flocks in the vicinity. Germlandia has no stocks of vaccine.

There are ~3000 commercial flocks in Germlandia, almost all of which are located in three regions; A, B and C (see Map). In each of the regions there are about 1000 flocks (there are <100 commercial flocks elsewhere in the country). The average number of birds in a flock is 10,000 and this does not differ by geographical location. The Ministry of Agriculture does not know the exact location of all of the commercial flocks in the country, although they are working on getting this information together as rapidly as possible. These separate geographical locations have links between them (through shared slaughterhouses and breeders). There are a few "back-yard" chickens in the 3 regions and elsewhere in the country, though they are very rare in the major urban areas (shown shaded on the map). It is about 150km from the mid-point (centroid) of Region A to Region B, and a further 50km to the midpoint of Region C (although rail and road links are actually better between Region A and C than Region A and B). Broiler birds live on average for 6 weeks before being slaughtered for consumption and typically stay on the farm in which they were reared. Laying hens live, on average, for about 1.5 years. About 20% of the flocks are laying hens. Most (80%) of the laying hen flocks are located in Region A, the rest being evenly split between Regions B and C (there are very few in other parts of the country).



Analysis of the 2003 H7N7 HPAI outbreak in the Netherlands [1,2] suggested that the risk of transmission fell rapidly with increasing distance between farms. Boender et al. [1] fitted a logistic function ($h(r) = h_0/[1+(r/r_0)^{\alpha}]$, where $h(r)$ is the transmission kernel describing how the risk of transmission is related to the distance, r, between farms). The best fitting estimates they derived for parameters h_0, r_0, α, where 0.002 (per day) 1.9 (per km) and 2.1 respectively, suggesting that the risk of transmission at 2 km distance is about 50% lower than at 0km distance, and that the risk at 10km is only 3% of the risk at 0km. The corresponding figures for

25, 50, 100, and 150km are only 0.44%, 0.1%, 0.02% and 0.01% respectively, though there is considerable uncertainty about these numbers, and estimates for the Netherlands in 2003 may not apply in Germlandia today. Stegeman et al. [2] estimated that the transmission parameter b was around $0.4*10^{-3}$ per flock per day before control measures were instituted and around $0.17*10^{-3}$ per flock per day after control measures were introduced. These measures included a transport ban on poultry and poultry products, tracing of dangerous contacts, institution of hygienic measures and culling of infected flocks. Again, it is not clear how generalisable these parameter values are.

We know that the incubation period for HPAI in a bird is approximately 2 days. Birds then onset with symptoms and that following onset of clinical signs the birds die within about 4 days. An infected flock is thought to be not infectious for about 2 days. After this a noticeable number of birds will die and the owner will contact a vet. It is estimated that from the time in which deaths are noticed in the flock it will take 2 days before infection is confirmed in the laboratory and a further 2 days from confirmation of infection to isolation of the flock when culling of birds will begin.

1. Using this information formulate a simple model to explore transmission of HPAI in the commercial poultry population in Germlandia.

Points to consider:

a) What is the appropriate unit to study?
b) What is the appropriate disease pathogenesis model (e.g. SI, SIR etc)
c) What detail of the population needs to be included in the model structure to answer the questions being posed?
d) Will you need to consider a spatial model, and if so, what structure will you adopt?
e) What parameters are known and which will you need to generate?
f) Should the model be stochastic?
g) What are the initial conditions?
h) Is transmission likely to be frequency or density dependent?

Note: Not all the information given needs to go into the model!

Part II: Model Analysis

Now open the file “Avian_Flu_Meta_Pop.mmd”. This contains a version of the model that you might have formulated. Check first that you understand how it works and whether it is similar to the model that you have written.

2. Some of the parameter values are missing from this version of the model and you will need to add appropriate values.

3. Implement your favoured transmission matrix, either by parameterising a pre-existing structure, or writing your own piece of code. Explore the behaviour of the model as systematically as time allows. Answer the following questions:

3a. When is the epidemic likely to peak?
3b. How many flocks are likely to be culled?

3c. How important are these findings to the proportion of within-region contact?

Part III: Impact of control measures

4. How effective is movement restrictions at limiting the epidemic size?
- *4a. Within the first-affected area?*
- *4b. Within the country as a whole?*
- *4c. How sensitive are your results to the initial conditions*

5. What other control policies could be implemented (in practice, and in theory – i.e. if Germlandia had been better prepared, for instance)? Implement a range of them in the model. How effective are they?

6. Are there any restrictions on the sort of control policies that you can look into that arise from the model structure? If so, how would you re-structure your model, and would you need any further data?

References

[1] Boender GJ, Hagenaars TJ, Bouma A, Nodelijk G, Elbers AR, de Jong MC, van Boven M. Risk maps for the spread of highly pathogenic avian influenza in poultry. PLoS Comput Biol. 2007 Apr 20;3(4):e71.

[2] Stegeman A, Bouma A, Elbers AR, de Jong MC, Nodelijk G, de Klerk F, Koch G, van Boven M. Avian influenza A virus (H7N7) epidemic in The Netherlands in 2003: course of the epidemic and effectiveness of control measures. J Infect Dis. 2004 Dec 15;190(12):2088-95.

Introduction to Infectious Disease Modelling and its Applications – 2018

Session 28: Sexually transmitted infections model

Practical

1 Objectives

In this practical we will set up a deterministic model of the dynamics of the curable STI *Neisseria gonorrhoea* (gonorrhoea).

- We will show that in the absence of heterogeneity in risk behaviour, unrealistic parameter values have to be assumed for gonorrhoea to become endemic.
- We will extend this model to allow us to model heterogeneity in sexual activity and use this model to show that risk-heterogeneity allows gonorrhoea to invade the population assuming much more realistic rates of partner change.
- We will extend this model to allow us to vary mixing pattern between the activity groups and use this model to explore the effect of changing the mixing pattern on gonorrhoea prevalence trends
- Finally, we will use this model to explore the likely impact of screening on equilibrium gonorrhoea prevalence

By the end of this practical you should

- Be able to set up simple models of a curable STI that assume:
    - Homogenous sexual activity and proportionate mixing
    - Heterogeneous sexual activity and proportionate mixing
    - Heterogeneous sexual activity and with-unlike, proportionate or with-like mixing
- Appreciate the importance of heterogeneity in risk behaviour for STI endemicity
- Appreciate the importance of mixing patterns on STI prevalence
- Appreciate some of the pros and cons of untargeted and targeted screening strategies for controlling curable STIs

2 A transmission model of gonorrhoea assuming homogenous sexual activity

The model we are going to set up is based on work done by *Hethcote et al* in the early 1980s. In the early 1970s, a gonorrhoea screening programme was introduced in the US to identify asymptomatic women by doing culture testing of as many women as possible. The work by Hethcote and colleagues explored this strategy and compared it to alternatives.

As we saw in the lecture, the simplest model ignores the differences in the natural history of gonorrhoea, men and women and risk behaviours. The population is split into the number of susceptible (S) and infectious and infected (I) individuals,

The equations for this model are:

$$\frac{dS(t)}{dt} = -\lambda(t)S(t) + rI(t)$$

$$\frac{dI(t)}{dt} = +\lambda(t)S(t) - rI(t)$$

The force of infection $\lambda(t) = c\ \beta_p\ p$.

The probability that partners are infectious = prevalence of infectious individuals, p = $\frac{I(t)}{N}$

The parameters are:

D = duration of infection (years)
c = partner change rate (partners/year)
β = transmission probability per (discordant) partnerships
N = population size

Qu 2-1 Sketch the flow diagram for this model in the space below:

Parameter description	**Value**
Population size (N)	20000000
Partner change rate, per year (c)	2
Transmission probability in discordant partnerships (β)	0.75
Duration of infection, years (D)	0.167 (2 months)

Qu 2-2 Using the formula in your lecture notes (Equation 1.6), calculate the basic reproduction number of gonorrhoea.

R_0 = _____

Qu 2-3 As such, what do you expect to happen when you run the model?

_____

Either set up this model from scratch in Berkeley Madonna or open model file '*STI_1.mmd*' or '*STI_1 flowchart.mmd*' to check your prediction.

Set the STOPTIME to be 100 (years). Don't forget to assume one individual is initially infected!

Plot the prevalence of infection over time.

Qu 2-4 Using your model, predict how high the partner change rate needs to be for transmission of gonorrhoea to persist in the population _____. You can best do this using a parameter plot.

Qu 2-5 Is this annual rate of partner change plausible for the general population?

_____

Qu 2-6 What reasons could there be for our failure to explain the spread of gonorrhoea?

3 A transmission model of gonorrhoea assuming heterogeneity in sexual activity

We now assume that the sexually active population is made up into two risk groups, a high activity group (given the subscript *H*) and a low-activity group (subscript *L*).

The equations for the model are:

$$\frac{dS_j(t)}{dt} = -\lambda_j(t)S_j(t) + rI_j(t)$$

$$\frac{dI_j(t)}{dt} = +\lambda_j(t)S_j(t) - rI_j(t)$$

Where $j = H$ for the high-activity group and $j = L$ for the low-activity group

The equation for the force of infection is

$$\lambda_j(t) = c_j \beta_p p$$

Where $j = H$ for the high-activity group and $j = L$ for the low-activity group.

The parameters are:

D = duration of infection (years)
c_H = partner change rate of high-activity members (partners/year)
c_L = partner change rate of low-activity members (partners/year)
β = transmission probability per (discordant) partnership
N_H = population size of the high-activity group
N_L = population size of the low-activity group
p, the probability that a selected partner is infectious

$i_H(t)$ = prevalence of infection in the high-activity group = $I_H(t)/N_H$
$i_L(t)$ = prevalence of infection in the low-activity group = $I_L(t)/N_L$

NOTE that lowercase '*i*' is used to denote the prevalence of infectious individuals and uppercase '*I*' is used to denote the number of infectious individuals.

Qu 3-1 Sketch the flow diagram for this model in the space below:

This model splitting the population into two activity groups has been set up for you.
Have a look at the way the model works by examining four key formulas:

Open the model file '*STI_2.mmd' or* '*STI_2 flowchart.mmd'.*

Formulae 1 and 2:
Look in the section "*Useful calculations".*
For now we assume proportionate mixing, ie the probability that a newly selected partner will be a member of the low (g_L) or the high (g_H) activity group is just the proportion of partnerships offered by each of these two groups (using Equation 1.53 shown in Panel 1-4 in the lecture notes):

$$g_H = \frac{c_H N_H}{c_H N_H + c_L N_L}$$

As the probability of selecting either a high or low-activity group member must sum to one:

$$g_L = 1 - g_H$$

From these two formulas, you can see that g_H and g_L will change as you change the ratio of the partner change rates in the high and low activity groups

Formula 3:
Again look in the section "*Useful calculations".*
The probability that a new partner is infectious, *p*, is not the overall prevalence of infectious individuals in the population ($I(t)/N$) as it was for the simple model, but now needs to account for the fact that the selected partner can be either a member of the high or low-activity group, and the prevalence of infectious individuals is likely to differ between these two groups (Equation 1-10 in the lecture notes):

$$p = g_H \times i_H(t) + g_L \times i_L(t)$$

where g_H and g_L are the probabilities that a selected partner is a member of the high or low-activity group, respectively, and $i_H(t)$ is prevalence of infection in the high-activity group and $i_L(t)$ is the prevalence of infection in the low-activity group.

Formula 4:
So we can make a valid comparison, we want to explore how the heterogeneity affects the spread of gonorrhoea, but we want to keep the average partner change rate constant.

We can do this by setting the mean partner change rate (c_{mean}) of the whole population equal to 2 partners per year, and add a formula for c_H so that the partner change rate of high-activity group members depends the rate of low-activity members (Equation 1.51 in Panel 1-3 in the lecture notes):

$$c_H = \frac{c_{mean} N - c_L N_L}{N_H}$$

You can see this formula in the section *"Initial conditions and parameters".*

The model is ready to run!

When you run you model you should see that you still don't see sustained transmission of gonorrhoea because you are still assuming that both groups have 2 partners per year.

Now using the slider, reduce the partner change rate among low activity individuals while watching the partner change rate in the high activity group and the prevalence of infection in the low and high activity groups.

Qu 3-2 What happens to the partner change rate in the high activity group as you lower the rate in the low activity group?

_____

Qu 3-3 At what rate of partner change in the low-activity group does gonorrhoea prevalence rise?

_____

Qu 3-4 What is the partner change rate in the high-activity group when this occurs?

_____

Set the partner change rate in the low-activity group to 1.4 partners per year for the rest of this section.

Qu 3-5 Using your model or Excel or a calculator, calculate the fraction of all partnerships provided by the high and low activity groups in this situation?

g_H = _____

g_L = _____

Qu 3-6 What is the equilibrium prevalence of gonorrhoea in the low-activity group, high-activity group and overall in this situation?

_____

Qu 3-7 Is the overall prevalence consistent with the ~2% prevalence seen in the sexually active population in the US?

_____

Qu 3-8 Is this partner change rate in the low-activity group more plausible than the rate we had to assume to get gonorrhoea to invade if we did not model heterogeneity?

_____

You can better see how the prevalence of gonorrhoea changes with the heterogeneity in the number of partners by running the *Parameter Plot* that has been set up for you in the model.

Qu 3-9 Run the parameter plot and describe the graph. What happens to the prevalence of infection in the low-activity group, high-activity group and overall as we increase heterogeneity in sexual activity between the two groups?

4 The impact of changes in mixing by sexual-activity on gonorrhoea prevalence

Open model file '*STI_3a.mmd*'.

The model allows you to vary the mixing pattern between the two sexual activity groups in the model. As we saw in the lecture *(Section 1.5.2)*, the *degree of mixing* can be summarised by *Q.*

For a two group model *Q* is equal to 0 when all partners are selected proportionately, equal to 1 when all partners are selected purely *with-like*, and *Q* is equal to -1 when partners are selected purely *with-unlike*.

Effects of mixing on rate of STI spread and equilibrium STI prevalence (for a given R_0 value)

Let's start by varying *Q* while keeping R_0 constant so that we can see the impact of varying the mixing pattern on the rate of STI spread and equilibrium STI prevalence.

If we vary the proportion of partnerships formed between high-activity group members (g_{HH}) between 0 and 1, Q varies from -0.46 (modelling the most *with-unlike* mixing pattern without altering partner change rates), through 0 (when mixing is proportionate), to 1 (when mixing is purely *with-like*).

Now make your own predictions of the overall prevalence of infectious individuals over time in a curable STI model with more *with-unlike* (*Q*= -0.4), proportionate (*Q*= 0), or more *with-like* (*Q*= +0.4) mixing between activity groups.

To keep R_0 equal to 1.36 in these 3 scenarios you will need to set the duration of infection (D) to be equal to:

a) D = 0.340 for *Q*= -0.4,
b) D = 0.167 for *Q*= 0
c) D = 0.097 for *Q*= +0.4 years

As before assume β_p = 0.75, c_L = 1.4 partners/year, c_H = 31.4 partners/year and 2% of the population belong to the high-activity group.

Qu 4-1 What mixing pattern results in the fastest spread of infection?

_____

Qu 4-2 Why?

Qu 4-3 What mixing pattern results in the highest endemic prevalence of infection?

Qu 4-4 Why?

Effects of mixing on equilibrium STI prevalence (for a given STI natural history and partner change rates)

Now let's explore what may happen if the mixing pattern changed but the STI natural history and the partner change rates do not change. Plausibly, this could be the intended or unintended consequence of a behaviour change intervention.

Do this by opening '*STI_3b.mmd*' and running the parameter plot in that model.

This parameter plot varies 'Q' between moderately with-unlike mixing (Q= -0.4), through proportionate mixing (Q=0), to purely with-like mixing (Q=1).

Qu 4-5 What would happen to overall gonorrhoea prevalence if a (intended or unintended) behaviour change intervention changed mixing by sexual-activity from proportionate (Q=0) to slightly with-like (Q=0.3)

Qu 4-6 What would happen to overall gonorrhoea prevalence if the an intervention changed mixing by sexual-activity from slightly with-like (Q=0.3) to mostly with-like (Q=0.9)

Qu 4-7 What would happen to overall gonorrhoea prevalence if the an intervention changed mixing by sexual-activity from proportionate (Q=0) to slightly with-unlike (Q= -0.3)

---

5 Screening strategies for gonorrhoea control

We will now see if our model can be useful in selecting gonorrhoea control strategies.

Open model file '*STI_4.mmd*' or '*STI_4 flowchart.mmd*'

In this model we are assuming that Q=0.

Screening is implemented very simply in this model by adding a term to the equations determining the rate of change of the infectious and susceptible individuals in the high (*ps_H*) and low (*ps_L*) activity groups, to simulate a higher rate of recovery.

These two rates are then altered so that the number of screenings is kept constant but the screenings are targeted at the high, or low activity group, or distributed randomly.

To do this go to the section called "*Screening*". There you will see which lines of code to comment out to target the screening at these groups.

The model equations are:

$$\frac{dS_j(t)}{dt} = -\lambda_j(t)S_j(t) + rI_j(t) + y_j I_j(t)$$
$$\frac{dI_j(t)}{dt} = +\lambda_j(t)S_j(t) - rI_j(t) - y_j I_j(t)$$

Where y_j is the screening rate of high ($j = {}_H$) and low ($j = {}_L$) activity group. For simplicity we assume perfect diagnostic tests and 100% cure if treated.

Qu 5-1 What's the overall endemic infection prevalence if 1,000,000 screenings per year are targeted at:

a) the low activity group: _____%

b) randomly: _____%

c) at the high-activity group: _____%

Qu 5-2 Targeting which group is most effective for controlling gonorrhoea?

_____

Qu 5-3 What problems might there be in implementing this strategy in practice?

Introduction to Infectious Disease Modelling and its Applications – 2018

Session 29: Models for the transmission dynamics and control of malaria

Practical

Overview and Objectives

By the end of this session you should:

1. Be familiar with the Ross-Macdonald model
2. Understand the derivation of the expression for R_0 for malaria
3. Understand how different control interventions, such as vector control, bed nets and chemotherapy will affect malaria transmission.

This practical is structured into 4 parts, one of which is optional:

Part 1 – Review of the Ross-Macdonald model (with "pen and paper") - ***Suggested time: 40 minutes***

1. Review of the structure and assumptions of the Ross-Macdonald model
2. Components of R_0 for micro-parasitic vector-borne infections
3. Impact of different control interventions on malaria transmission

Part 2 – Model exploration and extension in Berkeley Madonna - ***Suggested time: 30 min***

Part 3 – Explore the effect of the latent period in the vector on the transmission dynamics of malaria ***– Suggested time: 20 min***

Part 4 (optional) – Explore the effect of seasonality in vector density

The values of parameters and epidemiological variables used throughout are typical of the *Plasmodium falciparum* parasite transmitted by *Anopheles gambiae* mosquitoes in meso-to-high transmission endemic settings in sub-Saharan Africa (Anderson & May 1991, Dye 1992, Hay et al 2000).

Part 1: The Ross-MacDonald model

1. Model structure and assumptions

Population variables:

S_H = proportion of humans who are susceptible = 1- I_H

I_H = proportion of humans who are infected & infectious

S_M = proportion of (female) mosquitoes which are susceptible = 1- I_M

I_M = proportion of (female) mosquitoes which are infected & infectious

As you saw in the lecture, the equations describing the rate of change in the proportions of susceptible and infected humans and mosquitoes in the Ross-Macdonald model, without considering the vector latency period, are:

$$\begin{aligned}\frac{dS_H}{dt} &= rI_H - \lambda_H S_H \\ \frac{dI_H}{dt} &= \lambda_H S_H - rI_H \\ \frac{dS_M}{dt} &= \mu - \lambda_M S_M - \mu S_M \\ \frac{dI_M}{dt} &= \lambda_M S_M - \mu I_M\end{aligned} \qquad (1)$$

where the forces of infection acting on humans and mosquitoes are given by:

$$\begin{aligned}\lambda_H &= mabI_M \\ \lambda_M &= acI_H\end{aligned} \qquad (2)$$

Parameter definition and values are given in Table 1.

Table 1 – Input parameters of the model

Symbol	Value	Units	Definition	reference
m	10	--	Density of mosquitoes per human	site dependent
a	1/3	1/day	Biting rate per mosquito assuming mosquitoes blood-feed exclusively on humans	Molineaux and Gramiccia 1980
b	0.20	--	Probability of human infection per bite by an infected mosquito	Beier (1994)
c	0.05	--	Probability of mosquito infection per bite on an infected human	Githeko (1992)
r	1/50	1/day	Rate of human recovery	Molineaux and Gramiccia 1980
μ	1/10	1/day	Rate of mosquito death and birth	Gillies 1961

Note that at equilibrium the overall flow in and out of the human and mosquito populations is zero.

Q1.1 What assumptions does the model make about the following:

a) Human recovery and average duration that humans are infectious to mosquitoes.

b) *Human mortality. Do you think this assumption is justifiable? (hint: consider human lifespan and recovery times, and then malaria-induced and natural mortality).*

c) *Mosquito recovery from infection.*

d) *Average time between bites (blood-meals) of a female mosquito (gonotrophic cycle).*

e) *Human and mosquito population sizes (hint: do they change over time?).*

f) *Time to becoming infectious once infected (latency) in humans and in mosquitoes.*

g) *Human acquired immunity.*

2. The basic reproduction number R_0

We now re-derive the R_0 expression for the Ross-Macdonald model already presented in the lectures, but without accounting for the vector latency period. The indirect transmission of the parasite implies the following structure

$$R_0 = R_0(H \to V)\, R_0(V \to H) \qquad (3)$$

We derive each factor separately using the parameter values in Table 1. We assume (as in equations (1)) that mosquitoes blood-feed exclusively on humans (a realistic assumption for *Anopheles gambiae*) and that bites are homogeneously distributed among humans. You may imagine vector and human populations of sizes V=10,000 and H=1000 (but what really matters is the density m=V/H).

Q1.2 Consider one infectious human in otherwise fully susceptible human and vector populations. How many vectors will the human infect? We will consider this in several steps below.

a) *How many times will this infectious human be bitten per day, on average?*

b) *How many times will the infectious human be bitten during the entire infectious period?*

c) *On average how many of these bites lead to mosquito infections?*

d) Write the full expression for the human part of R_0 by combining the 3 expressions above, and calculate its value using the parameters in the Table 1

$R_0 = R_0(H \to V) =$

Now consider one infectious mosquito among otherwise fully susceptible human and mosquito populations. How many humans will this vector infect?

e) How many bites will the mosquito take on humans during its lifetime?

f) How many of these bites will lead to human infections?

g) Write the expression and value obtained for the vector part of R_0

$R_0 = R_0(V \to H) =$

h) Write down the final expression and value for R_0 combining the vector and human components:

$R_0 =$

i) Compare the final expression for R_0 with the one given the lectures:

$$R_0 = \frac{ma^2bcp^n}{r\mu}$$

3. Malaria control measures and sensitivity of R_0

Changes in the value of each parameter may lead to different changes in the value of R_0, depending on how R_0 depends on the parameter, e.g. linearly, inversely or exponentially.

Q1.3 Using the expression for R_0 given above (including the mosquito latent period, as in the lecture) and starting with the parameter values in Table 1 and n=9 days, calculate what happens to the value of R_0 if you change the parameters as described below, one at a time. You may find it helpful to answer this question by using Table 2 to write down the new value of the parameter and the new value of R_0 (using Excel if necessary).

You can assume that the mosquito rate of death μ of 1/10 approximates to a daily survival rate p of 0.9.

a) vector population is halved? (this entry is already done in Table 2 as an example)
b) vector population doubles?
c) vector biting rate is halved?
d) human infection probability per bite from an infectious mosquito is halved?
e) human infectious period is halved?
f) mosquito lifespan is halved?

g) mosquito latent period (n) doubles (e.g. in colder conditions)?

Table 2. Sensitivity of R_0 to changes in the parameters

	m	a (1/day)	b	c	1/r (1/day)	1/μ (1/day)	p (1/day)	n (day)	R_0
Table 1	10	1/3	0.2	0.05	50	10	-	-	5.55
model with latency	10	1/3	0.2	0.05	50	10	0.9	9	2.15
a)	5								1.08
b)									
c)									
d)									
e)									
f)									
g)									

Q1.7 Which parameters are most effective in reducing R_0?

Q1.8 Which parameters would change with the following control measures (indicate whether there is an increase or decrease in the value of the parameter):

a) Self-protection, e.g. skin protection, bed nets
b) Vector control: spraying with insecticides that kill larvae (larvicides)
c) Vector control: spraying interior walls with insecticides that kill adult mosquitoes, or using insecticide treated nets
d) Chemotherapy, i.e. treatment that kills parasites within humans
e) Transmission-blocking vaccines, which target the parasite stages (gametocytes) that are transmissible to mosquitoes

Q1.9 Which control measures are likely to be most effective in reducing R_0 for vector-borne infections?

Part 2: Model exploration and extension using Berkeley Madonna

In Part 1 you examined the assumptions of the Ross-Macdonald model and some of its implications for malaria control based on R_0. In this part you will use Berkeley Madonna to explore both temporal and equilibrium dynamics.

Open the BM file ***Ross-MacDonald.mmd***

The model's basic parameter values are as in Table 1. However, there are new parameters (seasonality and annual EIR) and some parameters are renamed in the BM code, so all the information you need for this part is in Table 3.

Table 3 – Input parameters in BM model

Symbol in BM	Sympol in Part 1	Value	Units	Definition
H		1000	--	Human population size
V		10,000	--	Mosquito population size
mt	m=V/H	10	per human	Mosquito density, can be seasonal
Ih0	I_H(time=0)	1/H	--	Initial human prevalence
gonotrophic_c	1/a	3	day	Gonotrophic cycle, time between mosquito blood-feeds
infectious_p	1/r	50	day	Human infectious period
mosquito_life	1/μ	10	day	Mosquito average lifespan
mosquito_surv	p	0.9	day	Mosquito daily survival probability
latent_p	n	0 or 9	day	Mosquito (fixed) latent period
Annual EIR	$b\lambda_H$		per human per year	Annual entomological inoculation rate, or annual number of infectious bites

There are 3 graph windows in the BM session:
Graph 1 – 'year' on the x-axis
Graph 2 – 'latent_p' on the x-axis
Graph 3 – 'gonotrophic_c' on the x-axis.

Click on 'run' in graph 1, which shows the proportion of infectious mosquitoes and humans over time. Don't worry about the other quantities available at the bottom of the graph unless you have time to explore.

The curves show an epidemic over time starting with a single infected human in human and vector populations of sizes H=1000 and V=10,000. The model is running for just over 3 years (=STOPTIME), which will be changed later.

Q2.1.

a) How long does it take to reach endemic equilibrium?

b) Suppose this represents the result of a human case migrating to a malaria-free community of 1000 people. If a larger number of human cases migrated to a similar community, what can we expect regarding the time to reach endemicity?

c) Does R_0 have the value you calculated earlier?

Q2.2. What is the equilibrium annual EIR?

Field estimates of EIR range from <1 to >1000 per host per year depending on site and season (Hay et al 2000); EIR>100 is regarded as an indication of very-high transmission. This suggests this area has very-high transmission. But remember that the predicted prevalence of infectious mosquitoes is far too high – to achieve a more realistic model we will look in Part 3 at the EIR predicted by the model with latency in the vector.

Find the 'sliders' box, which already contains all the parameters you will be asked to change.

Q2.3 Currently the epidemic starts with a single infected human. Do you think that changing this will affect the final equilibrium level? Check for yourself by changing Ih0 in the sliders box.

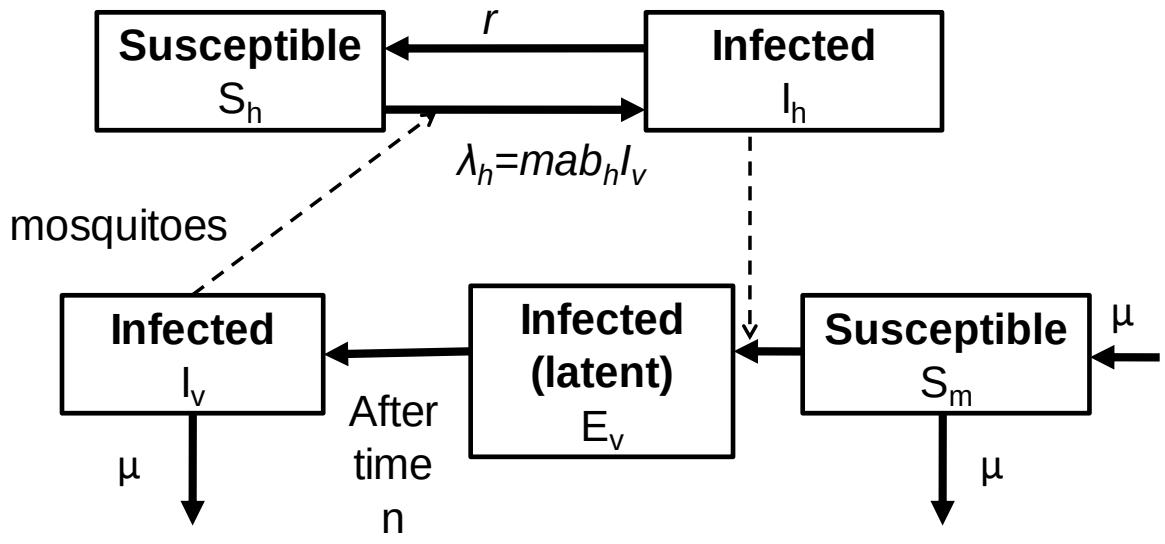
Reset Ih0 to its minimum value.

Q2.4 What do you expect to happen to prevalences and R_0 if the vector density (m) increases?
Change m from 10 to 20 and 100 in the sliders

Reset m to its initial value.

Part 3: Model with vector latent period

So far the BM model assumed that there is no vector latent period ('Latent_p'=n~0). Macdonald (1957) included a (fixed) latent period during which the mosquito is infected but not infectious to humans, as follows:



For a fixed latent period (i.e. not varying among mosquitoes), the equations for vectors in model (1) are modified as follows (the equations for humans are unchanged)

$$\frac{dS_M}{dt} = u - \lambda_M S_M - \mu S_M$$
$$\frac{dE_M}{dt} = \lambda_M S_M - \lambda_M(t-n)S_M(t-n)p^n - \mu E_M \quad (4)$$
$$\frac{dI_M}{dt} = \lambda_M(t-n)S_M(t-n)p^n - \mu I_M$$

where E represents the 'exposed', latent compartment, p^n is the probability of mosquito survival during the latent period n (assuming constant mosquito mortality) and $\lambda_M(t-n)S_M p^n(t-n)$ is the group of mosquitoes becoming infectious (i.e. completing the latent period) who were infected n days ago.

This latent period is ~9-10 days for typical Sub-Saharan conditions (Molineaux and Gramiccia, 1980, Anderson & May 1991 pp.378,387). Note that mosquitoes have an average lifespan (1/μ) that is comparable to n, but some mosquitoes can live longer than 1/μ and therefore survive beyond n.

Set 'latent_p' to 9 days in the 'sliders'.

Q3.1 What happens to the following:
- *a) mosquito and human endemic prevalence?*
- *b) R_0?*
- *c) The endemic EIR?*

Q3.2 Do you now regard the endemicity in this area as high (10<EIR<100) or very-high (EIR>100)?

Sensitivity analysis on latency

A 'sensitivity analysis' is a study of changes in model outcome with changes in the value of one or more parameters. Run a parameter plot by clicking 'run' on graph 2 (with 'latent_p' on the x-axis). Increase STOPTIME to 3000, because for some latent period values the time to reach equilibrium will be larger than the current STOPTIME (BM will run more slowly).

Q3.3 What do you find about prevalence and about R_0? (note the log scale on the x-axis). What does this mean in terms of mosquito infectivity and relationship to climatic conditions?

You can also investigate the relationship by using the 'sliders' and observing the temporal dynamics in graph 1 (possibly further increasing STOPTIME).

Sensitivity analysis on time between blood feeds on humans

Run a parameter plot by clicking 'run' on graph 3 (with label 'gonotrophic_c' on the x-axis).

Q3.4 What do you find about prevalence and about R_0? What does this suggest in terms of vector control (e.g. if vectors also feed on animals)?

Q3.5 In both sensitivity analyses, what is the striking feature when comparing prevalence of infection in humans and in mosquitoes?

Reset the latent period to 9 days. Keep STOPTIME = 3000 days.

Part 4 (Optional): Seasonality in vector density

In reality, many areas with endemic vector-borne infections such as malaria are seasonal in their rain pattern, and therefore have seasonal variation in vector density and endemic levels. Macdonald (1952) suggested that prevalence in malaria areas with high transmission areas would be less sensitive to fluctuations in mosquito density than areas with low transmission. In our model, mosquito density has the form:

$$m(t) = m_0\left[A_0 + A_1 \sin(2\pi t / 365)\right] \qquad (5)$$

where m_0=V/H and t is time in days. So far we have used A0=1 and A1=0, so m_0 is the parameter we have been calling m. In order to compare seasonal with previous (perennial) settings, we set A0 and A1 such that the annual average of m(t) equals our previous value of m. The sinusoidal function is set such that its annual average is zero (it oscillates between -1 and 1 over the period of 1 year). Hence, we keep A0=1 and change A1 between 0 and 1.

In 'sliders', the parameter 'seasonality' represents A1 – set seasonality =1.

Examine the dynamics of m(t) in the graph 1

Q4.1 What is the range of vector density over the year?

Q4.2 Similarly, examine the seasonal dynamics of R_0 and EIR. What do you find?

Q4.3 What happens to R_0 during the dry season?

Q4.4 Examine the dynamics of the prevalence of infection in humans and mosquitoes. What pattern do you observe in the magnitude of annual variation?

Q4.5 Why are the peaks in mosquito prevalence delayed relative to those for human prevalence?

References

Anderson RM and May RM (1991). *Infectious Diseases of Humans: Dynamics and Control*. Oxford, Oxford University Press:Chapter 14.

Beier JC et al. (1994) Plasmodium falciparum incidence relative to entomologic inoculation rates at a site proposed for testing malaria vaccines in western Kenya. *Am J Trop Med Hyg* 50: 529-536.

Dye C (1992). The analysis of parasite transmission by bloodsucking insects. *Annual Review of Entomology* **37:** 1-19.

Gillies MT (1961) Studies on the dispersion and survival of Anopheles gambiae Giles in East Africa, by means of marking and release experiments. *Bull Entomological Res* 52: 99-127.

Githeko AK et al. (1992) The reservoir of Plasmodium falciparum malaria in a holoendemic area of western Kenya. *Trans R Soc Trop Med Hyg* 86: 355-358.

Hay SI et al (2000). Annual *P. falciparum* entomological inoculation rates (EIR) across Africa. *Trans R Soc Trop Med Hyg* **94**, 113.

Molineaux L and Gramiccia G (1980). *The Garki project. Research on the epidemiology and control of malaria in the Sudan savanna of West Africa*. Geneva, World Health Organization (book).

Introduction to Infectious Disease Modelling and its Applications – 2018

Session 30: Network modelling

Practical

Overview

There are 2 parts to this practical. Part 1 looks at some models of epidemics on networks. Part 2 looks at methods of collecting social mixing data.

Part 1: Epidemics on Networks

In this session we'll use the NetLogo software package to explore some models of infection spreading through networks. There will be a chance to investigate in more detail some of the material covered in the previous lecture, and to use simulations to look at some other aspects of network disease dynamics.

This is an **exploratory session**, with time enough only to get a taste of some of the possibilities available. Some of the sections below include some optional questions that you may want to skip over if you're short of time. Although some questions ask for numerical solutions, the object of this practical session is **not** to find precise numerical values; rather, it is to **explore and compare** different modelling approaches. Don't get bogged down seeking precision.

Objectives

By the end of this part of the session you should:

- Understand the dynamics of epidemics spreading through networks.
- Understand the different behaviours of mass-action and network models.

Section 1: NetLogo

NetLogo is a modelling package developed to enable the programming of "multi-agent" systems – essentially anything where there are discrete interacting entities.

It can be freely downloaded from http://ccl.northwestern.edu/netlogo/ and runs on most machines (including Windows and Mac OS X). There is a large library of models available to play around with, or you can write your own – programming it is not too complicated.

It is possible to examine the code behind the models but, because of time constraints, we will mainly treat the model as a "black box" for the purposes of this practical; this does not stop anyone who is especially keen investigating further in due course.

To start up NetLogo

In the Applications folders, go to **SchoolApplications>StatisticalApplications** and click on the **NetLogo 5.0** icon.

> **Useful tip:** if at any point you ask the program to do something that it can't handle and it freezes, you can get out of it by going to **Tools>Halt** at the top of the window.

Section 2: Epidemic models

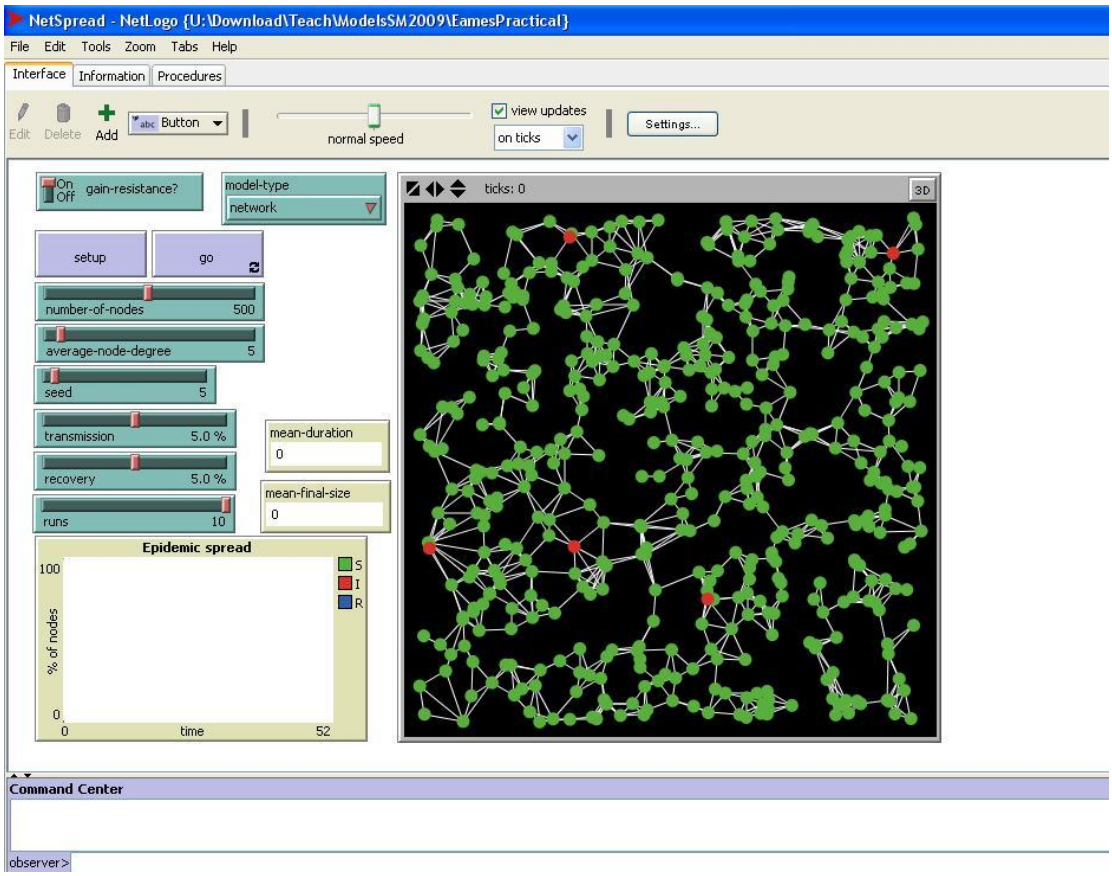
We will use NetLogo to look at epidemics on networks. We'll restrict our attention to a couple of simple network types; spatially local networks and small-world networks.

Go to folder for this session on the network drive. You'll see a few different models there, which we'll be working with for the rest of this practical.

> **Useful tip:** if you want to speed up your simulations, click on the "Code" button at the top of the window. When you click on "Interface" again, you'll see it's been running much faster when it doesn't have to update the plot every time step.

2.1 Comparing network and mass-action models

Open the program **NetSpread** and press **setup**. Your screen should look like this:

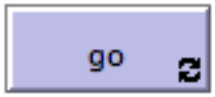


The structure of the NetLogo interface is as follows:

- **Green boxes**: parameters you can set (e.g. “number-of-nodes”)
- **Beige boxes**: output information, which may be simply a number (e.g. “mean-duration”) or may be a plot (e.g. “Epidemic spread”).
- **Purple boxes**: these run the model (e.g. “setup”, “rewire-one”).

You’ll see at the top of the window 3 buttons labelled “Interface”, “Info” and “Code”. If you want to get some background information about the model, press “Info”. If you want to see the code that runs the simulations, press “Code”. Otherwise, we’ll stick with “Interface”.

In general, buttons such as this:



with a pair of arrows in the bottom corner allow you to turn their operation on and off. Here, it means that you can pause the epidemic process.

Useful tip: the ‘go’ button is disabled until the ‘setup’ button has been pressed.

The program **NetSpread** allows you to simulate the stochastic spread of infection through a network of local contacts.

On the right is a network of contacts, with nodes coloured as follows:

Green=susceptible; Red=infected; Blue=recovered.

The controls allow you to do the following:

- You can adjust the properties of the network with the “average-node-degree” and “number-of-nodes” sliders.
- You can adjust the pathogen parameters using “transmission” and “recovery”. The values in these boxes give the per time step probability of the relevant event.
- You can use “seed” and “runs” to alter the seed size and the number of stochastic runs.
- The “gain-resistance?” button allows you to decide whether a node becomes immune following infection. This lets you switch between SIS and SIR infections.
- The “model-type” button allows you to model either a **network** (where infection only passes along links) or a **mass-action** model (where individuals interact according to how many links they have, but their force of infection is such that it is as though these interactions take place at random).

Important point: In the mass-action model, the links, although still shown, do **not** represent interactions through which infection spreads. Links are retained to show heterogeneities and to allow you to switch between mass-action and network models during a simulation.

The “Epidemic spread” graph will show the S, I, and R time series. The other output boxes will show the mean final size and mean duration of the epidemic.

Q 2.1.0: Take a few minutes to explore the model, testing the various options. Note that the local nature of spread within the network means that different parts of the population can behave very differently during an epidemic.

Q 2.1.1: Generate a network with 500 nodes of average degree 5. Set "recovery"=5% and "seed"=5. What value of "transmission" do you need to generate a ***network*** *SIR epidemic that infects approximately half the population?*

Q 2.1.2 Do the same for the ***mass-action*** *model. What is the difference? Why?*

You should see that a much lower level of transmission is needed for infection to spread effectively in a mass-action model. In a network, infection is more vulnerable to local depletion of susceptibles, saturation, and fade-out as correlations build up between connected individuals. An infected person in a mass-action model can, in theory, infect any other individual, whereas in a network he only contacts a few neighbours. Particularly in localised networks, as here, infection struggles to emerge from tightly connected local cliques.

Q 2.1.3 Keep "seed"=5 and "recovery"=5%. What value of "transmission" is needed for infection to take off in a ***mass-action*** *SIR model (you will need to decide how to define an epidemic that has "taken off" – several different definitions would be acceptable)?*

Q 2.1.4 What value of "transmission" is needed for an SIR infection to take off in the ***network*** *model? (optional) How does this vary with "average-node-degree"?*

Q 2.1.5 (optional) Knowing what you do about R_0 in network and mass-action models, could you have predicted the above results? If not, why not?

Not surprisingly, infection takes off more easily in a mass-action than a network model. You know that in a homogeneous mass-action model

$$R_0 = (\text{number of contacts}) \times (\text{transmission rate}) / (\text{recovery rate})$$

In Q 2.1.3, R_0=5 x "transmission"/5. Therefore, you would expect "transmission"=1% to give R_0=1. In a homogeneous unclustered network we've seen that

$$R_0 = (\text{degree}-2) \times (\text{transmission rate})/(\text{recovery rate}),$$

so we would expect "transmission"=1.67 to give R_0=1.

You might see that it's not this simple: clustering in the network reduces R_0, but heterogeneities due to not all nodes having the same degree increase R_0. Clustering makes no difference to the mass-action model, so in this model you might have seen that sometimes an epidemic can take off even when R_0 might be naively predicted to be below 1. In the network model, the clustering effects tend to dominate, meaning that epidemics are less likely to take off. This illustrates nicely the difficulty in defining a sensible R_0 for epidemics on networks.

Q 2.1.6 Run an SIS epidemic on a ***network*** *with 500 nodes of average degree 5, and "recovery"=5%, and note its equilibrium behaviour. Adjust "transmission" such that the equilibrium prevalence is approximately 50%. Since this epidemic does not die out (or, at least, not for a very long time), you'll need to press the "go" button again to stop it running.*

Useful reminder: you can adjust the parameters while the model is running; set an SIS model going then adjust "transmission" until you get the desired prevalence.

*Q 2.1.7 Using the parameters just determined, for "seed"=1, how long does it take for the **network** epidemic to reach equilibrium (if it doesn't fade out)?*

*Q 2.1.8 How long does it take for a **mass-action** model with the same equilibrium prevalence to reach equilibrium?*

Once again, the local aspect of infection on networks reduces prevalence and slows spread. Network epidemics last longer than mass-action epidemics (of the same final size) since the population is less "well-mixed". If you increase "seed" you'll see the difference between the two models is reduced.

The following three **optional** sections give you the chance to look at epidemics on small-world networks (section 2.2); the impact of vaccination (section 2.3); and strain competition on networks (section 2.4). You may well not have time to work through all these sections – feel free to pick the one that most interests you.

2.2 Small-world epidemics (optional)

We have already seen that adding long-range links to a local network reduces the clustering coefficient and path length; we will now seek to determine the effect on disease spread. Here, we will adapt the localised network to include some long-range connections, which we'll call **small-world links**.

Open the program **NetSpreadSW**. **NetSpreadSW** is identical to **NetSpread**, aside from the slider labelled "sw-links". Set this value to 5 and press **setup**. You should see some long-range connections appear in the network (shown as red lines).

Q 2.2.1 Generate a network with 500 nodes of average degree 5. Set "seed"=2, "recovery"=5%, and "transmission"=6.5%. Using an SIR model, determine the effect of small-world links on epidemic duration and final size: what if 0.5% of links are long range? What if it's 5%?

> **Hint:** For a network with N nodes of average degree k, there are N x k/2 links (since each link has 2 ends). Therefore, a network with 500 degree 5 nodes has 1250 links.

You should see that small-world links allow infection to escape from local cliques. This corresponds to their ability to reduce path lengths. Hence the presence of small-world links means that infection spreads more quickly and more widely.

Q 2.2.2 How does the impact of "sw-links" depend on the seed size? Try for "seed"=5 and "seed"=20.

The larger the seed, the smaller the effect of long-range connections: if infection is already seeded in all parts of the network, there is little effect of path length reductions. Hence we might expect small-world links to be most important when infection is first seeded into a population but to be less significant thereafter.

2.3 Vaccination (optional)

We'll now look at a model of vaccination within a network. Open **NetSpreadVacc** and take a few minutes to explore the model. As you'll see, this model allows you to vaccinate a proportion "vacc-frac" of the population (vaccinated nodes are shown in pink).

*Q 2.3.1 Use "transmission"=7%,"recovery"=5%, and "seed'=5, on a **network** with 500 nodes of average degree 5. Set "vacc-frac"=20%. Run the model a few times. Change "target-vacc?" from **off** to **on** and see what happens. What's the difference?*

You should spot two things: that now the nodes with largest numbers of contacts are the ones that are vaccinated, and that the epidemic has a lower impact (both in terms of final size and duration).

Q 2.3.2 Test the usefulness of targeted vaccination for a few different sets of infection parameters. How much difference does targeted vaccination make?

Q 2.3.3 (optional) Is targeting by degree the only way to target? How else might you want to target vaccination? How might you use simulations to help improve vaccination strategies?

*Q 2.3.4 Now return to the **mass-action** SIR model and use the parameter values found in Q 2.1.2 to give an epidemic with a final size of about 50%. Determine what level of vaccination is required to prevent infection spreading.*

Q 2.3.5 Could you have predicted this?

This returns us to the earlier R_0 discussion; if R_0 is reliably approximately by (number of contacts) x (transmission rate) / (recovery rate), then we can use this to estimate the critical level of vaccination for herd immunity as $1- 1/R_0$.

*Q 2.3.6 Repeat Q 2.3.4 for the **network** SIR model.*

*Q 2.3.7 Is targeted vaccination equally useful in the **network** and **mass-action** models? If not, why not?*

Targeting an intervention allows the most influential individuals to be removed, which we would expect to make a difference in any population that displays heterogeneities. In a network it has the added benefit of breaking the network up into unconnected components, preventing infection passing from component to component. The additional effect will not be evident in a mass-action model since, by its nature, this model is well-mixed.

*Q 2.3.8 (optional) Is the critical level of vaccination the same for an SIS **network** model as for an SIR **network** model? Is it the same for **mass-action** SIS and **mass-action** SIR models?*

You should see that in a mass-action model, the critical level of vaccination is the same for SIS and SIR infections. In a network, however, SIR infections require less vaccination to stop them taking off than SIS infections do; because of the localised nature of spread within networks, recovered individuals can physically block the spread of infection and thus act as a natural barrier. Here again we see the effect of the fundamental difference between network and mass-action models: in a network pairs of individuals interact repeatedly whereas in a mass-action model no lasting interactions occur.

2.4 Strain competition (optional)

We'll now look at strain competition. Open **NetSpreadStrain**. This program allows two different strains of infection to spread through the same network. For many infections, several different strains are observed to coexist: however, theory suggests that the fitter strain should drive the others to extinction, so the reasons for this coexistence are not always clear.
In the model here, a node can be infected by only one strain at a time; the model shows one strain as yellow, the other as red. Take a minute or two to familiarise yourself with the new features of the model.

*Q 2.4.1 Generate a network with 500 nodes of average degree 5, and "seed"=5. For an SIS **network** model, investigate for how long 2 identical strains (give them both "recovery"=5% and "transmission"=5%) coexist. For how long do they coexist in a mass-action model?*

Q 2.4.2 What if the strains are not identical? For instance, try increasing "transmission" to 5.1 for one of the strains.

*Q 2.4.3 (optional) Decide how you might determine which strain is the fittest. How much "less fit" does one strain have to be in order for it to fail to persist in a **network** model? What about in a **mass-action** model?*

You should see that coexistence is much more likely and lasts much longer within a network: the segregation provided by the network allows different strains to dominate in different parts of the network.

*Q 2.4.4 For strain 1 set "recovery"=2% and "infection"=1% while for strain 2 set "recovery"=10% and "infection"=5%. Which one would you expect to dominate? What happens when you run the **network** model?*

*Q 2.4.5 Does the outcome change in a **mass-action** model?*

Here, there is a fast strain and a slow strain that would, in a differential equation model, be perfectly matched. The fast strain is expected to be better at colonising new areas of the network, whereas the slow strain is expected to be better able to persist. It is not easy to predict what will happen in a finite, stochastic, model but we see that, generally, they coexist for a long time. In a mass-action model the fast strain is at an advantage, since it can colonise and exclude the slower strain, reducing the slow strain to lower prevalence and greater risk of stochastic fade-out.

Q 2.4.6 (optional) Look at the effect of small-world links on strain coexistence.

As you would expect, adding long-range links increases the interaction between different parts of the network and makes strains less likely to be able to find separate niches to persist in. Lack of interaction between different population sub-groups is one postulated reason for the coexistence of multiple strains of infection.

Part 2: Collecting Social Mixing Data

Here we look at some of the tools commonly used to collect the data required for network or age structured models.

There's no computing in this exercise. Instead, you'll be asked to complete two different "contact surveys" and think about the methods in detail.

Objectives

By the end of this part of the session you should:

- Be familiar with two methods of measuring social contact patterns.
- Understand the strengths and weaknesses of these methods.
- Identify potential sources of error and bias.
- Consider the quality of data commonly available, in relation to the complexity of models sometimes used.

Section 1: Contact surveys

1.1 POLYMOD

The POLYMOD study (Mossong J, Hens N, Jit M, Beutels P, Auranen K, et al. (2008) Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. PLoS Med 5: e74) was a project to measure social mixing patterns. It builds on work by Edmunds and others (e.g. Edmunds et al (1997) Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections. Proc R Soc B 264: 949-957).

The study uses paper questionnaires ("contact diaries") to gather information from study participants about who they meet over a given time period (usually 1 day). Participants are asked to record each person they encounter, and to give some details about that person and about the encounter.

The POLYMOD study involved 7,290 participants in 8 different European countries. The POLYMOD data have been used to parameterise age-structured models; the data suggest strong like-with-like mixing, particularly among school children. The main figure from the Mossong et al paper is shown below:

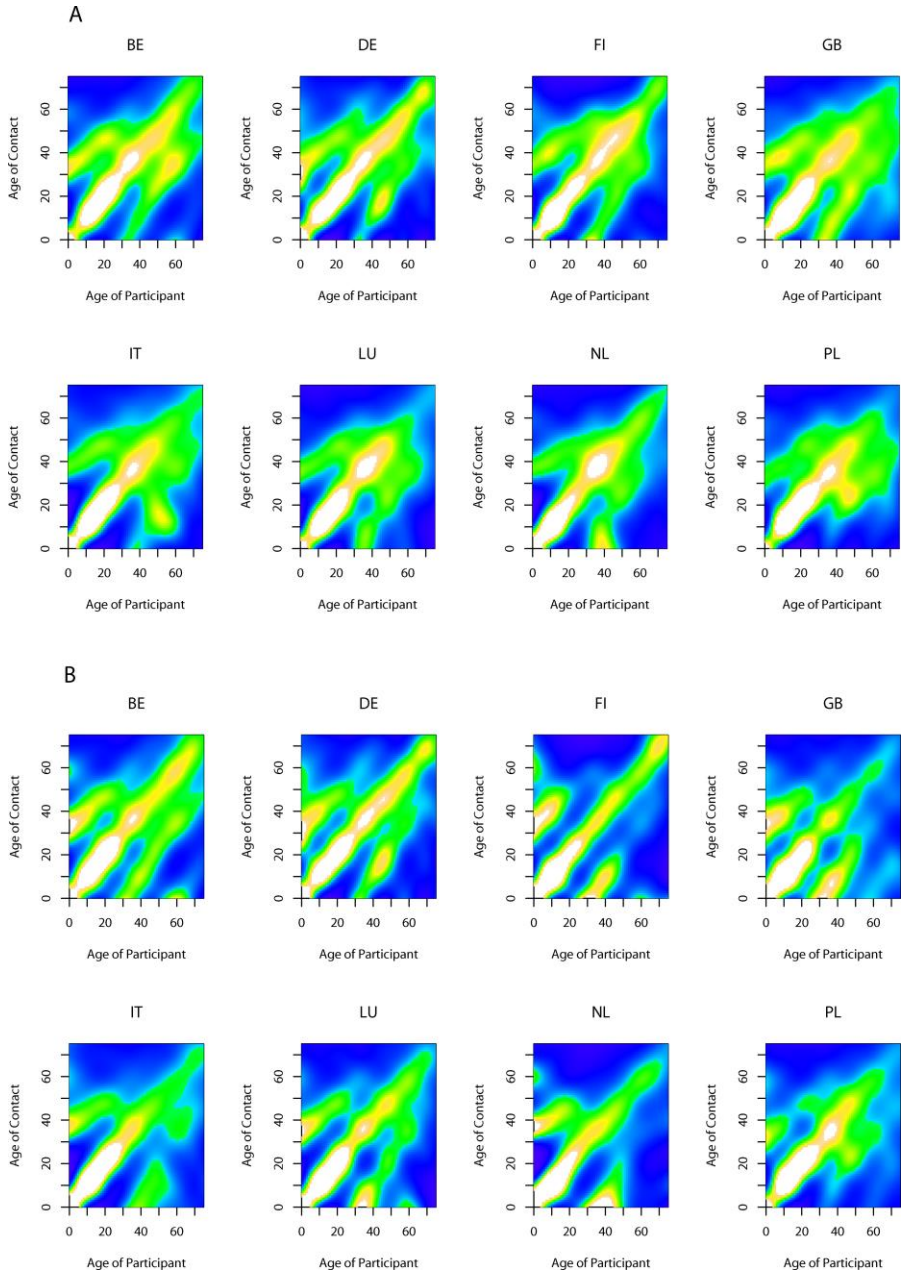


Figure. Smoothed Contact Matrices for Each Country Based on (A) All Reported Contacts and (B) Physical Contacts Weighted by Sampling Weights. White indicates high contact rates, green intermediate contact rates, and blue low contact rates, relative to the country-specific contact intensity. Fitting is based on a tensor-product spline to contact matrix data using a negative binomial distribution to account for overdispersion.

The contact diary used in the POLYMOD study can be found Section 3.1.

1.2: flusurvey

The UK flusurvey (www.flusurvey.org.uk) is an internet-based influenza-surveillance platform, in which members of the public are asked to register and report any flu-like symptoms (or lack of symptoms) each week. It was launched by Eames & Edmunds in 2009 as part of a European-wide study.

As well as asking participants about their symptoms, the flusurvey includes a “social contact” questionnaire (Eames KTD, Tilston NL, Brooks-Pollock E, Edmunds WJ (2012) Measured Dynamic Social Contact Patterns Explain the Spread of H1N1v Influenza. PLoS Comput Biol 8:

e1002425). Rather than asking about **each** person whom participants meet over the course of a day, the flusurvey social contact survey asks participants to record the **number** of people they met, categorised into several different age groups.

Because the flusurvey runs for several months each year, the contact data have been used to explore the difference in contact patterns between school term time and school holidays; a marked difference is found in the interaction between school age children, while contacts between other age groups are almost unchanged. The main figure from the Eames et al paper is shown below.

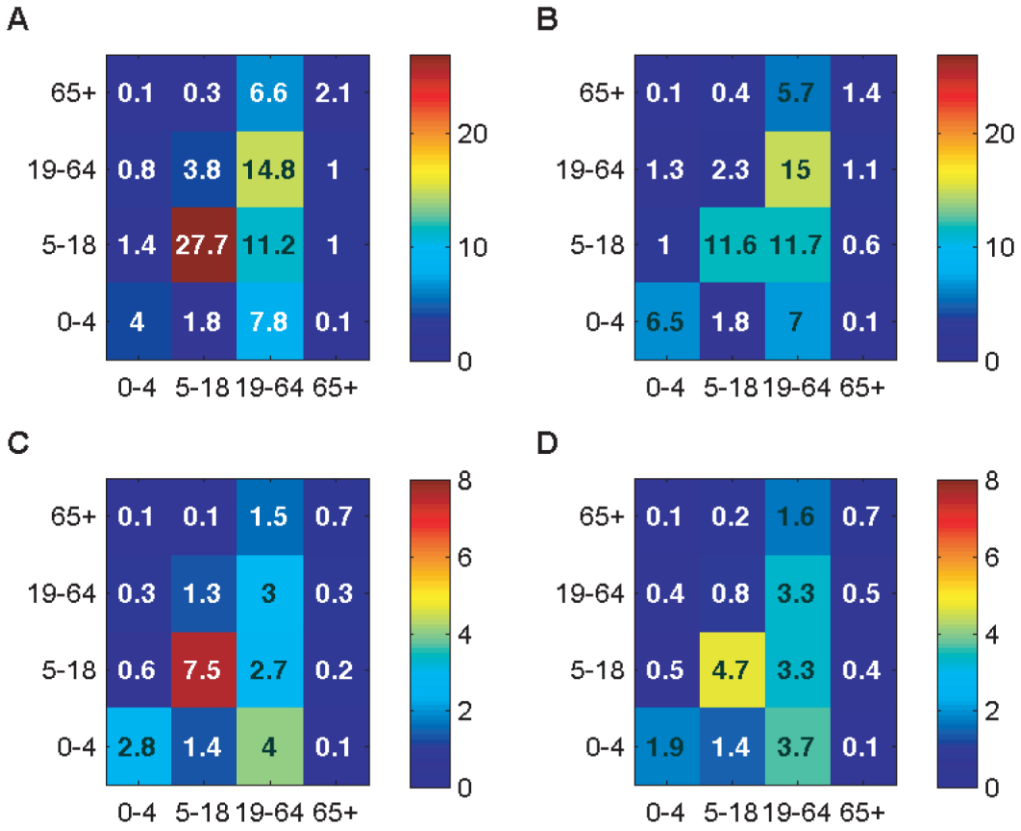


Figure. Social contact matrices. Values and colours show the mean number of contacts per day reported between each age group. In each panel, the participant's age group is shown on the vertical axis, that of their contacts on the horizontal axis. The four panels show patterns of A: conversational contacts during school term time; B: conversational contacts during school holidays; C: physical contacts during school term time; D: physical contacts during school holidays.

The contact survey used in the flusurvey can be found in Section 3.2.

Section 2: Practical exercise

Work through the two different contact surveys, filling them in as though you were a study participant.

Q 3.1 In groups, discuss the strengths and weaknesses of the two approaches. Think about:

- *Ease of completion.*
- *Accuracy of data generated.*
- *Sources of bias.*
- *Usefulness of data in models.*

Q 3.2 If you were a study participant, what changes would you make (if any) to the survey or the instructions?

Q 3.3 If you were carrying out a study to measure contact patterns, what changes would you make (if any)?

Q 3.4 We could attempt to measure contact networks using the POLYMOD contact diary and stipulating that participants should record the names of their contacts. What are the strengths and weaknesses of such an approach?

Section 3.1: POLYMOD contact diary (adapted)

- We would like you to record in the diary **every** person that you had **contact with** on the date specified by your tutor.
- A contact is defined as:
  - EITHER a **two-way conversation** with three or more words in the physical presence of another person,
  - OR physical **skin-to-skin contact** (e.g. a handshake, hug, kiss or contact sports).
- Write down **each person that you contact** during the day, regardless of whether the contact was long or short, and whether you know the person or not.
- Contacts made exclusively by **telephone** or **mobile phone** should **NOT** be recorded.
- If you contact the same person several times in the course of the day, only record him/her once, and record the total time you spent with that person over the entire day. Each person you have contact with should have one line in the diary: **one person, one line.**
- Please provide some information on your contact, namely:
  - Age.
  - Gender.
  - Whether there was skin-to-skin contact.
  - How often you have contact with this person in general.
  - Place(s) where contact(s) occurred (you may indicate several locations).
  - How long the contact with the person lasted over the entire day.
- If you don't know the exact age, please **estimate the age range** (e.g. 40-45), and try to make it as narrow as possible.
- **Estimate** the total **duration** you spent in the presence of the contact person that day. Example: 1-5 minutes for a contact in a shop or 1-4 hours caring for a child at home.
- Once you've filled in the diary, we suggest that you double check the entries; e.g. try to **remember** all of your **activities** to make sure you haven't missed any contact persons.
- The **order** in which you write down your contact persons is **not important**. The easiest is to use a **chronological order** according to when you met the person for the first time during each assigned day and then add anyone else who you remember as you go through your daily activities.

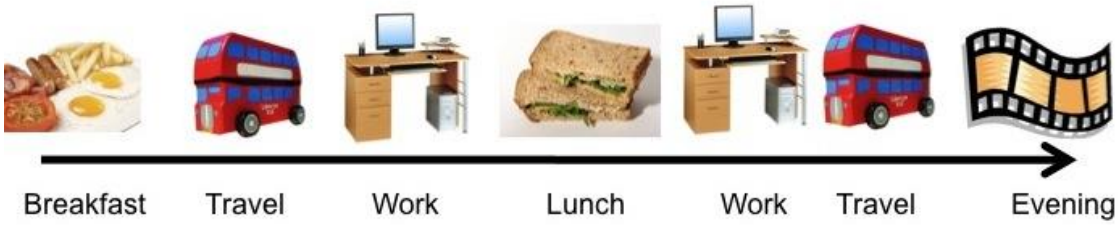
For the purposes of this study, the day starts at 5 a.m. on the morning of the day assigned, and ends at 5 a.m. the next morning.

Initials (or description) of contact	Age (or age range)	Gender (M or F)	Contact type Was there skin-to-skin contact? (Y or N)	Regularity In general, how often do you have contact with this person?						Location Where did this contact take place (tick any that apply)?						Duration Over the day, for how long were you in contact with this person?				
				Never met before	Less than monthly	Once or twice monthly	Once or twice weekly	Daily or almost daily		Home	Work	Travel	Social	Other		Under 5 mins	5-15 mins	15mins-1 hour	1-4 hours	Over 4 hours

Initials (or description) of contact	Age (or age range)	Gender (M or F)	Contact type Was there skin to skin contact? (Y or N)	Regularity In general, how often do you have contact with this person?					Location Where did this contact take place (tick any that apply)?					Duration Over the day, for how long were you in contact with this person?				
				Daily or almost daily	Once or twice weekly	Once or twice monthly	Less than monthly	Never met before	Home	Work	Travel	Social	Other	Under 5 mins	5-15 mins	15mins-1 hour	1-4 hours	Over 4 hours

Section 3.2: flusurvey contact diary (2013 version)

Flu is spread via social contacts. Measuring how we meet each other helps understand and predict flu epidemics. Think about the all people you met yesterday and where you met them.



Question 1: How many people did you have conversational contact with yesterday (talking face to face)?
N.B. the contacts survey on the flusurvey website has a dropdown menu for each box in this table: options 0; 1; 2; 3; 4; 5; 6; 7; 8; 9; 10; 11; 12; 13; 14; 15; 16-24; 25-49; 50-99; 100+

	0-4 years	5-18 years	19-44 years	45-64 years	65+ years
Home					
Work					
Other					

How should I answer it? Only record people in one setting. So if you saw someone at work and socialising, then only record them in the place where you spent most time together. "Home" means your home. "Work" means you place(s) of work (or school for younger participants). Anyone you have had a conversational contact AND a physical contact with should be recorded both here and in the next question – i.e. they should be included twice.

Question 2: How many people did you have physical contact with yesterday (skin-to-skin contact, e.g. handshake, kiss)?

	0-4 years	5-18 years	19-44 years	45-64 years	65+ years
Home					
Work					
Other					

Question 3: How much time did you spend on public transport (e.g. bus, train, underground) yesterday?

No time at all	
0-30 minutes	
30 minutes - 1.5 hours	
1.5 hours – 4 hours	
Over 4 hours	

Question 4: How long did you spend in an enclosed indoor space (e.g. office, classroom, bar, cinema) with more than 10 other people yesterday? (Not including public transport)

No time at all	
0-30 minutes	
30 minutes - 1.5 hours	
1.5 hours – 4 hours	
Over 4 hours	

Question 5: What was the furthest distance from home that you travelled yesterday?

Under 1 mile	
1-4 miles	
5-9 miles	
10-29 miles	
30-100 miles	
Over 100 miles	

Section 4.1: EXCEL file for collecting information on contact patterns

The folder for the models for this session includes an Excel file called "Survey.xlsx", which is designed to help you to compile the data from the Flusurvey and POLYMOD surveys and facilitate discussion in your group. We assume that small groups comprising at most 5 people have gone through the Flusurvey and the adapted POLYMOD questionnaires.

Each member should copy the answers given in Sections 3.1 and 3.2 of the practical into the Excel file. Please use only one Excel file per group when entering the data. The Excel file also provides some histograms and plots on distribution of contacts by age, gender, location etc. The analysis and results are similar to those conducted by researchers for analysing contact data. The following describes the file and any issues relating to entering the data in further detail.

The Excel file consists of 3 worksheets:

1. Flu Survey: this collects the survey answers from each member of the group and the results are summarised using histograms
2. Polymod: this worksheet reproduces the adapted Polymod survey as seen in the practical. This collects the information on contacts from each group member, with one line being used for each contact recorded
3. ResultsPolymod: this produces analyses of the data in the POLYMOD questionnaire.

We will begin by looking at the Flu Survey form.

Flu Survey Instructions

The worksheet is divided in 3 sections, as described in the practical notes. The first section (columns A-G) collects details, including the age ranges, of conversational contacts; the second section (columns K-Q) collects details on the physical contacts; the third (columns U-AG) collects the activity data (time spent indoor on public transportation and walking distance). To begin with, each table has 0 in each of the cells.

Each member of the group should first put her/his name in the green cell in column A for the corresponding respondent number. The name will then automatically appear in other parts of the worksheet. Each member should enter the number of contacts that they made on the specified day with others in given age groups and location in the appropriate row in columns B-G (conversational contacts) and L-Q (physical contacts). Similarly each person should complete the table in columns U-AA for the time spent in given locations.

Once you have entered the number of contacts that you have made, you will see that the plots and histograms in the Results and Plots section (row 33 onwards) will have been updated.

For the Conversational/Physical contacts:

- The line/scatterplot shows the number of contacts established by each respondent;
- The first histogram shows the fraction of contacts established at home/work or during other activity by each respondent (represented using different colours);
- The second histogram shows the age distribution of the contacts reported by each respondent.

For the activity part (3rd section) histograms are shown beside each table and show the distribution of the time that respondents have spent on public transport, indoors and movement distance.

Q.4.1 Do you observe any differences between the activities? Where do most of the contacts take place? Is this the same for all the respondents?

Polymod Instructions

In this worksheet, every member of the group should report details of the contacts that they had on the day proposed by the facilitator, noting the following.

1. Each line provides the details for a single contact.
2. In the first 3 columns (red ones) the person completing the survey should put his/her initials, his/her age and gender. This information should be repeated for every contact that the person made. For example, if you made 10 contacts on the specified day, you should copy this information down for a further 9 rows.
3. Columns D-F contain the same type of information as columns 1-3 but relate to the contact. Note that in the column relating to age, you should type in the age range and that 0-5/1-0 would be written as 00-05/05-10.
4. Column G relates to the type of contact made.
5. In the other columns for each contact, you should put 1 if the option mentioned is applicable; otherwise you should leave it blank.

After all the members in the group have completed this part, please look at the third (Results Polymod) Worksheet.

Instructions for "Results Polymod"

Column A is simply a copy of column A from the previous worksheet and contains the initials of the respondent or 0 when the cell is empty. If there are any unusual characters in these cells, copy the first cell down until all the entries of column A that are in the previous sheet appear.

Follow the steps below to see the summary results of the questionnaire:

1. Select all of Column A including the title.
2. Click on the Data button and select Advanced Filter
3. In the 'Advanced Filter' Menu select 'Copy to Another Location'
4. Click on 'List range' and select all the cells in column A until the first 0 appears. Remember to also include the header
5. In the 'Copy to' box, write B1
6. Select the tab 'Unique Records Only'
7. Click OK

Column B should now have a list of unique IDs corresponding to each person who completed the questionnaire.

Cell C2 has a formula for the number of contacts that the person listed in cell B2 made. Copy this formula down for every person listed in column B.. The line plot in the section called Plots and Histograms plots the values in column C. Below the line plot, you should see histograms of the Distribution of contacts by location, regularity and duration.

The “Pivot table” tool in Excel can be used to summarize information. In our case, we are interested in summarising how many contacts are established among different age groups, or between genders. The Pivot tables (columns L-R) report information on the contact pattern by age group and gender. For both pivot tables, the column labels correspond to the responder (person filling out the questionnaire), while the row labels correspond to the contacts. The number shown in the cell is the total number of contacts established by members in the 2 categories, for example, how many contacts respondent men have established with women. You can change the stratification of the information in the pivot tables by doing the following:

For each of the pivot tables:

1. Select the table.
2. Right click; then, in the Pivot table tab, click “Refresh”, and then select the option “refresh”.
3. Click the filter icon in the Row labels cell and de-select blank

The number appearing in the table cells correspond to the number of contacts between the age/gender group of the respondents and their contact.

Cells M4, M5, M18, and M21 contain possible filters. Select one or more of them to see how the contact pattern changes.

Q.4.2. Discuss possible differences in the contact patterns with other members of the group. Does the age-specific contact pattern differ between men and women? Is there any assortativity in gender and/or age?

Occasionally, depending on the version of Excel that you are using, the pivot table may show unusual results. If that occurs, do the following for the first Pivot table:

1. Select the Pivot table
2. In the Pivot table menu select Builder
3. Drag the Field Name “Your Age” in the column Labels area
4. Drag the Field Name “Contact Age” in the row Labels area
5. Drag the field names” “Your gender”, “Contact gender”, ”Type of contact” in the Filter area
6. Drag “Contact Type ” field name in the Values Area

For the second Pivot table follow the same procedure but using Your Gender and Contact Gender as Column and Row labels.

Introduction to Infectious Disease Modelling and its Applications – 2018

Session 32: The practical application of real-time modelling

Practical

Overview and Objectives:

The purpose of this practical is to provide you with some experience of fitting models to data in real-time. By the end you should:

- Understand the relationship between serial interval and R_0;
- Be able to fit models to emerging data;
- Understand some of the problems of real-time data analysis.

1. Estimating R_0 for SARS

There are many ways of estimating R_0, such as by looking at the inter-epidemic period, or the proportion left uninfected after an epidemic, or by just looking at the epidemic curve in a totally susceptible population. Although this is not a commonly used technique, as we rarely observe infections in a completely susceptible population, this was used in real-time to estimate the R_0 for SARS. Hence, we will attempt to do just that.

Open up the spreadsheet SARS & Flu_R0.xls.

The spreadsheet is organised as follows. There are 2 parameters in the top right corner that refer to the average incubation period, and the average duration of infection. We will assume that these are known (i.e. we will not be estimating them). We will also assume that we know the serial interval (the time between successive cases, or generation time) from contact tracing and that it equals 12 days. To make life easier, we assume that there is no variability in this, but you should be aware that this is an enormous (over)simplification. At the very beginning of such an epidemic we can neglect to keep track of the depletion in susceptibles as the epidemic progresses, as the number of individuals in the population that are not susceptible will be negligible. Under these circumstances a single case will generate:

R_0 cases after the first generation,
R_0^2 cases in the 2 generation
R_0^3 cases in the 3 generation, and so on.

That is, the number of cases t generations after the initial (It) case is given by

$$It = R_0^t$$

And the cumulative number of cases observed up to generation t is simply:

$$\text{Cumulative cases} = \sum It$$

We can utilise this to estimate R_0 from the epidemic curve. In the example given we use least squares estimates to derive a best guess of R_0. That is, we choose a value of R_0, then compare the sum of the squared differences between the model and the data. We keep

choosing values of R_0 until we can minimise the sum of the squared differences between the data and the model.

1) Change the values of R_0 and see what happens to sum of the squared differences. Also observe how this affects the graph (which compares the model results with the data). Try and choose a value of R_0 that minimises the squared differences between the model and data.

Excel has an add-in program called Solver, which will run through lots of values of a cell and choose the one that fulfils certain criteria (either maximises a value in another cell, or minimises it, or sets it to zero). To get Excel to do this we must first run Solver, tell it which cell it has to change, and which one it is trying to minimise (or maximise, or set to zero). In our case we want to minimise Cell G16 (sum of the squared differences) by changing cell I18 (R_0).

2) Run solver and estimate R_0.

3) What do you think of the fit of the model to the data? How could you improve this fit? What are the major limitations of this technique?

4) Were we right to be worried about SARS?

5) How do you expect your estimate of R_0 for SARS to be dependent on your estimate of the serial interval? What would you expect it to be if it were:
a) shorter (say 6 days)
b) longer (say 24 days)
Try and answer the above on paper. Then change the spreadsheet, and see whether this confirms your suspicions.

2. Estimating R_0 for influenza and comparing to SARS

This second part of the exercise extends this simple model to try and estimate the basic reproduction number for influenza using (again) real data. This time it is real data from the influenza pandemic of 1968/69.

Turn to worksheet “Flu”. The data are given in the top left of the worksheet and are plotted in the graph named “Data”. They consist of weekly consultation rates (per 100,000 per week), for Influenza-Like-Illness, or ILI, over the first few weeks of the pandemic and a few weeks before. Flu-like symptoms are pretty broad, so other diseases are easily confused for influenza. Thus there is a natural “background” rate at which individuals consult for ILI, which we will assume is fixed, though obviously it may vary by season, and even from year to year. Cell E4 gives an estimate of this background rate (per person per half-week).

We will assume that we know that the serial interval for flu is about half a week, and so adapt the previous spreadsheet for SARS, i.e. we will calculate how many cases occur from one generation to the next (a period of 3.5 days). Since we want to project how many cases occur over the course of the epidemic we will need to take account for the depletion of susceptibles (this is the major difference between these two models). The model itself is towards the bottom of the spreadsheet (below Row 26). You should spend a few minutes working out how the model works. Again, we will be comparing the observed data with our

model estimates, and we will be minimising the sum of squared residuals (Cell 026). Again, we will be using Solver to obtain the best fit (we will be trying to minimise the sum of the squared residuals, by changing the model parameters).

There are a number of parameters that we will be assuming that we know, and we will be estimating the others. Initially, we will assume that no-one is immune (almost everyone is susceptible), and (as mentioned above), that the serial interval is 3.5 days.

We do not, however, know when the first case was introduced into this population (or indeed, whether more than one case was introduced). Thus we will have to estimate the initial number of cases, I_0 (Cell B19).

1) If $I_0 < 1$, then what does this mean?

Not everyone who has influenza develops symptoms, and not everyone who does will consult a physician. The model, on the other hand, tracks infectious cases, irrespective of their disease status and consulting behaviour. [Note that this is an important feature of transmission models, that often distinguishes them from models used in, say, health economics. In transmission dynamic models we are interested in infection, therefore we need to keep track of all infectious cases]. As we do not know what proportion of pandemic flu cases that visit a physician we will have to estimate this as well (Cell B18).

We are interested in estimating R_0 (Cell B16), and while we are at it we may as well see if we can get a decent estimate of the background ILI consultation rate (Cell B17).

There are 2 graphs on the right hand side of the worksheet, which show a comparison of the expected ILI consultation rate, with the observed data. The first one just shows the time period for which we have data, the second graph projects the estimate over the remainder of the epidemic.

2) Spend some time changing the parameters (Cells B16:B19) and seeing what effect it has on the epidemic curve.
    a. What effect does changing I_0 have, and why?
    b. What effect does changing R_0 have, and why?
    c. What effect does changing the % who consult have, and why?
    d. What effect does changing the background ILI consultation rate have, and why?

Estimating 3 or 4 parameters by changing them individually is close to impossible, so we will be employing Solver again, to try and get a reasonable estimate of these parameters.

3) Use Solver to estimate these 4 parameters, and look at the graphs comparing your model to the data, and the projections over time.
    a. When does the model predict that the epidemic peaks?
    b. How high is the peak?
    c. What is your estimate of R_0? Does this surprise you (given what you know about influenza), and if so, then what is the explanation?

4) How stable are your estimates? I.e. how sensitive are they to their initial values?

a. Change the initial estimates of cells B16:19 and re-estimate the parameters.
b. Keep a track of each of your "best-fit" estimates, by copying the value in Cell 026 and Cells B16:19 as well as the estimates of the number of ILI cases (Column K) into the bottom part of the spreadsheet (there is the backbone of a table provided from Row 76). You will need to "Paste Values", when you paste the values.
c. After you have done this a few times (say, 10) plot the results, in terms of the projections. How, certain are you about the height and the timing of the peak?
d. Which parameters are most important?
e. Which parameters are more accurately estimated?
    i. Could you simplify the fitting procedure by fixing any of the parameters, and if so, which ones would you choose, and how would this help (try it)?

Throughout the above, we have been assuming that everyone is susceptible at the beginning of the pandemic (which is an often-used assumption, but may not be completely true).

5) What if you relax this assumption, and assume that say 10% or 20% or 30% of the population are immune at the beginning of the epidemic?
    a. How does this affect your estimates of R_0, and why?
        i. Do you think that R_0 is estimable this way, if we don't know the proportion who are susceptible, and if not, then what can we estimate instead?
    b. How does this affect your projections, and why?

6) Can you come up with any other factors that might affect our estimates of the timing and size of the peak?

7) On "Sheet 1", data for the rest of the epidemic is given. Include data up to the end of week 52 into the spreadsheet (i.e. until just before the peak).

    a. How does this affect your estimates of the size of the epidemic, and your parameter values?
    b. Compare your model results with the observed data for the rest of the epidemic. How well does your model do now?
    c. What if you include another week of data (up to end of week 1, Jan 1970)? How does this affect things?
    d. Do all datapoints (i.e. weeks) have the same importance in terms of fitting, or is the data that emerges at certain points in the epidemic more important than others?

8) What does all this make you think about claims for "real-time" model projections?

Introduction to Infectious Disease Modelling and its Applications – 2018

Session 33: Further modelling (optional)

Lecture

This optional session will discuss some of the different software that is available for developing models and will describe further modelling courses that are available at LSHTM.