Block 4:

The applications of modelling

Lecture notes

Introduction to Infectious Diseases Modelling and its Applications – 2018

Session 26: Modelling the transmission dynamics of *M tuberculosis*

Lecture

Overview and Objectives

This lecture is designed to introduce you to the area of modelling the transmission dynamics of *M tuberculosis*

By the end of this session, you should:

- be aware of the complex natural history of tuberculosis
- understand how it affects the development of models for its transmission dynamics
- be aware of some of the areas of application of TB models

Brief overview of the epidemiology of tuberculosis

Tuberculosis is usually caused by the tubercle bacillus, *Mycobacterium tuberculosis*. It can also be caused by other mycobacteria, such as *M bovis* etc. Disease can either be respiratory (also known as pulmonary TB), affecting the lungs or extrapulmonary, affecting sites other than the lungs. Extrapulmonary forms are generally non-infectious. Treatment for tuberculosis (involving isoniazid, streptomycin, PAS) became available during the 1950s; before then the case-fatality depended on the type of tuberculosis (e.g. 100% for tuberculous meningitis). In most developed countries, tuberculosis incidence (and mortality) declined dramatically during the 20th century, even before the introduction of any specific treatment, which has been attributed to various factors (see below). In recent years, the epidemiology of tuberculosis has changed again with the advent of the HIV epidemic (e.g. HIV affects the risks of developing tuberculosis). In many developed countries, the contribution to tuberculosis morbidity from the poor and homeless and from high incidence settings has increased in recent years, such that the decline in incidence has slowed, or even reversed.

The general structure of a TB model

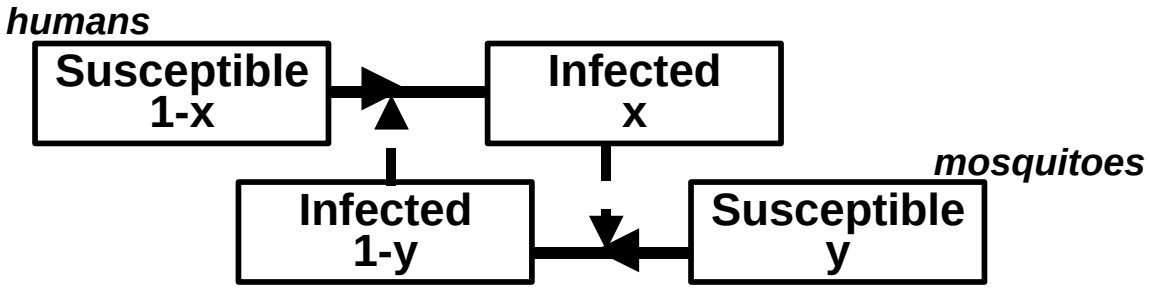
As you will have seen during the course, the general structure of a model of the transmission dynamics of a disease should reflect its natural history. The following shows the general structures of the models of the diseases considered so far:

a) Simple immunizing infections (e.g. measles, mumps, rubella)

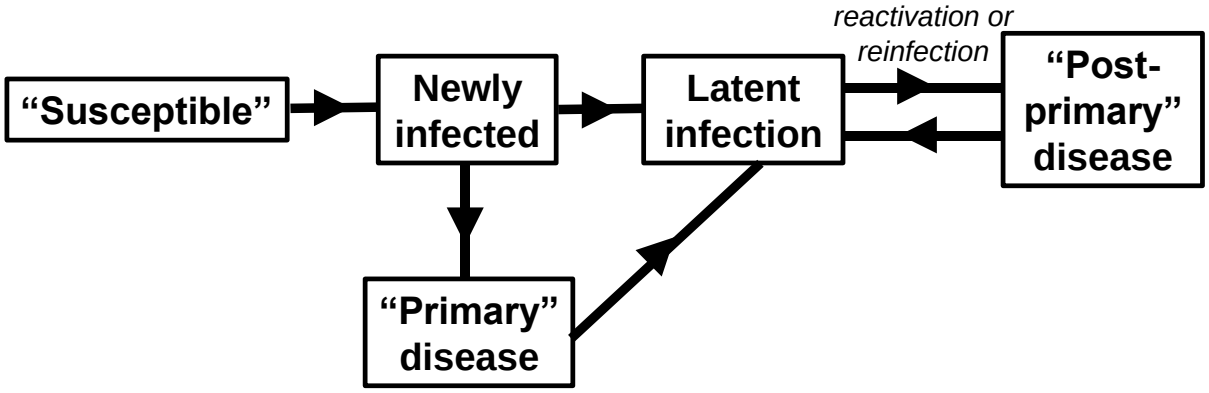
b) STDs

c) AIDS

d) Malaria



For TB, the design of a model is made complicated by the fact that there is much debate about its natural history. Disease can occur either soon after initial infection (defined as primary disease), or many years thereafter through (endogenous) reactivation of an earlier infection or following (exogenous) reinfection. Disease which occurs many years after initial infection is often referred to as "post-primary" disease. This categorization results in the following general structure for a model for TB:



However, the relative contribution of the different forms of disease is unknown and controversial, with some workers emphasizing the importance of disease attributable to reactivation, and others emphasizing the importance of disease attributable to reinfection. The debate has continued in recent years, as a result of increasing numbers of molecular studies, measuring the proportion of cases in a population who share identical isolates, which suggest that disease attributable to recent transmission is more common than previously thought.

There are several other complications in the epidemiology of TB which need to be considered when designing a model.

Complications in the epidemiology of TB

1. The annual risk of infection with *M tuberculosis* varies between settings and has declined dramatically in many industrialized countries, as illustrated in Figure 1. This decline has

mirrored the decline in the incidence or, equivalently, the mortality rate (before treatment became available during the 1950s) from TB.

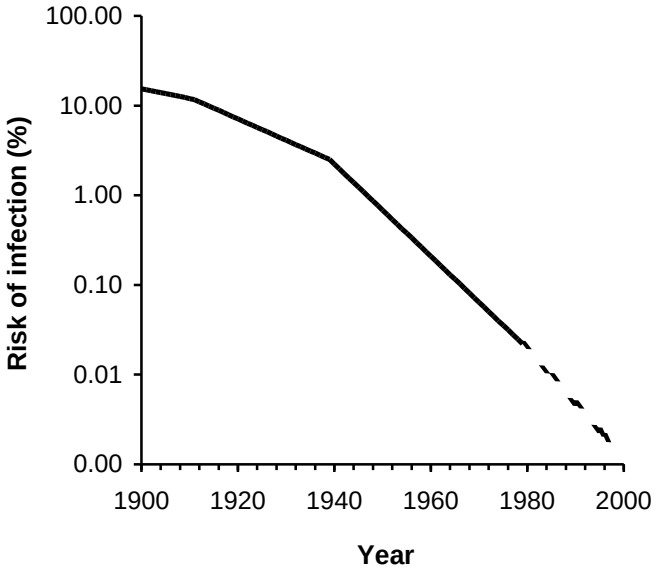


Figure 1: Estimated annual risk of tuberculous infection in The Netherlands [1-2]

2. The risks of disease are age-dependent. This is illustrated in Figure 2, which shows the mortality rate from respiratory tuberculosis experienced in each year of life for individuals born during the period 1891-1900 and 1911-1915 in the UK. (NB Before the availability of treatment during the 1950s, TB mortality provides a good indication of the TB incidence). This figure suggests that the risk of developing tuberculosis increases during the young adult years; the increase in tuberculosis morbidity during adolescence is made more dramatic by the fact that the overall annual risk of infection with *M tuberculosis* declined during the lifetime of the birth cohort. Similar age-specific patterns have been found in other settings (e.g. Massachusetts, Finland, and New Zealand)

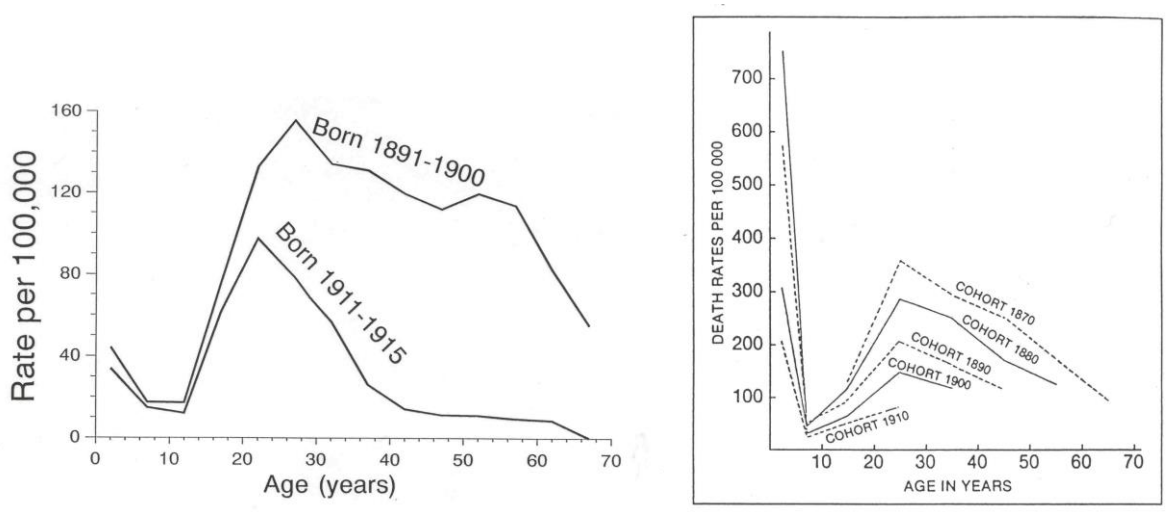


Figure 2: Mortality rates from respiratory tuberculosis plotted by birth cohort for England and Wales[3] and Massachussetts[4-5]

Age also determines whether or not individuals are infectious, as illustrated in Figure 3.

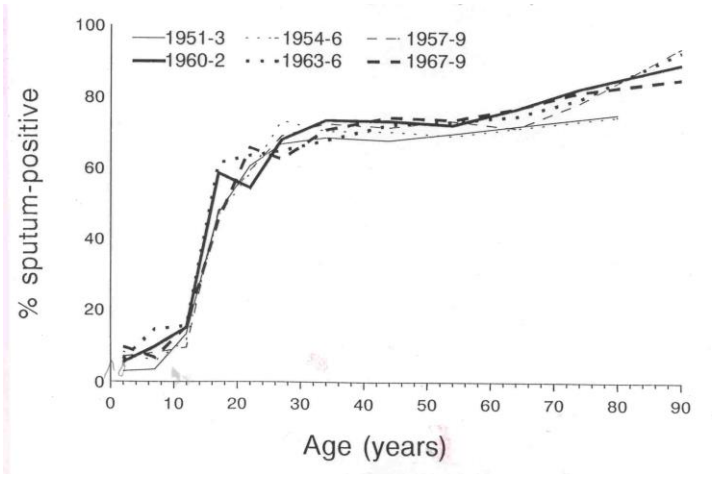


Figure 3: Proportion of respiratory tuberculosis cases in Norway who were found to be infectious (sputum smear/culture positive) between 1950 and 1970 (source Bjartveit and Styblo)

3. The risk of disease depends on the time since infection. This is illustrated in Figure 4, which shows the proportion of individuals who developed disease at different time intervals after "infection" (defined as conversion to tuberculin positivity) during a major BCG trial during the 1950s in the UK[6].

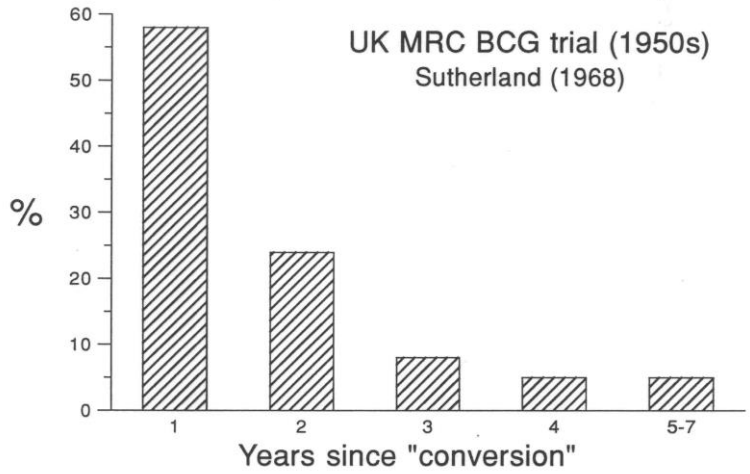


Figure 4: Distribution of the time interval between tuberculin conversion ("infection") and disease onset among individuals during the UK MRC BCG trial during the 1950s[6].

5. Contact between individuals has changed dramatically during a time period equivalent to the serial interval (time interval between successive cases in a chain of transmission). This is illustrated in Figure 5, which shows the ratio between predictions of the prevalence of infectious tuberculosis cases and the annual risk of infection in The Netherlands. This ratio is interpretable as the number of individuals infected by each tuberculosis case during the infectious period and suggests that, at least in some western countries, an infectious case may have contacted about 20 others (on average) during the 1950s, but relatively few in recent years.

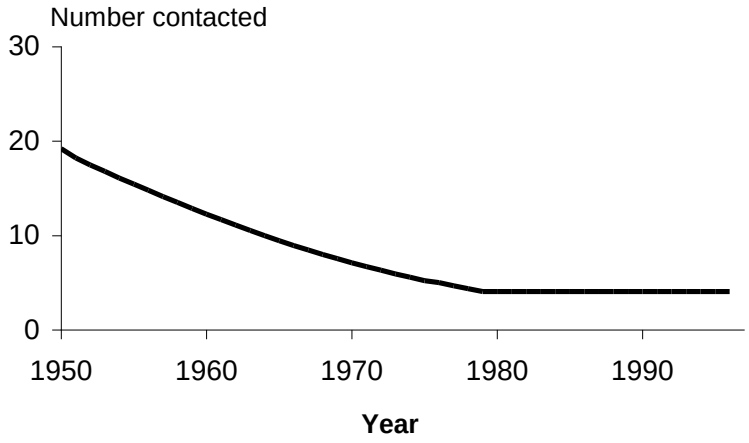


Figure 5: Estimates of the number of individuals contacted by each infectious case during his/her infectious period in The Netherlands [7-8]

More recently, the epidemiology of tuberculosis has changed again in industrialized populations, with a large proportion of cases in industrialized settings occurring among the homeless and disadvantaged individuals, and immigrants from high tuberculosis incidence settings. Thus the overall disease incidence will depend on contact between these and other individuals.

6. The epidemiology of tuberculosis differs between developed and developing countries, as illustrated by the fact that BCG vaccination protects in some areas, and not in others[9]. It is likely that the risks of developing disease also differ between settings, though this question has received little attention to date. The HIV epidemic has also altered the epidemiology of tuberculosis: individuals who are infected with HIV have an increased risk of developing tuberculosis, though its effect on the risks of developing disease soon after initial infection, reinfection or through reactivation are unknown.

Implications of these complications for the basic and net reproduction numbers for TB

As you saw in earlier lectures, the basic and net reproduction numbers are probably the most important parameters underlying the transmissibility of an infectious agent and are were critical for developing models of the transmission dynamics of infections.

For simple immunizing infections, the basic reproduction number can be calculated as:

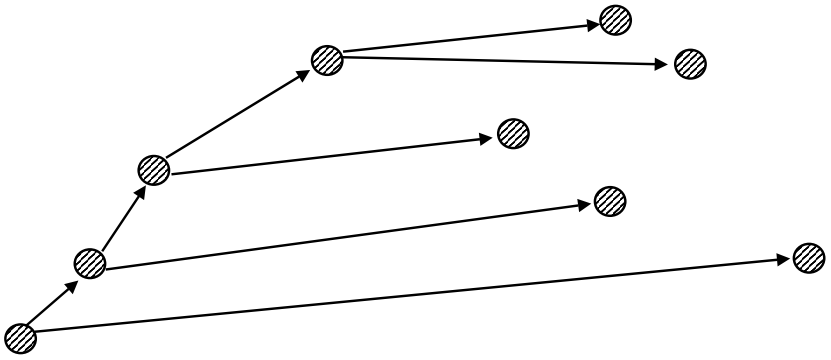
$$R_0 = \beta \times \text{“Size of the susceptible population”} \times \text{“duration of infectiousness”}$$

For tuberculosis, the calculation of the basic reproduction number is made complicated by the facts that:

- The probability that 2 specific individuals come into contact per unit time, β, has probably changed over time
- The definition of a susceptible population has to be refined to account for the occurrence of reinfection
- The duration of infectiousness can be long and variable; in addition, individuals can develop disease more than once.

As a result of these complications, the basic reproduction number has probably changed over time...
For simple immunizing infections, the magnitude of the net reproduction number correlates with the trend in disease incidence. However, for tuberculosis, since individuals can develop disease many years after initial infection, the current value of the net reproduction number is more likely to correlate with the trend in incidence many years from now and not with its current trend. Thus the net reproduction number could be equal to or above 1 at the same time as the disease incidence is in decline.

Exercise:



Question: What is the net reproduction number in this population? What is the trend in incidence?

In summary, it is possible to calculate "basic" and "net" reproduction numbers for tuberculosis; however, the interpretation of these statistics may be unconventional (and not always useful...)

Major areas of application of models for the transmission dynamics of *M tuberculosis*

Despite the above complications, many models of the transmission dynamics of *M tuberculosis* have been developed. They have been applied in the following areas:

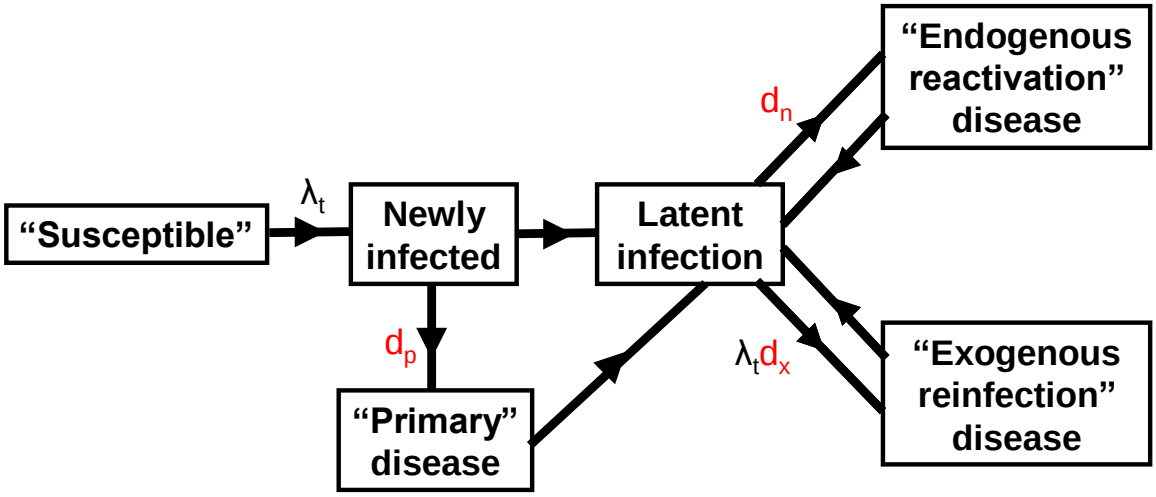
1. Estimating the relative contribution of disease attributable to recent transmission
2. Interpreting the decline in tuberculosis in developed countries
3. Modelling the impact of control strategies

We describe these areas below.

1. Estimating the relative contribution of disease attributable to recent transmission

There have been several studies which have used mathematical approaches to estimate the relative importance of disease attributable to recent transmission. The earliest study was that of Sutherland et al during the 1970s[10-11]. Using data on the annual risk of infection in The Netherlands, the study estimated the proportion of individuals in 5 year age groups in the range 15-69 years in each year between 1951 and 1960 who had been recently infected or reinfected, and the proportion who would have been (re)infected many years previously.

The following shows the general structure of the model (the parameters are defined below):



For a given annual risk of infection λ_t at time t, these proportions can be calculated roughly as follows:

$e_{a,t}$ = Proportion already infected by age a at time t =
1 – (proportion who escaped infection by age a at time t)
$= 1- (1- \lambda_{t-a})(1- \lambda_{t-(a-1)})(1- \lambda_{t-(a-2)}) \ldots (1- \lambda_t)$

$p_{a,t}$ = Proportion newly infected at age a at time t =
risk of infection at time t ×(proportion who escaped infection by age a)
$= \lambda_t (1- e_{a,t})$

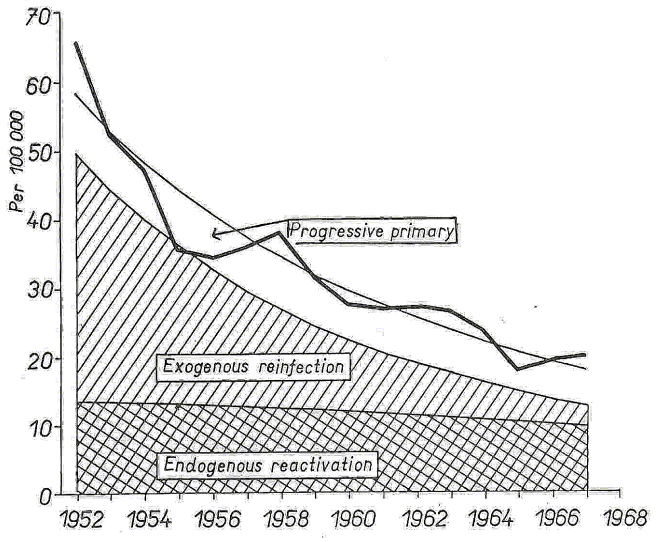
$x_{a,t}$ = Proportion newly reinfected at age a at time t =
risk of infection at time t ×(proportion already infected by age a at time t)
$= \lambda_t e_{a,t}$

The overall disease incidence in a given age group a in any given year t can then be expressed as the values of these proportions, multiplied by the population size in age group a at time t ($n_{a,t}$) and the risks of developing the different forms of disease soon after initial infection (d_p), through reactivation (d_n) and following reinfection (d_x):

$$n_{a,t} (p_{a,t} d_p + e_{a,t} d_n + x_{a,t} d_x)$$

By fitting model predictions of the disease incidence to the observed notifications in The Netherlands, Sutherland et al estimated the risk of developing disease during the 5 years after initial infection to be ~22%, whereas the annual risk of developing disease through reactivation was 0.025% and the cumulative 5 year risk of disease following reinfection was ~5%. The work was an important contribution to the epidemiology of tuberculosis, in that it provided the first ever estimates of the size of these risks and led to estimates of the relative contribution of disease attributable to recent transmission.

Fig. 2. Contribution of the three different types of infection to the total morbidity from pulmonary tuberculosis at ages 45-49 in the Netherlands from 1952 to 1967.



Source: [11]

This was especially important given that it is usually impossible to determine empirically whether or not an individual has ever been reinfected and thus there are no empirical estimates of the risk of developing exogenous disease.

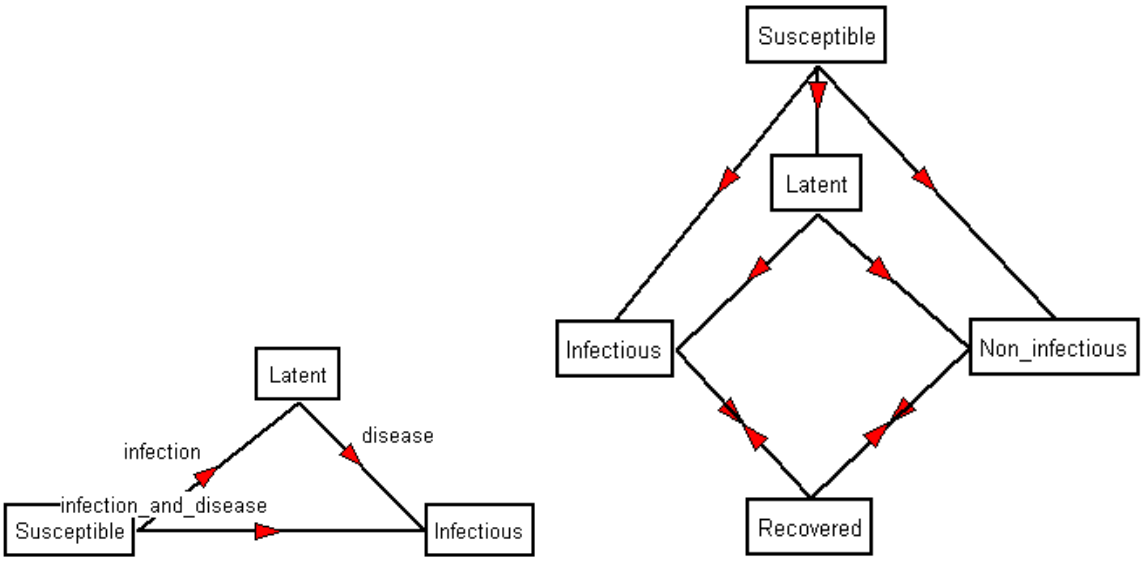
Exercise:
Using the facts that the proportion of individuals who are newly infected at a given at time t is given by $p_{a,t} = \lambda_t (1- e_{a,t})$, the proportion newly reinfected at age a at time t is given by $x_{a,t} = \lambda_t e_{a,t}$, and the annual risk of infection went down over time in the Netherlands, explain why the proportion of disease attributable to *recent infection* remained similar over time for 45-49 year olds, whereas the proportion of disease attributable to *reinfection* declined over time.

More recently, a similar approach has been used to estimate age-specific risks of developing disease following recent infection, reactivation, and recent reinfection in the UK[12]. The risks of developing disease through reactivation and following reinfection in the UK were similar to those estimated for The Netherlands, whereas as the estimated risk of developing disease following initial infection was lower than the Dutch risk, possibly as a result of differences in diagnosing TB between The Netherlands and the UK.

2. Interpreting the decline in tuberculosis in developed countries

There have been several attempts to use modelling to discuss the reason for the decline in tuberculosis in developed countries during the last century.

The first attempt, by Blower et al[13], used the following models to calculate the basic reproduction number for tuberculosis to discuss this question.



The authors concluded that the decline in tuberculosis represented part of a long decline in an epidemic spanning several centuries. The model did not stratify individuals by age and assumed that reinfection did not occur and it was not tied to any specific population. These simplifying assumptions complicated the authors' conclusions, given that reinfection would have been important at least in the early part of this century in developed countries, when the annual risk of infection with *M tuberculosis* was estimated to have been over 10% (see Figure 1).

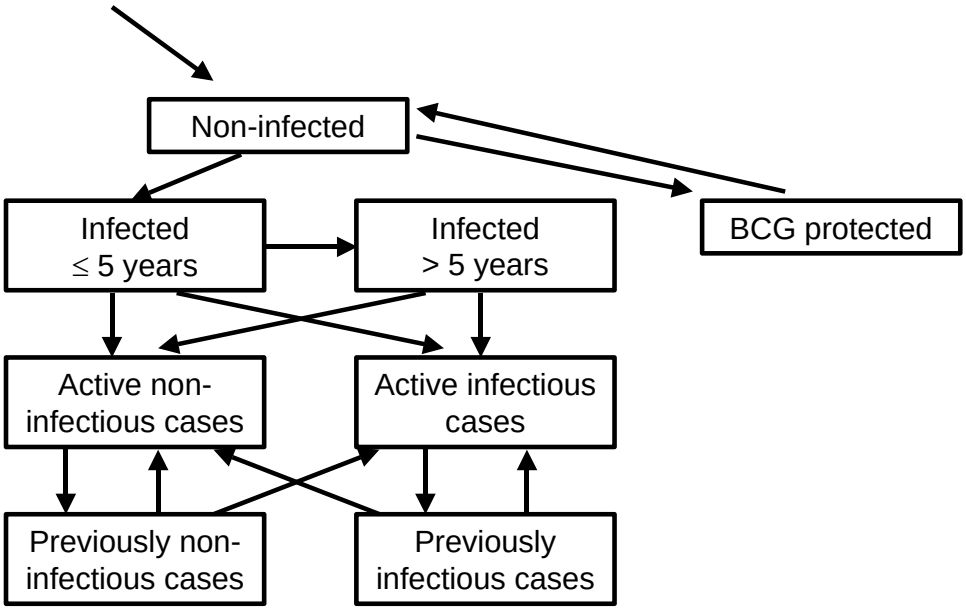
The paper highlights the problems with extending the logic which is applicable for modelling simple immunizing infections to complicated diseases such as tuberculosis, as discussed above in relation to the basic and net reproduction numbers.

Other modelling studies have concluded that the decline in tuberculosis disease incidence was attributable to a decline in contact between individuals and/or changes in the risks of disease.

3. Modelling the impact of control strategies

Before 1990

There has been much modelling work exploring the effect of different control strategies on the incidence of tuberculosis in various settings. Among the first models to do this was that of Brogger et al in 1967[14], which explored the effect of BCG vaccination, case-finding and treatment in Thailand. Other early models in this category include those of Waaler developed during the 1960[15-17]. These stratified infected individuals into those at risk of developing disease soon or many years after infection, though it did not allow individuals to develop disease following reinfection. The work evaluated the cost-effectiveness of BCG vaccination in both Norway and South India at the time when a major trial of BCG vaccination was being carried out in South India, and concluded that with a vaccine efficacy of 30%, 66% coverage among 0-20 year olds would reduce the incidence of bacillary cases to less than a third of its initial level within 40 years. The predictions were based on the assumption that individuals in South India face the same risks of developing disease in South India as those in Norway, and (more obviously) overestimated the vaccine efficacy.



1990s-mid 2000s and HIV

During the 1990s there was much important modelling work carried out by Dye et al and Murray on the effectiveness of various treatment strategies, such as DOTS and case-finding for tuberculosis and the implications of drug-resistance[18-20]. Both models were among the most detailed created until then, considering age structured populations, and allowing for reinfection to occur. The work of Dye in particular highlighted the role of DOTS, suggesting that in a setting with a stable disease incidence, a control programme which reached WHO targets of 70% case detection and 85% cure would reduce the incidence rate by 11% per year, but would have a smaller effect in settings in which the incidence has been declining for several years[18].

The HIV epidemic has complicated control of tuberculosis and several models have explored the effects of HIV on tuberculosis and the implications of control.

The first model was that of Schulzer et al, published in 1997[21-22]. The risk of developing disease among HIV-positive individuals was assumed to be about 8 times greater than that among HIV-negative individuals during each year after infection (which may be unrealistic, since, in their model, it resulted in a 42.1% risk among HIV-positive individuals during the first year after *M tuberculosis* infection). The model also did not consider the role of reinfection. The model predicted that with an annual risk of tuberculous infection of 1%, a tuberculous infection prevalence of 45% and a 2% prevalence of HIV infection in 1989, there would be a 68% increase in the number of new cases of smear-positive tuberculosis among adults by the year 2000.

The models of Dye et al[18] and Murray et al[19] which examined the impact DOTS in different settings on tuberculosis trends during the late 1990s also incorporated HIV. During the 2000s, the models of Dye at al have been extended to look at the impact of antiretroviral therapy for HIV on tuberculosis trends. One of the models[23] linked data on decline of CD4 counts to data on the risk of tuberculosis at different stages of immunosuppression and inferred that the relative risk of tuberculosis among HIV-positive individuals, as compared with those who are HIV-negative increases gradually from 1 immediately after infection with HIV, to reach about 7 seven years thereafter. Antiretroviral therapy among HIV-positive individuals was assumed to reduce the risk of tuberculosis to the level of that shortly after seroconversion.

The study concluded that a large proportion of tuberculosis cases could be prevented using antiretroviral therapy if it was provided early (e.g. shortly after HIV seroconversion) and the levels of coverage and compliance were to be high. For example, even if ARVs were provided when CD4 counts had reached 500cells/μL, the tuberculosis incidence would only be halved over a period of 20 years with an 85% effective coverage. The importance of providing ARVs as soon as possible after HIV infection was emphasized in subsequent work, which extended this model, and considered nine African countries[24].

After the mid 2000s

Since the mid 2000s, there has been much interest in the potential for meeting the global STOP TB targets of reducing TB incidence to less than 1 per million by 2050 and halving the death and mortality rates by 2015 from the level seen in 1990[25]. The effects of different interventions on the long-term TB incidence are difficult to predict using traditional epidemiological studies but have been explored using mathematical modelling. These have considered the effects of using existing technology (treatment of latent *M tuberculosis* infection, treatment of disease, antiretroviral therapy) as well as technology that has only recently become available (improved diagnostics, such as GeneXpert MTB/RIF[26-27]) and technology which has not yet become available (improved vaccines, new drug regimens to treat either infection or disease[28-29]).

Studies to date have suggested that combining interventions (e.g. vaccination, improving diagnostics and treatment) is likely to have a bigger effect than that of a single intervention[29]. However, the impact of any given intervention is likely to depend on many factors and may vary between settings.

For example, one study suggested that introducing new diagnostics, which result in cases being detected earlier than they might detected with existing diagnostics (sputum-smear and culture), could lead to substantial reductions in TB incidence [30] with 132,000 cases (95% CI: 55,000-284,000) being averted in southern Africa during the ten years after its introduction. However, the impact also depended on the underlying HIV prevalence and findings were not necessarily generalisable to other settings. The authors also noted that introducing GeneXpert MTB/RIF could also lead to increased costs, resulting from factors such as the improved survival of HIV-positive TB cases, which might occur if they are detected earlier, and increased detection of drug-resistant cases. Other studies[27] have highlighted that additional factors affect the impact, including the initial patient default after detection, treatment success.

With the emergence of new interventions for TB and interest in reducing TB incidence, such discussions are likely to continue in the future...

Further reading (optional)

Vynnycky E and White RG (2010) An introduction to infectious disease modelling. Oxford University Press. Chapter 9, pp285-296.

Summary

The natural history of tuberculosis has complicated the development of models describing the transmission dynamics of *M tuberculosis*. To date, most of the models have either

explored the relative contribution of recent transmission to the disease incidence, the reasons for the decline in tuberculosis in the past, impact of control strategies, and the impact of HIV. The accompanying practical illustrates how a simple model might be set up in Excel, taking account of the different pathways by which individuals can develop disease.

References

1. Styblo K, Meijer J, Sutherland I. Tuberculosis Surveillance Research Unit Report No. 1: the transmission of tubercle bacilli; its trend in a human population. Bull Int Union Tuberc. 1969 Aug;42:1-104.
2. Sutherland I, Bleiker MA, Meijer J, Styblo K. The risk of tuberculous infection in The Netherlands from 1967 to 1979. Tubercle. 1983 Dec;64(4):241-53.
3. Adelstein A. Mortality from tuberculosis. Population Trends. 1977;8:20-3.
4. Frost WH. The age selection of mortality from tuberculosis in successive decades. 1939. Am J Epidemiol. 1995 Jan 1;141(1):4-9; discussion 3.
5. Frost WH. The age selection of mortality from tuberculosis in successive decades. Am J Public Health 1939;30:91-6.
6. Sutherland I. The ten-year incidence of clinical tuberculosis following "conversion" in 2,550 individuals aged 14 to 19 years. . The Hague, The Netherlands: KNCV; 1968.
7. Borgdorff MW, Nagelkerke NJ, van Soolingen D, Broekmans JF. Transmission of tuberculosis between people of different ages in The Netherlands: an analysis using DNA fingerprinting. Int J Tuberc Lung Dis. 1999 Mar;3(3):202-6.
8. Vynnycky E, Nagelkerke N, Borgdorff MW, van Soolingen D, van Embden JD, Fine PE. The effect of age and study duration on the relationship between 'clustering' of DNA fingerprint patterns and the proportion of tuberculosis disease attributable to recent transmission. Epidemiol Infect. 2001 Feb;126(1):43-62.
9. Fine PE. Variation in protection by BCG: implications of and for heterologous immunity. Lancet. 1995 Nov 18;346(8986):1339-45.
10. Sutherland I, Svandova E, Radhakrishna S. The development of clinical tuberculosis following infection with tubercle bacilli. 1. A theoretical model for the development of clinical tuberculosis following infection, linking from data on the risk of tuberculous infection and the incidence of clinical tuberculosis in the Netherlands. Tubercle. 1982 Dec;63(4):255-68.
11. Sutherland I, Svandova E, Radhakrishna S. Alternative models for the development of tuberculosis disease following infection with tubercle bacilli. Bull Int Union Tuberc. 1976;51(1):171-9.
12. Vynnycky E, Fine PE. The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. Epidemiol Infect. 1997 Oct;119(2):183-201.
13. Blower SM, McLean AR, Porco TC, Small PM, Hopewell PC, Sanchez MA, et al. The intrinsic transmission dynamics of tuberculosis epidemics. Nat Med. 1995 Aug;1(8):815-21.
14. Brogger S. Systems analysis in tuberculosis control: a model. Am Rev Respir Dis. 1967 Mar;95(3):419-34.
15. Waaler HT. Model simulation and decision-making in tuberculosis programmes. Bull Int Union Tuberc. 1970 Jun;43:337-44.
16. Waaler HT, Piot MA. Use of an epidemiological model for estimating the effectiveness of tuberculosis control measures. Sensitivity of the effectiveness of tuberculosis control measures to the social time preference. Bull World Health Organ. 1970;43(1):1-16.
17. Waaler HT, Piot MA. The use of an epidemiological model for estimating the effectiveness of tuberculosis control measures. Sensitivity of the effectiveness of tuberculosis control measures to the coverage of the population. Bull World Health Organ. 1969;41(1):75-93.

18. Dye C, Garnett GP, Sleeman K, Williams BG. Prospects for worldwide tuberculosis control under the WHO DOTS strategy. Directly observed short-course therapy. Lancet. 1998 Dec 12;352(9144):1886-91.
19. Murray CJ, Salomon JA. Modeling the impact of global tuberculosis control strategies. Proc Natl Acad Sci U S A. 1998 Nov 10;95(23):13881-6.
20. Dye C, Williams BG, Espinal MA, Raviglione MC. Erasing the world's slow stain: strategies to beat multidrug-resistant tuberculosis. Science. 2002 Mar 15;295(5562):2042-6.
21. Schulzer M, Fitzgerald JM, Enarson DA, Grzybowski S. An estimate of the future size of the tuberculosis problem in sub-Saharan Africa resulting from HIV infection. Tuber Lung Dis. 1992 Feb;73(1):52-8.
22. Schulzer M, Radhamani MP, Grzybowski S, Mak E, Fitzgerald JM. A mathematical model for the prediction of the impact of HIV infection on tuberculosis. Int J Epidemiol. 1994 Apr;23(2):400-7.
23. Williams BG, Dye C. Antiretroviral drugs for tuberculosis control in the era of HIV/AIDS. Science. 2003 Sep 12;301(5639):1535-7.
24. Williams BG, Granich R, De Cock KM, Glaziou P, Sharma A, Dye C. Antiretroviral therapy for tuberculosis control in nine African countries. Proc Natl Acad Sci U S A. 2010 Nov 9;107(45):19485-9.
25. Raviglione MC, Uplekar MW. WHO's new Stop TB Strategy. Lancet. 2006 Mar 18;367(9514):952-5.
26. Dowdy DW, Chaisson RE, Maartens G, Corbett EL, Dorman SE. Impact of enhanced tuberculosis diagnosis in South Africa: a mathematical model of expanded culture and drug susceptibility testing. Proc Natl Acad Sci U S A. 2008 Aug 12;105(32):11293-8.
27. Lin HH, Dowdy D, Dye C, Murray M, Cohen T. The impact of new tuberculosis diagnostics on transmission: why context matters. Bull World Health Organ. 2012 Oct 1;90(10):739-47A.
28. Abu-Raddad LJ, Sabatelli L, Achterberg JT, Sugimoto JD, Longini IM, Jr., Dye C, et al. Epidemiological benefits of more-effective tuberculosis vaccines, drugs, and diagnostics. Proc Natl Acad Sci U S A. 2009 Aug 18;106(33):13980-5.
29. Dye C, Williams BG. Eliminating human tuberculosis in the twenty-first century. J R Soc Interface. 2008 Jun 6;5(23):653-62.
30. Menzies NA, Cohen T, Lin HH, Murray M, Salomon JA. Population Health Impact and Cost-Effectiveness of Tuberculosis Diagnosis with Xpert MTB/RIF: A Dynamic Simulation and Economic Evaluation. PLoS Med. 2012 Nov;9(11):e1001347.

Introduction to Infectious Disease Modelling and its Applications – 2018

Session 27: Applications in veterinary epidemiology: Spatial transmission and meta-population models

Lecture

This lecture will introduce the main concepts in mathematical models for the spread of veterinary diseases. The lecture will use examples from two major UK outbreaks – Bovine Spongiform Encephalopathy (BSE) in the late 1980s and early 1990s and the Foot-and-Mouth Disease (FMD) outbreak in 2001 to illustrate some of the key concepts. Additionally we will draw on examples from recent Avian Influenza outbreaks in poultry.

The concepts to be covered include:

- Within-farm models
- Between-farm models
- Reasons for each type of model
- Important veterinary terminology
- Spatial metapopulation models and spatial kernels
- Impact of animal survival
- Control of veterinary outbreaks – slaughtering of infected premises, culling, movement restrictions and ring vaccination
- Issues in the use of vaccines to control veterinary outbreaks

By the end of this lecture and the associated practical you should:

- Have an understanding of the different epidemiological units (animals, farms) used in veterinary models
- Understand when each unit would be appropriate to incorporate in a model
- Be aware of spatial metapopulation models and understand the concept of a spatial kernel
- Appreciate the importance of incorporating animal survival in models
- Understand the types of controls and how they can be implemented in the models

References:

Anderson RM, Donnelly CA, Ferguson NM et al. (1996) Transmission dynamics and epidemiology of BSE in British cattle. Nature 382 779-788.

Ferguson NM, Donnelly CA, Anderson RM (2001). The foot-and-mouth epidemic in Great Britain: pattern of spread and impact of interventions. Science 292 1155-1160.

Ferguson NM, Donnelly CA, Anderson RM (2001). Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain. Nature 413 542-548.

Keeling MJ, Woolhouse ME, Shaw DJ, Matthews L, Chase-Topping M, Haydon DT, Cornell SJ, Kappey J, Wilesmith J, Grenfell BT (2001). Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. Science 294 813-917.

Keeling MJ, Woolhouse ME, May RM, Davies G, Grenfell BT (2003). Modelling vaccination strategies against foot-and-mouth disease. Nature 421 136-142.

Tildesley MJ, Savill NJ, Shaw DJ, Deardon R, Brooks SP, Woolhouse ME, Grenfell BT, Keeling MJ (2006). Optimal reactive vaccination strategies for a foot-and-mouth outbreak in the UK. Nature 440 83-86.

Introduction to Infectious Disease Modelling and its Applications – 2018

Session 28: Sexually Transmitted Infections
Lecture

1.1 Overview and objectives

This lecture introduces you to models of sexually transmitted infections (STIs). By the end of this lecture, you should:

- Understand the important characteristics of sexually transmitted infections and how they differ from the infections modelled so far
- Use simple deterministic compartmental models to explore the transmission dynamics of short-duration curable STIs such as gonorrhoea to,
    - Explore the importance of heterogeneity in sexual activity for STI invasion and endemic prevalence (insights from the *'Hethcote-Yorke'* models)
    - Appreciate the importance of mixing patterns on R_0, the rate of STI spread, the equilibrium STI prevalence and the utility of '*Q',* a summary measure of mixing
    - Explore the importance of heterogeneity in sexual activity and mixing patterns for STI control
    - Appreciate the similarity between a heterosexual STI model and a host-vector infection model

1.2 Characteristics of sexually transmitted infections

There are some important differences between STIs and the infections modelled so far. In contrast with the directly transmitted respiratory infections modelled in earlier sessions, when modelling infections that require intimate contact such as STIs (Holmes, Sparling et al, 2008), the rate at which new infections are generated does not depend on the population density (Hethcote and Yorke, 1984). Close crowding of individuals does not necessarily lead to increases in the rate at which individuals make sexual contact whereas it is likely to increase the rate at which individuals make contacts which are sufficient to transmit respiratory infections. Also, the population at risk of an STI is a subset of the community. For a purely sexually transmitted infection, transmission will only occur between sexually active individuals and so these groups will need to be identified. However, data on sexual behaviours are often difficult to collect and subject to significant bias and therefore need to be interpreted carefully (Fenton, Johnson et al, 2001). There is often great heterogeneity in the sexual activity of individuals within and between different populations which will be important for predicting the spread and control of STIs (Schneeberger, Mercer et al, 2004).

The natural history of infection differs markedly between different STIs (Holmes, Sparling et al, 2008). STIs are also more likely to be asymptomatic in women than in men, which has implications for identifying control strategies. Some short-duration curable STIs, such as gonorrhoea, chancroid and chlamydia infection, don't induce a significant level of immunity in the host, while long duration syphilis infection results in latent disease and may induce immunity (Holmes, Sparling et al, 2008). Before widespread treatment with antibiotics syphilis infection in adults was a significant cause of death. A more important cause of death nowadays worldwide is the relatively new STI, HIV, which kills in the absence of effective treatment and has become pandemic (Coffin, Haase et al, 1986; Holmes, Sparling

et al, 2008). Some STIs enhance the probability of the transmission of other STIs (Fleming and Wasserheit, 1999; Laga, Nzila et al, 1991; Rottingen, Cameron et al, 2001; Wasserheit, 1992). Most work in this area has focussed on the transmission enhancing effects on HIV of these other STIs, particularly those that cause genital ulcers such as chancroid, syphilis and HSV-2 infection. Such STIs are referred to as *cofactor* STIs.

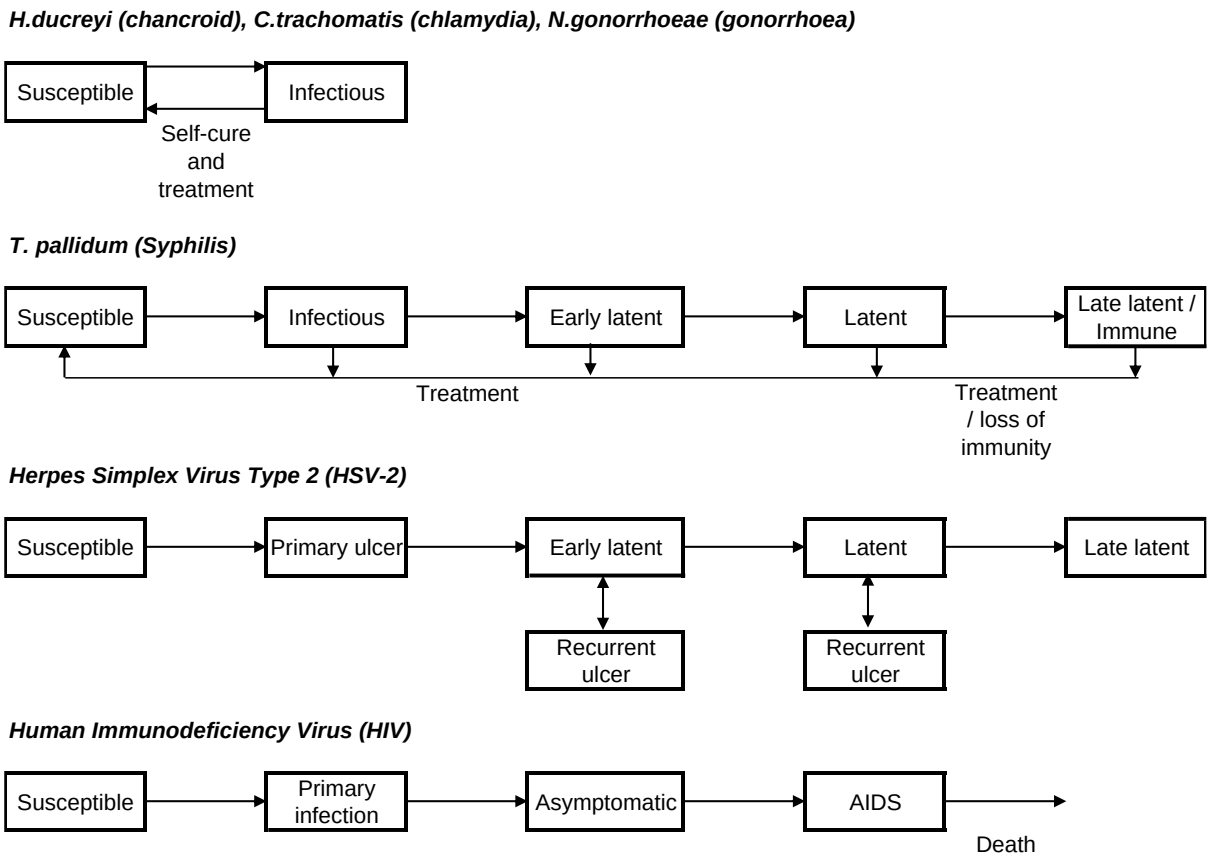


Figure 1-1 Examples of the natural history of key STIs assumed in modelling studies (Based on White, Orroth et al. 2004)

These characteristics need to be considered when deciding how to set-up models of STI transmission. Figure 1-1 gives examples of the natural history of key STIs that have been assumed in modelling studies. Although model representations are typically based on extensive literature review, some arbitrariness remains in the categorisation of STI stages (particularly “early latent”, “latent” and “late latent”).

An important period in the modelling of STIs occurred during the 1970 and 1980s with the work of Cooke and Yorke, who were the first to develop a mathematical model of gonorrhoea transmission that incorporated many of the characteristics described above (Cooke and Yorke, 1973). In this paper and in a subsequent monograph (Hethcote and Yorke, 1984), they used deterministic compartmental models of the transmission dynamics of gonorrhoea, together with data on the dramatic increase in the reported number of gonorrhoea cases between 1950 and 1980 in the US (Figure 1-2), to explore the transmission dynamics of gonorrhoea and evaluate possible control strategies (Hethcote and Yorke, 1984).

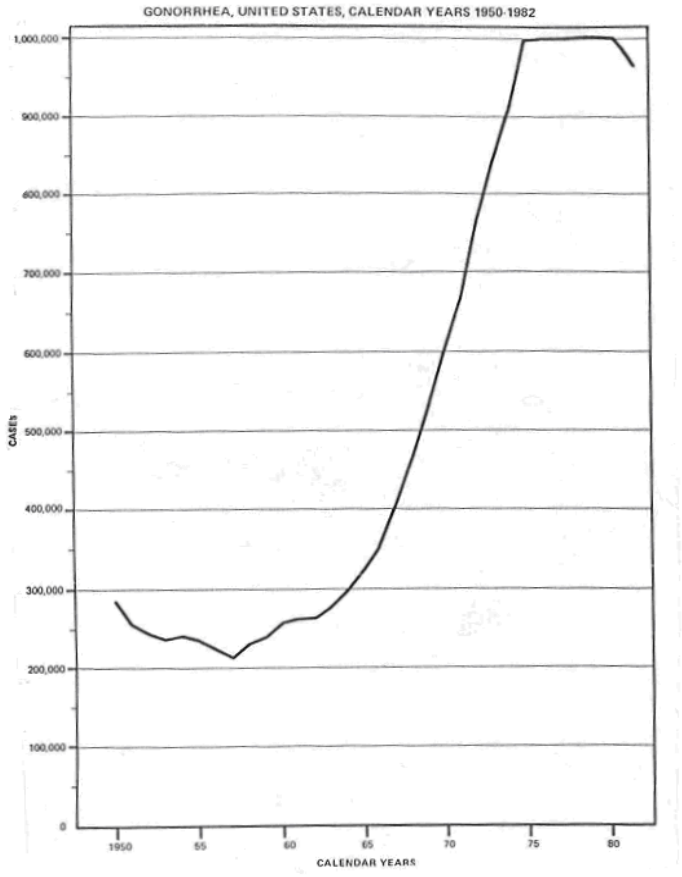


Figure 1-2 Reported cases of gonorrhoea in the US from 1950 to 1982. Cited in (Hethcote and Yorke, 1984). Original source: Statistical Services Section, Division of Venereal Disease Control, Centers for Disease Control, USA.

There was great concern about how best to control this rapid increase. This rise had been attributed to increases in sexual risk behaviour and the introduction of hormonal contraception, replacing condom use, but was probably also partly due to improved surveillance. The elegant and well parameterised modelling studies by Hethcote and colleagues provided important insights into the transmission and control of gonorrhoea in the US and were used to inform US STI control policy.

We begin by describing some of Hethcote and Yorke's models, the insights these models provided, and illustrate how extensions to these models have led to an improved understanding of the transmission and control of STIs.

1.3 What prevalence of gonorrhoea infection might we expect? Insights from the Hethcote-Yorke model

Based on the data in Figure 1-2 and an estimate of the size of the sexually active population in the US, it was estimated that the prevalence of individuals infected with *N. gonorrhoea,* the bacteria that causes gonorrhoea, in the sexually active population was around 2%.

Let us start our exploration of the transmission dynamics of gonorrhoea by seeing if we can reproduce this estimated prevalence using a transmission model. To do this, we shall (as Hethcote and Yorke did) first consider the simplest model possible, describing a population that consists of individuals that are not differentiated by gender, age, sexual behaviour or any other characteristic except whether or not they are infected. We assume all infected individuals are infectious (Figure 1-3).

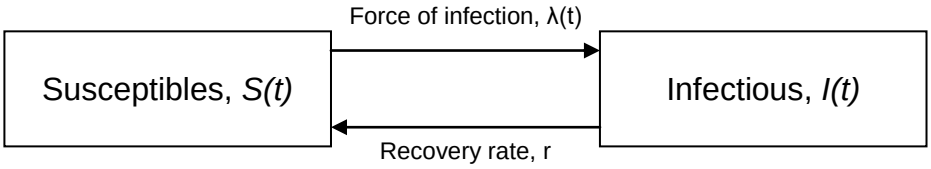


Figure 1-3: A simple model of STI transmission dynamics

The equations for our simple model of STI transmission in Figure 1-3 are

$$\frac{dS(t)}{dt} = -\lambda(t)S(t) + rI(t) \qquad \text{Equation 1.1}$$

$$\frac{dI(t)}{dt} = +\lambda(t)S(t) - rI(t) \qquad \text{Equation 1.2}$$

Where *S(t)* and *I(t)* are the numbers of susceptible and infectious individuals at time t, respectively, *λ(t)* is the force of infection at time *t*, and *r* is the recovery rate per year, assumed equal to *1/D* where *D* is the duration of infection in years. The population size, *N,* equals the sum of the number of susceptible and infectious individuals. We assume the population is closed and therefore *N* does not change over time, ie *N = S(t) + I(t)*

Let us also assume random mixing and assume that each individual changes sexual partners at a constant rate per year, *c* , and the probability of transmission of the STI during a sexual partnership between STI discordant partners is β_p,

In earlier lectures where we modelled infections transmitted via the respiratory route, we described two formulations for the force of infection expression assuming *density* or *frequency dependence*

$$\lambda(t) = \beta I(t) \qquad \textit{density dependence (pseudo mass action)} \qquad \text{Equation 1.3}$$

$$\lambda(t) = c_e \frac{I(t)}{N(t)} \qquad \textit{frequency dependence (true mass action)} \qquad \text{Equation 1.4}$$

where *β* was defined differently as the rate at which two specific individuals come into effective contact per unit time and c_e was the average number of individuals effectively contacted by each person per unit time.

Assuming *density dependence* means we assumed the force of infection increases with increased population size while assuming *frequency dependence* means we assumed the force of infection does not change with population size. For STIs and other infections that require intimate contact between individuals, increasing the number of individuals in a room

(for example) won't increase the rate at which individuals change sexual partners, and therefore the *frequency dependence* assumption is the most appropriate.

Furthermore, the two implicit components of the transmission parameter ('β ' in early sessions) - the contact rate and the transmission probability per contact - are usually separated into two separate parameters when modelling STIs. This is primarily because, for STIs, it's believed the contact rate can be measured more easily than for respiratory infections by, for example, asking the question '*How many sexual partners have you had in the last 12 months?'.*

STI transmission can be modelled at the *per-act* or the *per-partnership* level. Modelling at the *per-act* level, the correct contact rate parameter is the coital frequency per unit time and the correct transmission probability parameter is the transmission probability per sex-act β_a. Modelling at the *per-partnership* level, the correct contact rate parameter is the partner change rate, *c,* and the correct transmission probability parameter is the transmission probability per partnership β_p. In this session we model STIs at the per-partnership level as this is most common, but all the models can be reformulated at the per-act level. The relationship between β_a and β_p is discussed in more detail in Panel 1-1.

If we assume random mixing, the equation for the force of infection $\lambda(t)$ is defined as the product of the partner change rate, *c*, the transmission probability per partnership, β_p, and the prevalence of infection at time *t* , $\frac{I(t)}{N}$.

$$\lambda(t) = c\,\beta_p \frac{I(t)}{N} \qquad \text{Equation 1.5}$$

Using Equation 1.5 and the equations for our simple model of STI transmission (Equation 1.1 and Equation 1.2) we can now make predictions for the spread of gonorrhoea in our population.

However, if we assume plausible parameter values, an average partner change rate, *c,* of *2* partners per year, a transmission probability per partnership, β_p *,* of 0.75, and a duration of infection with some treatment services in the population*, D,* of 2 months (0.167 years) we quickly encounter problems because gonorrhoea fails to spread (Figure 1-4).

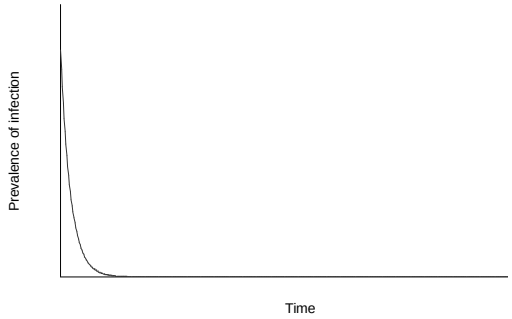


Figure 1-4 Prediction of the prevalence of infectious individuals over time after the introduction of one infectious case assuming β_p = 0.75, *D* = 0.167 years, *c* = 2 partners/year.

To understand why this occurs we can calculate the R_0 for this infection in this population. The R_0 for this infection in this population is derived in Panel 1-2 and is given by the expression:

$$R_0 = c\beta_p D \qquad \text{Equation 1.6}$$

Now we can see why gonorrhoea failed to spread. R_0 was well below one.

$$R_0 = 2 \times 0.75 \times 0.167$$
$$= 0.25$$

Note that the level of sexual activity that is necessary to raise R_0 above one ($c \geq 1 / \beta_p D$) will differ for different STIs because β_p *and* D vary between STIs (Boily and Masse, 1997).

By rearranging Equation 1.6 we can see that if we assume $\beta_p = 0.75$, if gonorrhoea is to invade the population we would need to assume an average partner change rate in the whole sexually active population of 8 partners per year, or duration of infection of 8 months, or intermediate values for both. These relatively high parameter values are unlikely.

The primary reason we are failing to predict that gonorrhoea will spread in population, is not the parameter values we are assuming, but that we are ignoring one of the important characteristics listed above - heterogeneity in human sexual behaviour. Surveys of the sexual behaviour of STI clinics attendees and the general population show that over any period of time most people have relatively low numbers of sexual partners but some have much higher numbers (Figure 1-5). This heterogeneity is very important for the understanding the spread and control of STIs.

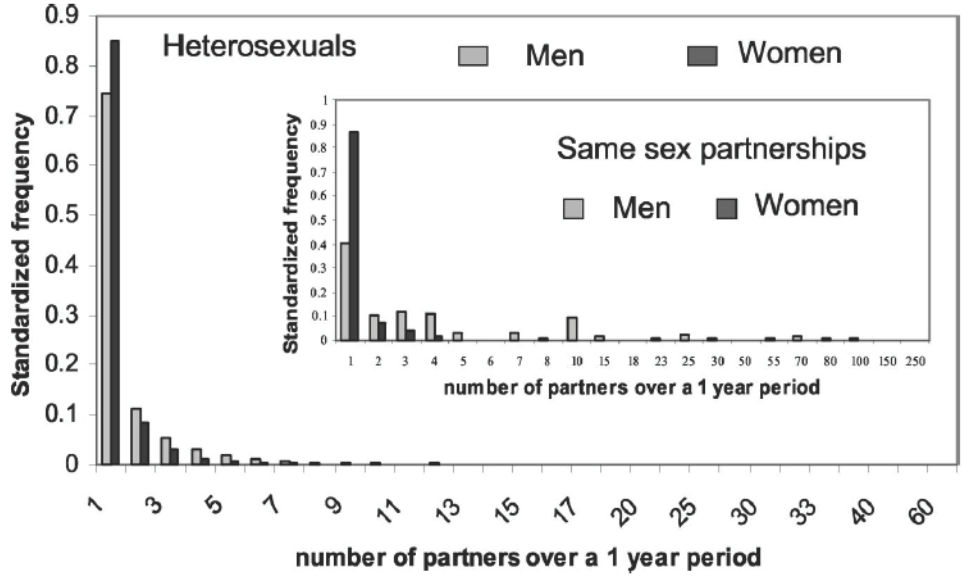


Figure 1-5 The distribution of reported annual number of partners for heterosexual and homosexual men and women in Britain from the National Survey of Sexual Attitudes and Lifestyles, 2000. The plots show the relative number of partners and excludes individuals who report zero partners (Schneeberger, Mercer et al, 2004).

1.4 The importance of heterogeneity in sexual activity for STI transmission dynamics

We can explore the implications of heterogeneity in sexual activity for the spread of a short-duration STI by adapting our model. In line with the findings of Hethcote and Yorke, modelling heterogeneity in sexual activity allows us to predict that gonorrhoea will invade our population while assuming more plausible values for the rate at which individuals change partners than we had to assume using the simple model shown in Figure 1-3. We first describe the key equations and assumptions for the model.

1.4.1 Incorporating risk heterogeneity in the Hethcote and Yorke model – key assumptions

We split the population into two groups based on the frequency at which individuals change their partners. More highly sexually active individuals are members of the high-activity group and less active individuals are members of the low-activity group.

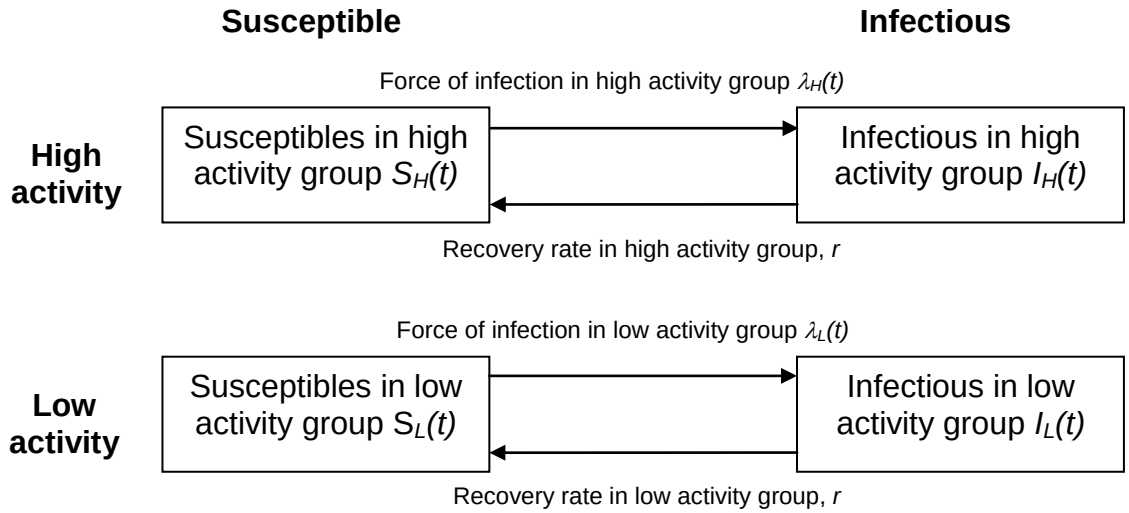


Figure 1-6 Model of gonorrhoea transmission incorporating heterogeneity in activity

The equations for rate of change of the number of susceptible and infectious individuals per unit time in this model are very similar to Equation 1.1 and Equation 1.2:

$$\frac{dS_j(t)}{dt} = -\lambda_j(t)S_j(t) + rI_j(t) \qquad \text{Equation 1.7}$$

$$\frac{dI_j(t)}{dt} = +\lambda_j(t)S_j(t) - rI_j(t) \qquad \text{Equation 1.8}$$

Where $j = {}_H$ for the high-activity group and $j = {}_L$ for the low-activity group

To maintain comparability with our simple model (Figure 1-3) we keep the total number of partnerships formed per year equal in both models but increase the difference (heterogeneity) between the partner change rates in the high and low-activity groups (see

Panel 1-3 for details of how to do this). Following *Hethcote* and *Yorke*, we assume 2% of the population belong to the high-activity group and 98% belong to the low-activity group. Therefore if we assume the partner change rate in the low-activity group, c_L , is 1.4 partners per year, then, to keep the mean rate of partner change at 2 partners per year the partner change rate in the high-activity group, c_H , has to be 31.4 partners per year.

Initially we assume *proportionate mixing*, ie partners are chosen with a probability that is proportional to the number of partnerships that they generate (Garnett and Anderson, 1994), and also assume that the other parameter values are unchanged, so β_p = 0.75 and D = 0.167 years. Also, for simplicity, we assume that the per-partnership transmission probability is the same in high and low-activity groups (as illustrated in Panel 1-1), although in reality, a high-activity individual is likely to have fewer acts per partnership and therefore β_p may be lower.

The force of infection experienced by a high-activity group member will be higher than for a low-activity group member because of their higher partner change rate (Equation 1.9). However, because we are assuming proportionate mixing, each time a new partner is selected both high and low-activity group members will have the same probability that they select an infected partner, *p(t)*

$$\lambda_j(t) = c_j \beta_p p(t) \qquad \text{Equation 1.9}$$

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

Note that, *p(t),* the probability that a partner selected according to proportionate mixing is infectious, is not the overall prevalence of infectious individuals in the population ($I(t)/N$) as it was for the simple model, but 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.

$$p(t) = g_H \times i_H(t) + g_L \times i_L(t) \qquad \text{Equation 1.10}$$

where g_H and g_L are the probabilities that a partner selected according to proportionate mixing will be a member of the high or low-activity group, respectively, and $i_H(t)$ and $i_L(t)$ are the prevalence of infectious individuals in the high and low-activity groups, respectively. $i_H(t) = I_H(t)/N_H$ and $i_L(t) = I_L(t)/N_L$ where N_H and N_L are the total number of individuals in the high and low-activity groups, respectively.

Panel 1-4 shows the probability that a partner selected according to proportionate mixing will be a member of the high or the low-activity group, is equal to the proportion of the total number of partnerships, generated by each group each year. In our example, this means 31% of partnerships are generated by high-activity group members ($g_H = 0.31$) and 69% are generated by low-activity group members ($g_L = 0.69$).

1.4.2 Calculating R_0 in a population with heterogeneity in sexual activity

To estimate whether gonorrhoea will invade the population we first need to calculate the number of secondary infections generated in a totally susceptible population by an infected high-activity, R_H, or low-activity, R_L, group member:

$$R_H = c_H \beta_p D = 31.4 \times 0.75 \times 0.167 = 3.93 \quad \text{Equation 1.11}$$

$$R_L = c_L \beta_p D = 1.4 \times 0.75 \times 0.167 = 0.18 \quad \text{Equation 1.12}$$

Therefore, at invasion, around 22 times more infections will be caused by each infected high-activity group member than by each infected low-activity group member.

However, we can't tell from these values whether gonorrhoea will invade because R_H is above one, and R_L is below one.

To find out if gonorrhoea will invade our single-gender model with heterogeneity in activity and proportionate mixing we need to calculate R_0. As described in earlier sessions, we can calculate R_0 using several approaches and these are shown in Panel 1-5 and Appendix 1.1.

Panel 1-5 shows that R_0 is the weighted average of the number of secondary infections from high-activity and low-activity members, where the weights, g_H (0.31) and g_L (0.69), are the probabilities that a partner selected according to proportionate mixing will be a member of the high or low-activity group respectively.

$$R_0 = g_H R_H + g_L R_L = 0.31 \times 3.93 + 0.69 \times 0.18 = 1.36 \quad \text{Equation 1.13}$$

Importantly, this means that by introducing heterogeneity in sexual activity into our model, while we have kept the mean rate of partner change per year constant, we have been able to predict that gonorrhoea will invade the population (Figure 1-7).

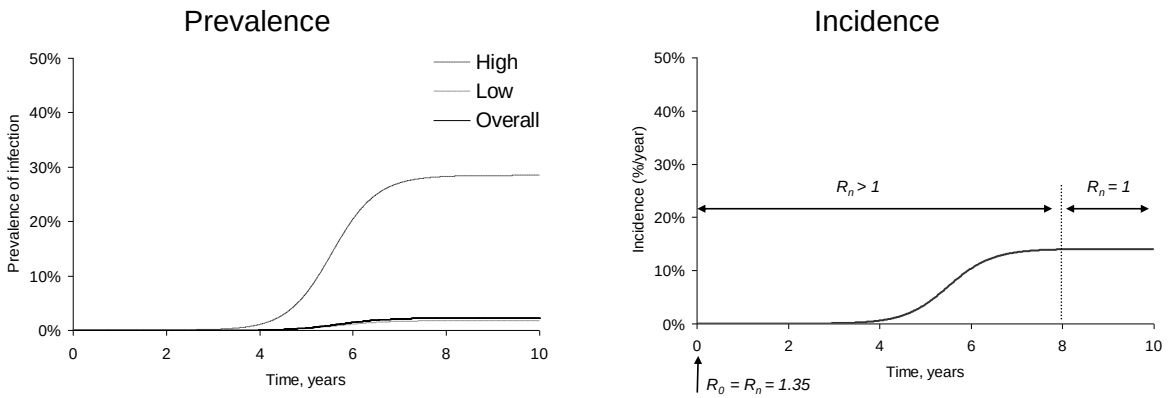


Figure 1-7 Predicted prevalence and overall incidence of gonorrhoea in a model with heterogeneity in sexual activity. The right hand figure also shows the relationship between the basic (R_0) and net (R_n) reproduction numbers. β_p = 0.75, D = 0.167 years, c_L = 1.4 partners/year, c_H = 31.4 partners/year. 2% of the population belong to the high-activity group.

Figure 1-7 also shows that introducing heterogeneity in sexual activity has also enabled us to predict a low overall infection prevalence of 2.3% that is in line with the 2% prevalence estimated by Hethcote and Yorke for the sexually active US population in the 1980s. This was achieved using parameter values for the overall sexual behaviour of the population that are much more plausible than those used in the model that did not incorporate heterogeneity in sexual activity. Other heterogeneities also make it more likely infections will persist (Anderson and May, 1984). To explore this further see (Garnett and Anderson, 1993).

Figure 1-8 shows more generally what happens if we increase the difference between the partner change rates in the high and low-activity groups. Once heterogeneity increases above a critical level, R_0 rises above one and gonorrhoea can invade. In our example this occurs when the partner change rate is around 1.5 per year in the low-activity group and therefore around 26 per year in the high-activity group.

Figure 1-8 also shows that if we continue to increase heterogeneity in sexual activity then gonorrhoea prevalence in the overall population first increases and then decreases. In our example, the overall prevalence peaks at around 0.8 and 61 partners per year in the low and high-activity groups respectively, after which further increases in heterogeneity leads to a reduction in the overall prevalence.

This is because the impact of increasing heterogeneity on reducing the partner change rate in the low-activity group starts to exceed the impact of increasing heterogeneity on increasing the prevalence of infection in the high-activity group.

In the extreme case in which the partner change rate in the low-activity group becomes zero, gonorrhoea becomes extinct in the low-activity group.

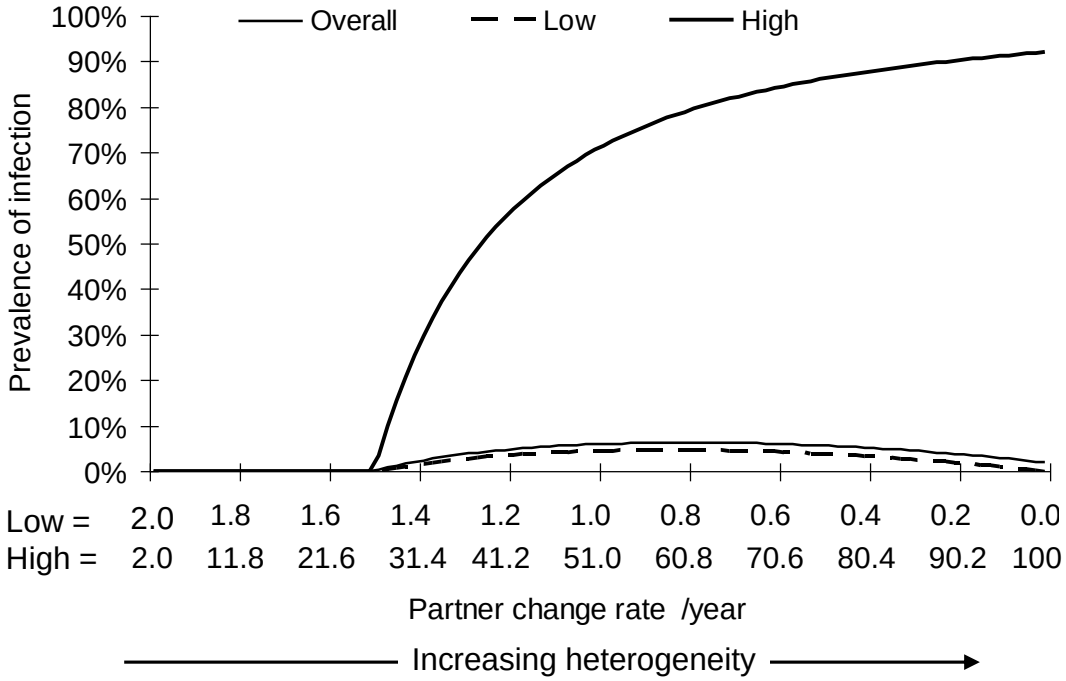


Figure 1-8 The effect of increasing heterogeneity in sexual activity in a population on equilibrium STI prevalence. The mean rate of partner change remains constant at 2 partners per year. $\beta_p = 0.75$, $D = 0.167$ years. 2% of the population belong to the high-activity group.

Heterogeneity in sexual activity is also important in explaining why the infection can be maintained at a low and relatively stable prevalence in the population as a whole. In the absence of immunity, contact with already-infected individuals (called *pre-emptive saturation*), is the only mechanism to prevent the prevalence of infection rising to 100% (see Panel 1-6). However, in a population without heterogeneity in sexual activity and low infection prevalence, the proportion of potentially effective contacts that would be 'wasted' on infected individuals would be very low, and therefore it is an implausible mechanism for limiting transmission. However as we have seen, heterogeneity in sexual activity means that prevalence is very much higher in the high-activity group than in the overall population (Figure 1-8) and therefore in the high-activity group the effects of saturation are much stronger. In our example the prevalence is around 30% in the high-activity group and therefore nearly 1 in 3 contacts between infectious individuals and high-activity group members will be 'wasted' on already-infected individuals. As such, *pre-emptive saturation* becomes a much more plausible mechanism to limit transmission.

More generally, what happens after the infection invades a population depends not only on the value of R_0 but also on whether infection is followed by re-susceptibility, immunity or death. For infections such as gonorrhoea that do not invoke an effective immune response in the host, infection prevalence rises steadily to its equilibrium level (Figure 1-7). As this happens the net reproduction number, R_n , falls from its initial value of R_0 (1.36 here) to 1 when equilibrium prevalence is reached (Figure 1-7). Whereas, if infection leads to immunity or death the prevalence of infected and infectious individuals can rise and then fall..

1.4.3 Proportion of infections due to the high-activity group

As we have seen, although they make up only 2% of the population, the high-activity group generates 31% of the partnerships in the population, suggesting they may generate 31% of new infections, but they actually generate much more because they are also more likely to be infected than low-activity individuals. We can calculate the proportion of all new cases that are generated by high-activity group members when gonorrhoea is in equilibrium using an equation derived by *Hethcote* and *Yorke* in their monograph (Hethcote and Yorke, 1984):

$$\begin{aligned} h_{eq} &= \frac{g_H i_H(\infty)}{g_H i_H(\infty) + g_L i_L(\infty)} \\ &= \frac{0.31 \times 28.4\%}{0.31 \times 28.4\% + 0.69 \times 1.7\%} \\ &= 88\% \end{aligned}$$

Equation 1.14

Where $i_H(\infty)$ and $i_L(\infty)$ are the equilibrium prevalence of gonorrhoea infection in the high and low-activity group, respectively. These values of $i_H(\infty)$ and $i_L(\infty)$ used in Equation 1.14 were obtained from the model (see in Figure 1-7).

So in our example, the high-activity group is only 2% of the population, but it generates 31% of the partnerships and 88% of new infections (Figure 1-9), highlighting that high-activity individuals are a priority population for STI control efforts because they are at high-risk of STI infection and also for onward STI transmission.

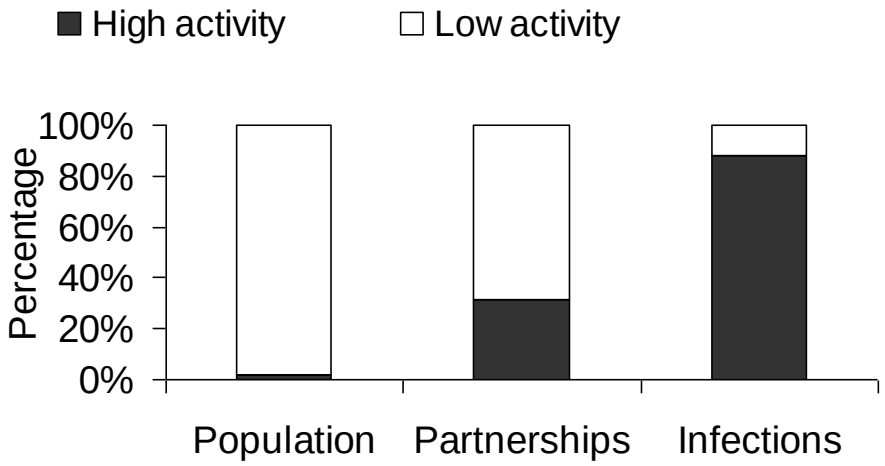


Figure 1-9 Percentage of population, sexual partnerships and infections generated by the high and low-activity groups in simple SIS model of the transmission of gonorrhoea assuming proportionate mixing, heterogeneity in sexual behaviour and the parameter values shown in the text.

1.5 Mixing by sexual activity

In earlier sessions we saw that the degree to which the various groups within a population contact each other strongly determines the impact of interventions designed to control respiratory infections. This is also the case for sexually transmitted infections.

1.5.1 Mixing patterns and mixing matrices

Up until now we have been assuming *proportionate* mixing by sexual activity, but this is only one possibility. Mixing between population groups can be broadly categorised into three types*:*

- *Proportionate mixing*, so called because partnerships are formed between population groups based on the proportion of all partnerships generated by these groups. This is also sometimes called *random mixing*, but this can be confused with randomly selecting partners based on the proportion of *individuals* in each group. Therefore describing this type of mixing as *random* is discouraged.
- *With-like mixing,* in which individuals preferentially form partnerships with individuals who have characteristics like their own, for example those with many sexual partners preferentially choose partners who also have many sexual partners. This is also sometimes called *assortative mixing.*
- *With-unlike mixing,* in which individuals preferentially form partnerships with individuals who have characteristics unlike their own, for example those with many sexual partners preferentially choose partners with few sexual partners. This is also sometimes called *disassortative mixing.*

Mixing can be modelled between subgroups defined by any characteristic of the population that is considered important in explaining the spread or control of the infection, such as age (as you saw in earlier sessions), race (Turner, Garnett et al, 2004), or gender as we show in section 1.6.

Similar to the methods used when we considered respiratory infections, contact patterns can be summarized using *mixing matrices* in which each of the elements g_{jk} in the matrix is the probability that someone in group *k* forms a partnership with someone in group *j*. In line with the convention used in the previous sessions, the second subscript, k , refers to the 'choosing' partner and the first subscript, j , refers to the 'chosen' partner. These matrices are similar to the WAIFW matrices described in earlier sessions. They differ because the matrix elements in earlier sessions represented the probability of an effective contact and therefore included the transmission probability per contact, whereas for STIs the transmission probability per contact is typically considered separately.

For *proportionate* mixing sexual partners in a given group are selected randomly in proportion to the number of partnerships that the group generates. For this assumption the probability that someone in the high-activity group forms a partnership with someone in the high-activity group, g_{HH} , is the same as the probability that someone in the low-activity group forms a partnership with someone in the high-activity group, $g_{HL.}$ i.e. $g_{HH} = g_{HL} = g_H$. In our example g_H equals 0.31 (see Panel 1-4). Similarly, the probability that someone in the low-activity group forms a partnership with someone in the low-activity group, g_{LL} , is the same as the probability that someone in the high-activity group forms a partnership with someone in the low-activity group, g_{LH} , i.*e.* $g_{LH} = g_{LL} = g_L$. Here g_L equals 0.69 (Panel 1-4).

Therefore the proportionate mixing matrix for the partner change rates and group sizes we have assumed is:

$$\text{Partner } j \begin{matrix} \\ H \\ L \end{matrix} \begin{matrix} \text{Partner } k \\ H \quad\quad L \end{matrix} \begin{pmatrix} g_{HH} & g_{HL} \\ g_{LH} & g_{LL} \end{pmatrix} \qquad \text{Equation 1.15}$$

$$\textit{Proportionate} = \begin{matrix} \\ H \\ L \end{matrix} \begin{matrix} H \quad\quad L \end{matrix} \begin{pmatrix} 0.31 & 0.31 \\ 0.69 & 0.69 \end{pmatrix}$$

Note also that the probabilities in each column of the matrix sum to one because the partner of individual k must be chosen from either the high or low-activity group. Formally this can be written as:

$$\sum_j g_{jk} = 1 \qquad \text{Equation 1.16}$$

The mixing matrices describing purely *with-like* and purely *with-unlike* mixing are:

$$\textit{Purely with-like} \quad \begin{matrix} \\ H \\ L \end{matrix} \begin{matrix} H \quad L \end{matrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \text{Equation 1.17}$$

$$\textit{Purely with-unlike} \quad \begin{matrix} \\ H \\ L \end{matrix} \begin{matrix} H \quad L \end{matrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \text{Equation 1.18}$$

1.5.2 A summary measure of mixing - *Q*

The *degree of mixing* can be summarised using a statistic that has been given the symbol *Q* by *Gupta* and colleagues (Gupta, Anderson et al, 1989). *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 negative when partners are selected purely *with-unlike*.

Q depends on the elements of the mixing matrix that measure the proportion of partnerships that are formed within groups. These are the elements on the top-left to bottom-right diagonal of the mixing matrix and in our example these are g_{HH} *and* g_{LL} *.*

The formula to calculate *Q* is:

$$Q = \frac{\left(\sum_{j=k} g_{jk} - 1\right)}{b - 1} \qquad \text{Equation 1.19}$$

where *b* is the number of groups. *Q* is scaled so that it is equal to 0 when all partners are selected proportionately, 1 when all partners are selected purely *with-like*, and $\frac{-1}{b-1}$ when all partners are selected purely *with-unlike*. So for our two activity-group example, *b*=2 and

therefore $Q = \frac{-1}{2-1} = -1$ when partners are selected purely *with-unlike*. If there were 3 activity groups, Q would equal $Q = \frac{-1}{3-1} = -1/2$ if mixing was purely *with-unlike*.

Although commonly used, *Q* has limitations. *Q* only measures mixing within identical groups, ignoring mixing between similar groups. *Q* also equally weights mixing within different population groups regardless of the size of these groups. An alternative measure of mixing that may overcome these limitations has recently been proposed by *Keeling* and colleagues. '*q'* weights mixing between similar groups by the size of the group so that mixing within a smaller group would contribute less to the value of *q* than mixing within a larger group (Keeling and Rohani, 2008). To date, however, this measure has not been used extensively.

1.5.3 Data on mixing by sexual activity

Data on the mixing between different sexual activity groups are uncommon because data are required on the characteristics of sexual partners. The data shown in Figure 1-10 suggest that, on average, the individuals in these surveys tended to form sexual partnerships with individuals that were more similar to themselves in terms of sexual activity than the proportionate mixing assumption would predict, so *Q* is just above 0. The data summarised in the top three rows in this figure are based on data from contact tracing studies among US STI clinic patients collected towards the end of the last century. These individuals are likely to be higher risk than other members of the general population and therefore these data do not necessarily generalise to the rest of the US population. However the data summarised in the bottom row were collected in a survey of the US general population conducted in 1992 and these data also indicate slightly *with-like* mixing. That mixing by sexual activity is slightly *with-like* is intuitively plausible, as in general people tend to socialise with people who are similar to themselves (McPherson, Smith-Lovin et al, 2001).

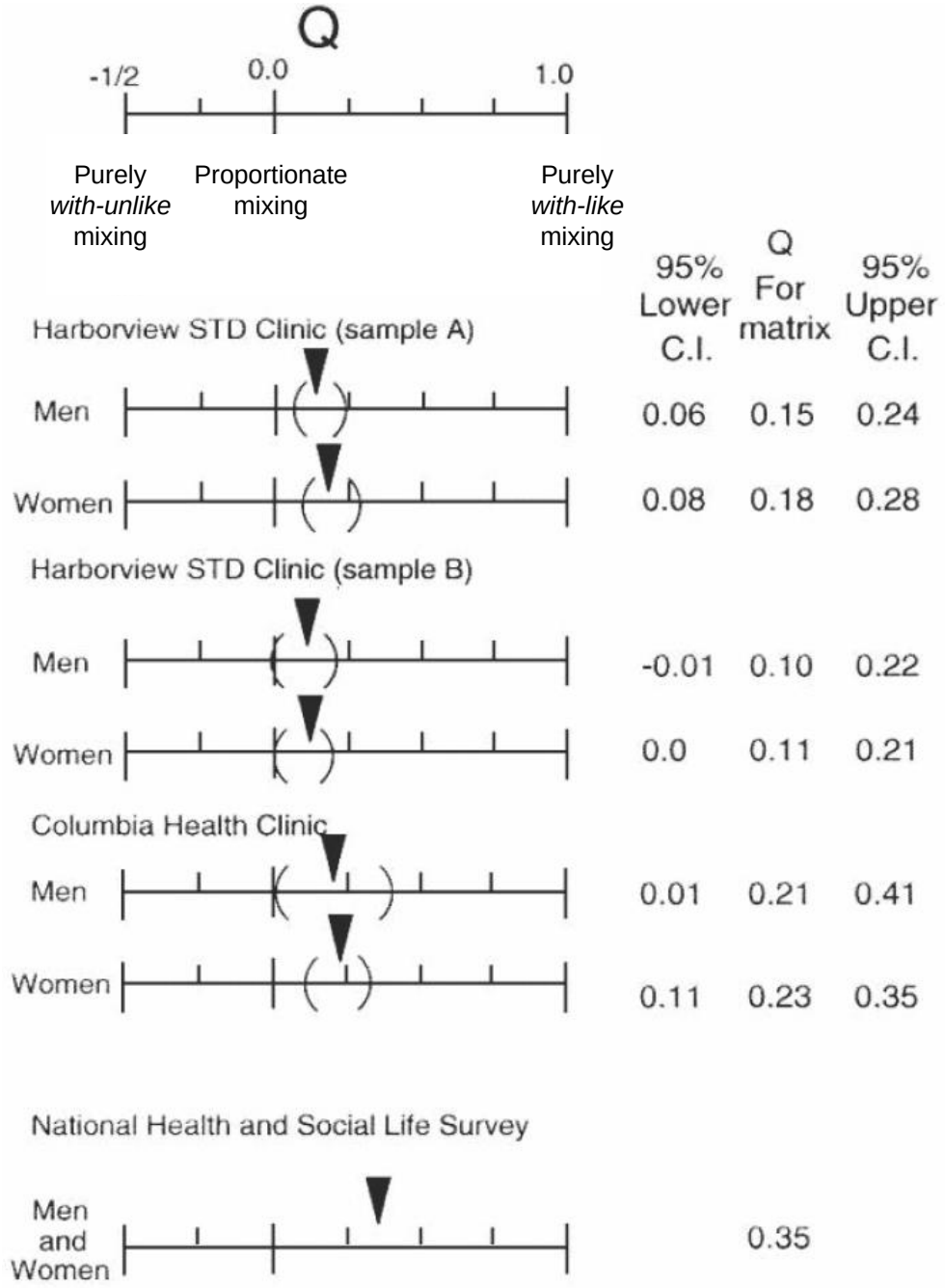


Figure 1-10: The value of the measure *Q* for mixing between sexual activity groups in four studies of sexual behaviour in the US, three of which were based on contact tracing in STI clinics, and one (the National Health and Social Life Survey) was based on a survey of the general population. Figure based on (Garnett, Hughes et al, 1996). Note in this example *Q* was calculated on data categorised into three groups, and therefore purely *with-unlike* mixing was indicated by a Q value of -1/2 rather than -1 as in our two-activity group example.

1.5.4 Modelling mixing by sexual activity

Many methods have been proposed to model mixing between different population groups (Boily and Anderson, 1991; Garnett and Anderson, 1994; Gupta, Anderson et al, 1989; Hallett, Gregson et al, 2007). The method we use is based on the method proposed by *Gupta* et al (Gupta, Anderson et al, 1989). This method allows a range of mixing patterns

between activity groups (*with-unlike,* proportionate, *with-like*) to be modelled using one model parameter for a two-activity group model.

To start we note that the number of partnerships formed between activity group *k* with activity group *j* must equal the number of partnerships formed by activity group *j* with activity group *k*. Here we have two activity groups low ($_L$) and high ($_H$), so:

$$g_{HL}c_L N_L = g_{LH}c_H N_H \qquad \text{Equation 1.20}$$

Using Equation 1.16 and Equation 1.20, if we set the g_{HH}, the probability that someone in the high-activity group forms a partnership with someone in the high-activity group, we can write down equations for g_{HL} , g_{LH} and g_{LL} in terms of g_{HH} :

***g_{LH}* :** Using Equation 1.16, $g_{HH} + g_{LH} = 1$*, so*

$$g_{LH} = 1 - g_{HH} \qquad \text{Equation 1.21}$$

g_{HL} : Rearranging Equation 1.20 for g_{HL} gives:

$$g_{HL} = g_{LH}\frac{c_H N_H}{c_L N_L} \qquad \text{Equation 1.22}$$

Replacing g_{LH} in Equation 1.22 using Equation 1.21, we obtain:

$$g_{HL} = (1 - g_{HH})\frac{c_H N_H}{c_L N_L} \qquad \text{Equation 1.23}$$

***g_{LL}* :** From Equation 1.16 we also know:

$$g_{LL} = 1 - g_{HL} \qquad \text{Equation 1.24}$$

Substituting g_{LH} in Equation 1.23 into Equation 1.24:

$$g_{LL} = 1 - (1 - g_{HH})\frac{c_H N_H}{c_L N_L} \qquad \text{Equation 1.25}$$

These four equations can be summarised as a mixing matrix in terms of g_{HH} :

$$\begin{pmatrix} g_{HH} & g_{HL} \\ g_{LH} & g_{LL} \end{pmatrix} = \begin{pmatrix} g_{HH} & (1 - g_{HH})\dfrac{c_H N_H}{c_L N_L} \\ 1 - g_{HH} & 1 - (1 - g_{HH})\dfrac{c_H N_H}{c_L N_L} \end{pmatrix} \qquad \text{Equation 1.26}$$

We can now calculate the values for the elements of the mixing matrix and for *Q* if the high-activity group selects partners proportionately (ie g_{HH} = 0.31), purely *with-like* (g_{HH} = 1), or

purely *with-unlike* (g_{HH} = 0), and the low-activity group supplies partners to meet these mixing preferences if possible.

To illustrate this, let's assume that the population size is 1000 and therefore there are 20 and 980 individuals in the high and low-activity groups, respectively:

Mixing matrix (Using Equation 1.26)	**Q value** (Using Equation 1.19)

Purely *with-like* mixing by high-activity group

$$\begin{pmatrix} 1 & (1-1)\left(\frac{31.4\times 20}{1.4\times 980}\right) \\ 1-1 & 1-(1-1)\left(\frac{31.4\times 20}{1.4\times 980}\right) \end{pmatrix} = \begin{matrix} \\ H \\ L \end{matrix}\begin{matrix} \begin{matrix} H & L \end{matrix} \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{matrix}$$

Q = (1 + 1 - 1) / (2 -1)
= 1 / 1
= 1

Equation 1.27

Proportionate mixing by high-activity group

$$\begin{pmatrix} 0.31 & (1-0.31)\left(\frac{31.4\times 20}{1.4\times 980}\right) \\ 1-0.31 & 1-(1-0.31)\left(\frac{31.4\times 20}{1.4\times 980}\right) \end{pmatrix} = \begin{matrix} \\ H \\ L \end{matrix}\begin{matrix} \begin{matrix} H & L \end{matrix} \\ \begin{pmatrix} 0.31 & 0.31 \\ 0.69 & 0.69 \end{pmatrix} \end{matrix}$$

Q = (0.31 + 0.69 - 1) / (2 -1)
= 0 / 1
= 0

Equation 1.28

Purely *with-unlike* mixing by high-activity group

$$\begin{pmatrix} 0 & (1-0)\left(\frac{31.4\times 20}{1.4\times 980}\right) \\ 1-0 & 1-(1-0)\left(\frac{31.4\times 20}{1.4\times 980}\right) \end{pmatrix} = \begin{matrix} \\ H \\ L \end{matrix}\begin{matrix} \begin{matrix} H & L \end{matrix} \\ \begin{pmatrix} 0 & 0.46 \\ 1 & 0.54 \end{pmatrix} \end{matrix}$$

Q = (0 + 0.54 - 1) / (2 -1)
= - 0.46 / 1
= - 0.46

Equation 1.29

Thus, we can see that for these partner change rates and group sizes, purely *with-like*, proportionate, or purely *with-unlike* mixing by *high-activity group* members is possible (Equation 1.27, Equation 1.28 and Equation 1.29). We can also see that purely *with-like* or proportionate mixing by the high-activity group members results in purely *with-like* or proportionate mixing in the population overall, and as expected, *Q* equals 1 when we model purely *with-like* mixing and *Q* equals 0 if we model proportionate mixing.

However in our example, purely *with-unlike* mixing by high-activity group members did not result in purely *with-unlike* mixing by low-activity group members (Equation 1.29). This is because the number of partnerships required by the two activity groups differs. To model purely *with-unlike* mixing, partner change rates must be altered so that the numbers of partnerships required by both activity groups balance (see Panel 1-7).

In the following section we vary *Q* between -0.46 and +1 so we don't have to alter the partner change rates in the model. This means we can more clearly see the effect of mixing on $R_{0,}$ and the effect of mixing on the spread of STIs.

1.5.5 Effects of mixing on R_0 , rate of STI spread and equilibrium STI prevalence

Let's start by varying *Q,* while keeping partner change rates and biological parameters (*D* and β_p) constant, to explore the effect of mixing on R_0. 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*).

As Figure 1-11 shows, R_0 increases as the mixing pattern becomes more *with-like*. This is because increasing *with-like* mixing means higher-activity individuals tend to contact other higher activity individuals more frequently. As higher-activity individuals are more likely to be infected, increasing *with-like* mixing leads to an increase in the probability a 'typical-infectee' will be a member of the high-activity group. Higher-activity individuals have higher partner change rates and therefore generate more secondary infections in a completely susceptible population. Therefore the value of R_0 increases. The derivation of the equation for R_0 that can be used to calculate these R_0 values for any pattern of mixing is shown in Panel 1-8. This derivation is more general than that shown for proportionate mixing in Panel 1-5.

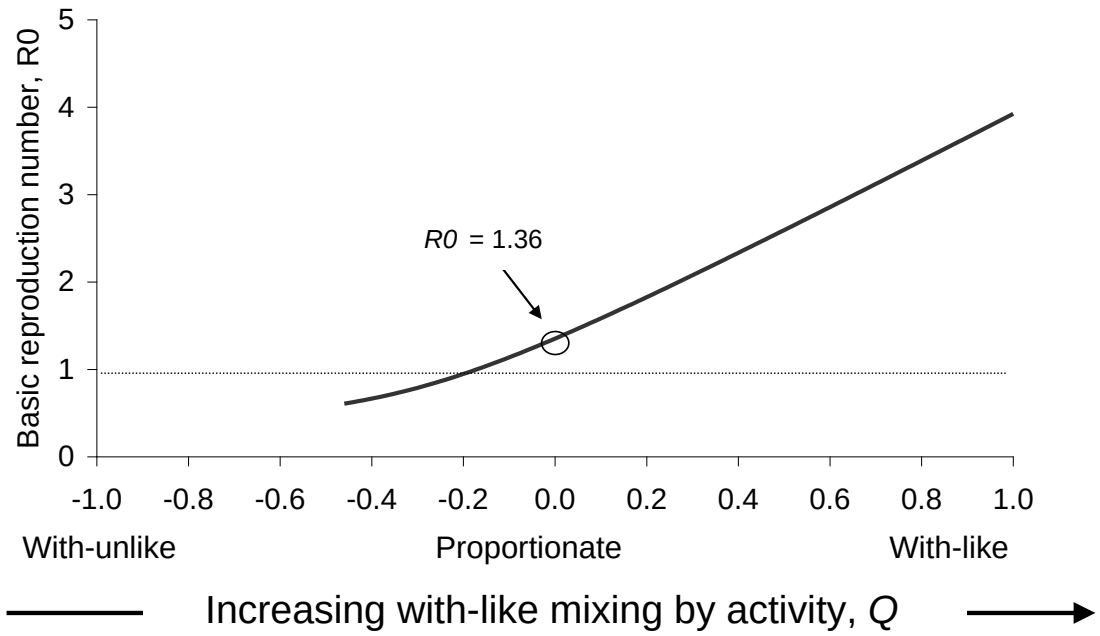


Figure 1-11 Predictions for the basic reproduction number, R_0 , for a curable STI in a single-sex, two-activity group model as the mixing pattern between activity groups varies from 'More with-unlike' (Q= -0.46, the most *with-unlike* mixing possible without altering partner change rates), through proportionate (Q= 0), to purely *with-like* mixing (Q= 1). β_p = 0.75, D = 0.167 years, c_L = 1.4 partners/year, c_H = 31.4 partners/year. 2% of the population belong to the high-activity group. An arrow shows the value of R_0 (1.36) assuming proportionate mixing as illustrated in section 1.4. A horizontal line highlights R_0 = 1.

Note that in the limit when Q=1, the two activity groups cease interacting, and we are modelling two independent populations, each with its own R_0 value. The R_0 plotted in Figure 1-11 for Q=1 is the R_0 for high-activity group. The R_0 value for the lower activity group will be smaller.

Note also that in this example, the value of R_0 falls below one when mixing becomes moderately *with-unlike* (Q < -0.17) and prevalence will fall ultimately to zero.

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

Increasing *with-like* mixing tends to lead to the infection invading more rapidly, but rather surprisingly, for a given R_0, it also tends to result in lower equilibrium prevalence (Figure 1-12 and (Garnett, Swinton et al, 1992).

The STI invades more rapidly as *with-like* mixing increases because infected higher-activity individuals are more likely to transmit infection to other higher-activity individuals, who generate new infections more quickly than lower activity individuals.

The equilibrium STI prevalence is higher in scenarios with less *with-like* mixing because, for a given R_0, more transmission has to occur in the low-activity group. The lower prevalence in the low-activity group means fewer of these contacts are 'wasted' on infected individuals than if these contacts had been with higher activity individuals. Therefore prevalence rises in the low-activity group until a sufficient proportion of contacts are again 'wasted' on infected individuals to reduce the number of secondary infections from 1.36 (at STI introduction) to 1 (at equilibrium). Therefore, at equilibrium prevalence is higher in the, larger, low-activity group, and therefore the overall equilibrium prevalence is higher.

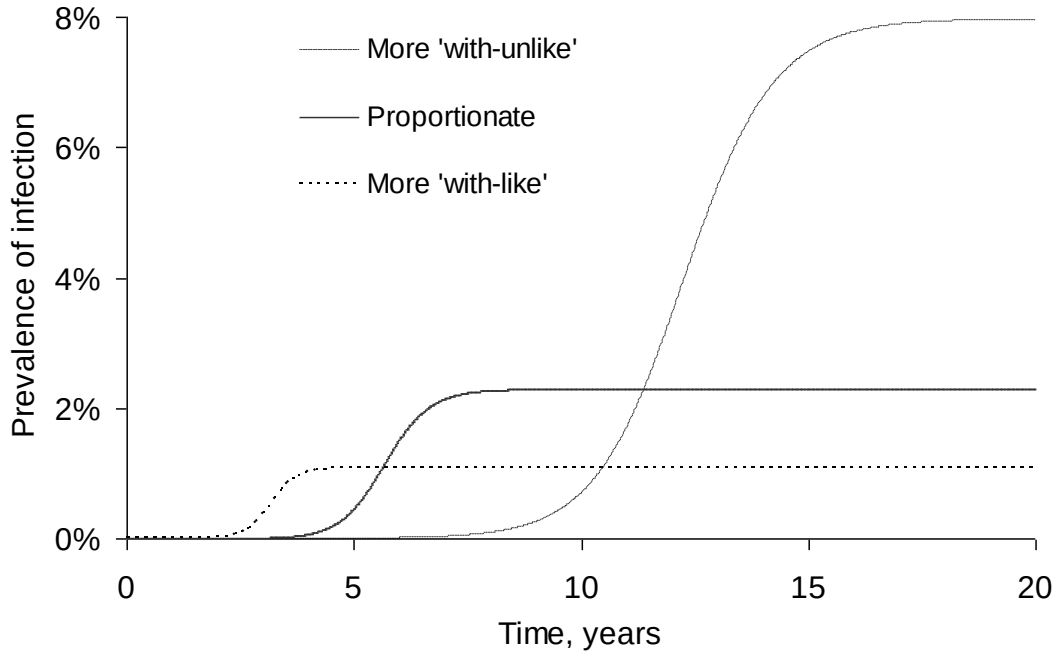


Figure 1-12 Predictions of the overall prevalence of infectious individuals over time in a curable STI model with more *with-unlike* (Q= -0.4), proportionate (Q= 0), or more *with-like* (Q= +0.4) mixing between activity groups. R_0 was set equal to 1.36 in all scenarios by varying the duration of infection (D = 0.340, 0.167 and 0.097 years, respectively). β_p = 0.75, c_L = 1.4 partners/year, c_H = 31.4 partners/year. 2% of the population belong to the high-activity group.

Note however that in models with heterogeneity in risk activity, the equilibrium prevalence tends to be lower and the effects of saturation tend to occur earlier than in models without heterogeneity. The is clearly true in the example because if homogenous mixing was assumed, an R_0 of 1.36 would result in an equilibrium prevalence of 26% (using $i(\infty) = 1 - 1/R_0$). This is much higher than the prevalence in any of the scenarios shown in Figure 1-12 with heterogeneity in risk behaviour.

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

Fig 1-12 illustrated the effect of changing the mixing pattern if R_0 was held constant by increasing the duration of infection of the STI. It is also interesting to 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.

Figure 1-13 shows that increasing *with-like* mixing may not always lead to a reduction in equilibrium prevalence. For lower values of R_0, increasing *with-like* mixing may actually allow the infection to invade a population and lead to a rise in overall prevalence (Figure 1-13, top). Conversely, for higher values of R_0, increasing *with-like* mixing will always lead to a fall in overall prevalence because more contacts are 'wasted' on already-infected individuals in the high-activity group (Figure 1-13, middle). For moderate values of R_0, both

effects may be seen, so that increasing *with-like* mixing may lead to an initial rise, but subsequent fall in overall prevalence (Figure 1-13, bottom).

R_0 in proportionate mixing scenario (Q=0)

Low (R_0= 0.50)

High (R_0= 4.00)

Moderate (R_0= 1.36)

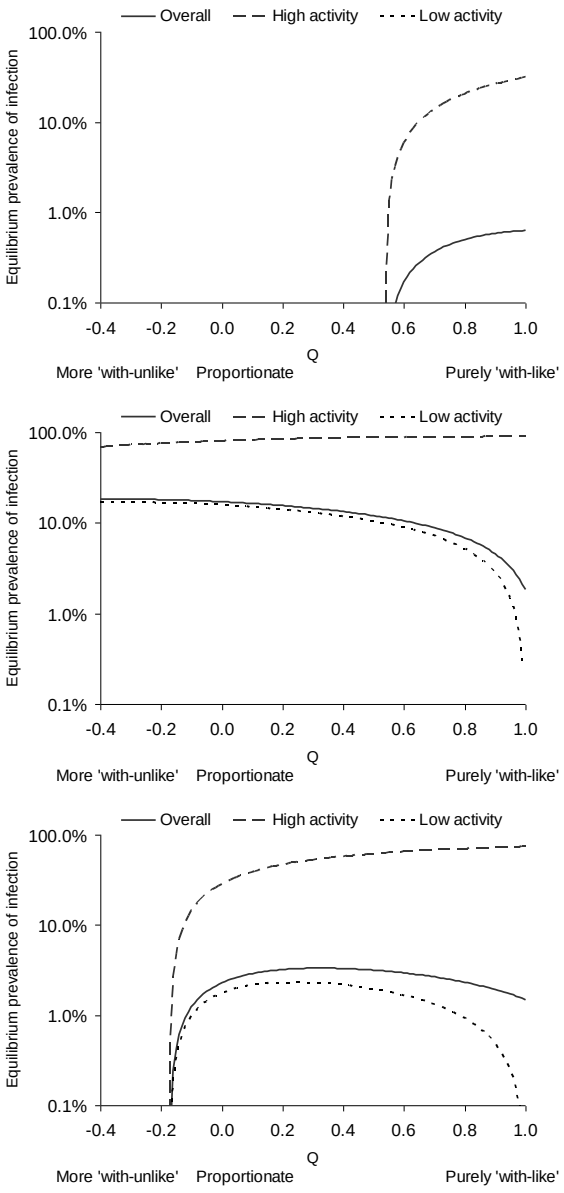


Figure 1-13 Equilibrium prevalence of a curable STI by mixing pattern and R_0. Assumes partner change rate in high and low-activity groups are 31.4 and 1.4 partners per year, respectively, the high-activity group is 2% of the population, the per-partnership transmission probability is 0.75. In the low, moderate and high R_0 scenarios, the durations of infection are 0.062, 0.167, and 0.493 years, respectively. Note a $\log_{10}$ scale is used on y-axis for clarity. Note that only two lines appear in the low R_0 scenario because the prevalence of infection in the low-activity group remains below 0.1% for all values of Q.

1.5.8 Implications of heterogeneity in sexual activity on STI control

Heterogeneity in sexual activity and different mixing patterns between activity groups also has implications for the control of STIs. Figure 1-14 shows how the equilibrium prevalence of gonorrhoea may be affected by a screening program. Screening is implemented very simply in the model by adding a term to the equations determining the rate of change of the infectious and susceptible individuals in the high and low 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. Model equations Equation 1.7 and Equation 1.8 become:

$$\frac{dS_j(t)}{dt} = -\lambda_j(t)S_j(t) + rI_j(t) + y_j I_j(t) \qquad \text{Equation 1.30}$$

$$\frac{dI_j(t)}{dt} = +\lambda_j(t)S_j(t) - rI_j(t) - y_j I_j(t) \qquad \text{Equation 1.31}$$

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.

Figure 1-14 shows how the equilibrium prevalence may change with the annual number of individuals screened and treated for gonorrhoea each year. Three scenarios show the impact of random screening or screening targeted exclusively at high or low-activity group members.

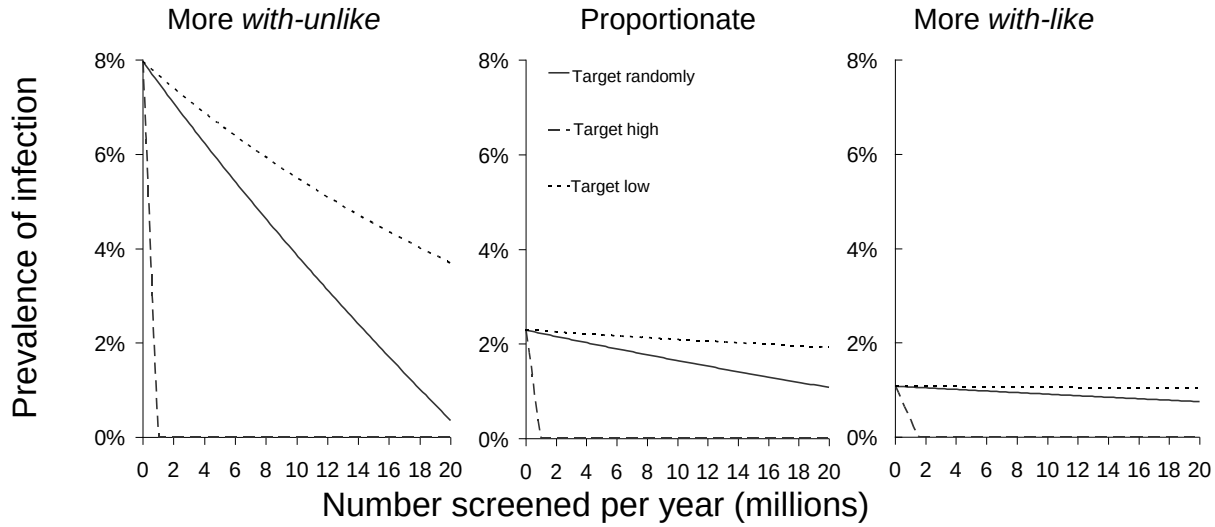


Figure 1-14 The effect of screening the population on equilibrium curable STI prevalence in the overall population, by mixing pattern and target group. The mixing patterns shown are: more *with-unlike* (Q= -0.4), proportionate (Q= 0), more *with-like* (Q= +0.4). Screening is targeted either randomly, or at high or low activity groups. R_0 was set equal to 1.36 in all scenarios by varying the duration of infection (D = 0.340, 0.167 and 0.097 years, respectively. Other parameter values are as shown in the heading to Figure 1-12. The assumed population size is 20 million.

Looking at each of the panels in Figure 1-14 independently shows that prioritising higher-risk members of the population is markedly more effective per-person-screened than prioritising

individuals at random or prioritising members of the low-activity group. For example, if the mixing pattern is proportionate (Figure 1-14, middle), one million screenings per year does not markedly affect prevalence if randomly distributed or targeted at the low-activity group, but eradicates the infection if targeted at the high-activity group.

Looking across all three panels in Figure 1-14 shows that increased *with-like* mixing tends to make STIs more difficult to control. The same number of screenings has a smaller impact on the equilibrium STI prevalence in populations with more *with-like* mixing. In our example, 20 million screenings per year randomly distributed would reduce the relative equilibrium prevalence by around a third in the more *with-like* mixing scenario, by around a half in the proportionate mixing scenario but would almost eradicate the infection in the more *with-unlike* mixing scenario.

1.6 Mixing on gender (heterosexual mixing model)

In the next model in this session, we focus exclusively on high-activity individuals and explore the implications of differences in the natural history of gonorrhoea in males and females. We calculate the number of secondary infections from a single infection in each gender, and the basic reproduction number for gonorrhoea in this population.

Consider the two-gender model shown in Figure 1-15.

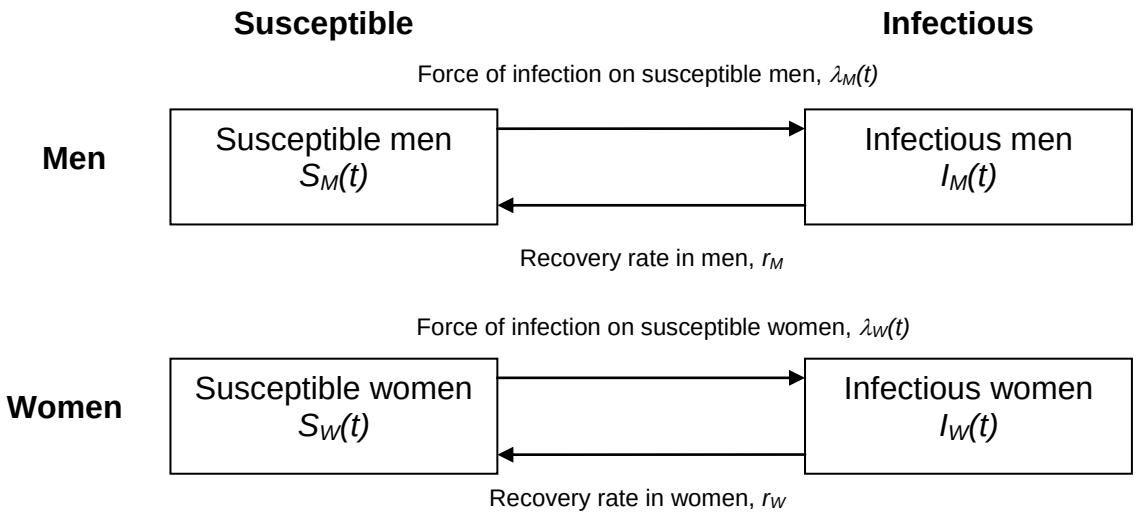


Figure 1-15: A two-gender SIS model of gonorrhoea transmission

In this model we assume exclusively heterosexual activity. Therefore we assume completely *with-unlike* mixing between the two genders. Note that in a model for a heterosexual population in which individuals are grouped by gender and activity, we are likely to assume moderately *with-like* mixing by activity and *with-unlike* mixing by gender. A model that assumes some same-gender sexual activity could be parameterised to model moderately *with-like* mixing by activity and moderately *with-unlike* mixing by gender. We also assume the male to female ratio is 1:1 and our population is *closed,* and therefore the numbers of partnerships formed by men and by women are equal.

The equations for rate of change of the number of female susceptible and infectious individuals and the force of infection on females at time *t* in this model are:

$$\frac{dS_W(t)}{dt} = -\lambda_W(t)S_W(t) + rI_W(t) \qquad \text{Equation 1.32}$$

$$\frac{dI_W(t)}{dt} = +\lambda_W(t)S_W(t) - rI_W(t) \qquad \text{Equation 1.33}$$

$$\lambda_W(t) = c_W\beta_{WM}i_M(t) \qquad \text{Equation 1.34}$$

Where c_W is the average rate at which women change partners, β_{WM} is the per partnership transmission probability from men to women, and $i_M(t)$ is the infection prevalence in males at time *t*. The equations and parameters for males are obtained by replacing 'W's with 'M's and vice versa.

The left hand side of Figure 1-16 illustrates heterosexual mixing in which one infectious woman generates four secondary infections in men, and each infection in males generates, on average, 0.5 secondary infections in women (ie four infections → two infections). So overall, one infection in women leads to two tertiary infections in women and therefore we would expect the STI to invade the population.

Figure 1-16: Schematic of STI transmission between men and women (left) and infection transmission between host and vector (right).

This looks very similar to host-vector modelling (Figure 1-16-right). Ronald Ross first realised this similarity when he was developing his models for Malaria in 1911 (Ross, 1911). Ross recognised that one gender could be thought of as the 'vector' to the other gender. Indeed, *Hethcote* and colleagues based the equations for their two-gender STI model (Hethcote, 1976) on the equations and ideas developed by Ross for his host-vector modelling of malaria.

1.6.1 Calculating R_0 for a heterosexually mixing population (host-vector)

As before, to calculate the basic reproduction number for the population as a whole first we need to calculate the number of secondary infections generated by each subgroup in the model. In this case these are men and women:

The *next generation matrix* of the numbers of secondary infections that would result from a single infection introduced into a susceptible heterosexually mixing two-gender model population is:

		Infectious individual: Man	Infectious individual: Woman	
Susceptible individual	Men	0	R_{MW}	Equation 1.35
	Women	R_{WM}	0	

Where R_{MW} is the number of secondary infections in men due to an infected women and R_{WM} is the number of secondary infections in women due to an infected man. Since, in this purely heterosexual transmission model, there is no direct transmission between males and males or between females and females $R_{MM} = R_{WW} = 0$. Further, R_{WM} and R_{MW} may differ because of variation in sexual behaviour, transmission probability and duration of infection between males and females.

We can use any of the methods described in earlier sessions to calculate the R_0 . Here we calculate R_0 by solving the simultaneous equations of the average number of secondary infections caused by a single infection in a typical infectee. The basic reproduction number of gonorrhoea in this population is the average number of secondary infections caused by a single infection in a *typical infectee* introduced into a susceptible population. The typical infectee is some theoretical average of a high and low-activity individual. If we let *x* be the probability the typical infectee is a member of the high-activity group, and *1-x* be the probability the typical infectee is a member of the low-activity group, R_0 is the maximum value that satisfies the following matrix equation:

$$\begin{pmatrix} 0 & R_{MW} \\ R_{WM} & 0 \end{pmatrix}\begin{pmatrix} x \\ 1-x \end{pmatrix} = R_0\begin{pmatrix} x \\ 1-x \end{pmatrix} \qquad \text{Equation 1.36}$$

The two implicit equations can be solved for R_0, eliminating the unknown term, '*x*'

$$R_{MW}(1-x) = R_0 x \qquad \text{Equation 1.37}$$

$$R_{WM}x = R_0(1-x) \qquad \text{Equation 1.38}$$

Rearranging Equation 1.37 for (1-x) and substituting (1-x) into Equation 1.38 yields

$$R_{WM}x = \frac{R_0 R_0 x}{R_{MW}}$$

Cancelling the remaining *x* terms and rearranging gives the equation for R_0 for a two-gender heterosexually mixing model of as SIS infection like gonorrhoea:

$$R_0 = \sqrt{R_{WM}R_{MW}} \qquad \text{Equation 1.39}$$

This is a particular case of a more general result for the R_0 of an infection in a host-vector model such that the basic reproduction number is equal to the geometric mean of the number of secondary infections from each of the groups (Dietz, 1993). Here we have two groups, males and females, so the geometric mean is the square root of the product of R_{WM} and R_{MW} .

The equations for R_{WM} and R_{MW} are:

$$R_{WM} = c_M \beta_{WM} D_M \quad \text{Equation 1.40}$$

$$R_{MW} = c_W \beta_{MW} D_W \quad \text{Equation 1.41}$$

Where c_W and c_M are the annual partner change rates in women and men, respectively, β_{MW} and β_{WM} are the transmission probabilities per partnership from women to men and men to women, respectively, and D_W and D_M is the duration of infection in women and men, respectively.

As we are modelling the highly active group, we assume the partner change rates among females and males, c_W and c_M , are both 31.4 partners per year. To model (crudely) the lower proportion of infections that are symptomatic in women than men (Holmes, Sparling et al, 2008) and therefore the likely lower treatment seeking rates among women than men, we assume the average duration of infectiousness with some treatment in the population is longer for females than for males (D_W = 3 months and D_M = 1 month). Finally, we assume the male to female transmission probability per partnership is higher than the female to male transmission probability per partnership (β_{WM}=0.9 and β_{MW} =0.6) as is commonly observed for STI transmission (Holmes, Sparling et al, 2008).

Using these parameter values:

$$R_{WM} = c_M \beta_{WM} D_M = 31.4 \times 0.9 \times \frac{1}{12} = 2.4 \quad \text{Equation 1.42}$$

$$R_{MW} = c_W \beta_{MW} D_W = 31.4 \times 0.6 \times \frac{3}{12} = 4.7 \quad \text{Equation 1.43}$$

Using these values and Equation 1.39 we can calculate the basic reproduction number for this STI in this high-activity population.

$$\begin{aligned} R_0 &= \sqrt{R_{WM} R_{MW}} \\ &= \sqrt{2.4 \times 4.7} \\ &= 3.3 \end{aligned} \quad \text{Equation 1.44}$$

Thus, our simple model has highlighted the potential importance of gender differences in the natural history and transmission probability of STIs. Although the probability of gonorrhoea transmission from female-to-males is lower than from males-to-females, our model predicted that females may generate more secondary infectious than males because of the longer average duration of infection in females than in males.

1.7 Summary

The early deterministic compartmental modelling work of gonorrhoea by *Cooke, Hethcote* and *Yorke* highlighted some of the important characteristics of sexually transmitted infectious that are important for their transmission and control. They showed that heterogeneity in human sexual activity was critical in explaining the invasion of short-duration non-immunising bacterial STIs and their relatively low but stable equilibrium prevalence. They coined the term *pre-emptive saturation* for the process that limits transmission for non-

immunising infections in which contacts of infectious individuals are 'wasted' (from the pathogen's perspective) on already-infected individuals. Using only slightly more complicated models than those detailed above, *Hethcote* and *Yorke* showed the importance of asymptomatic individuals in the spread of infection, and that strategies that best identified either higher-activity individuals or asymptomatics, or ideally both, would be most effective at controlling infection. Their modelling helped inform US STI control policy in the 1980s.

We have also explored the likely impact of mixing by sexual activity on the transmission and control of STIs. In general *with-like* (also called 'assortative') mixing facilitates the invasion of an infection, because fewer contacts are 'wasted' on low activity individuals, but it limits the overall spread of infection because it tends to protect lower-activity individuals as it reduces contact between activity groups. More *with-like* mixing also tends to make infections more difficult to eradicate because it creates a 'core' of high-activity individuals who continually infect and re-infect each other.

Panel 1-1 *Per-act* and *per-partnership* transmission probabilities

STIs are transmitted during sexual contacts or acts and therefore it is natural to try to model at the level of the sexual act. However the data on transmission probabilities are commonly available from studies of sexual partnerships. If we have an estimate of the number of acts within a partnership and we assume that the per-act transmission probability does not vary over time (which may not be true), we can easily calculate the per-act transmission probability from the per-partnership probability using a binomial model (Kaplan, 1990):

$$\beta_p = 1 - (1 - \beta_a)^n \qquad \text{Equation 1.45}$$

Where β_p is the transmission probability per sexual partnership, β_a is the transmission probability per sexual act, and n is the number of sex acts in the partnership after infection of the index case. This equation is analogous to the Reed-Frost equation discussed in the stochastic modelling session and can be understood by considering the constituent parts in turn:

$1 - \beta_a$ is the probability of **not getting** infected in 1 sex act
$(1 - \beta_a)^n$ is the probability of **not getting** infected in n sex acts
$1 - (1 - \beta_a)^n$ is the probability of **getting** infected in n sex acts

STIs with high transmission probabilities, such as gonorrhoea and chlamydia infection will quickly transmit to sexual partners, while STIs with low transmission probabilities such as HIV in the asymptomatic stage and in the absence of other cofactors for transmission will take much longer to transmit to sexual partners. The relationship between β_p and β_a is shown for a range of transmission probabilities in Figure 1-17.

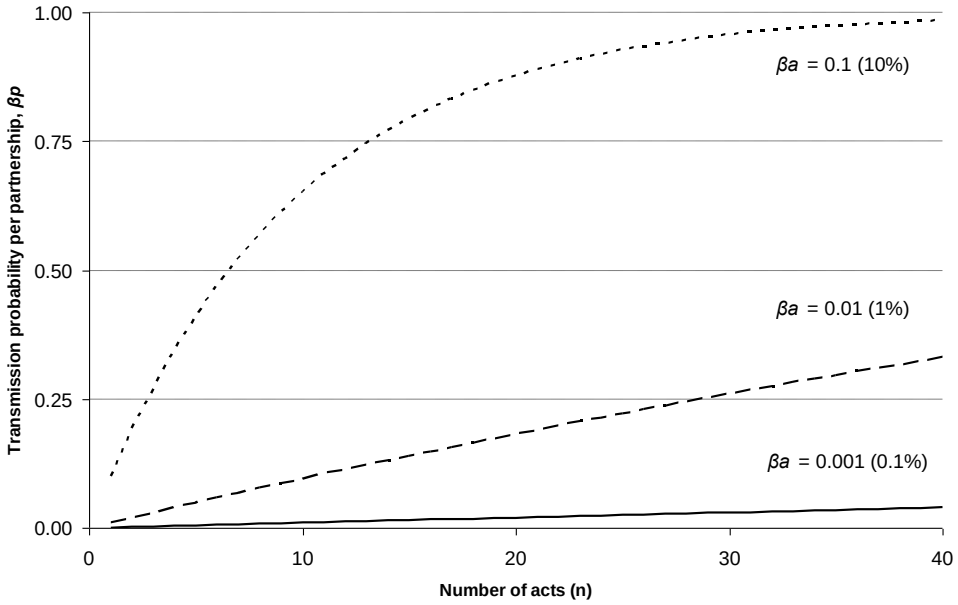


Figure 1-17 The relationship between the transmission probability per partnership (β_p) and per sex-act (β_a) for n sex acts.

This has important implications for the impact of condom use in preventing infection (Garnett and Anderson, 1995; Pinkerton and Abramson, 1996). For highly infectious STIs such as gonorrhoea, chlamydia and HIV during the primary stage of infection, inconsistent condom use has little effect in preventing transmission. For less infectious STIs such as HIV during the asymptomatic stage and in the absence of other cofactors for transmission, even inconsistent condom use can still have a significant impact.

Panel 1-2 Derivation of R_0 for simple gonorrhoea model

We can obtain R_0 for this infection using a similar approach to that used in earlier sessions. Using Equation 1.2 and Equation 1.5 we can write down an equation for the rate of change of the number of infectious individuals

$$\frac{dI(t)}{dt} = \lambda(t)S(t) - rI(t)$$
$$= c\beta_p \frac{I(t)}{N} S(t) - rI(t)$$

If the STI is to spread then the rate of change of the number of infectious individuals must be greater than zero, ie

$$c\beta_p \frac{I(t)}{N} S(t) - rI(t) > 0 \qquad \text{Equation 1.46}$$

Rearranging, we get

$$c\beta_p \frac{I(t)}{N} S(t) > rI(t) \qquad \text{Equation 1.47}$$

Cancelling *I(t)* from both sides of the equation and recognising that when *t*=0, *S(0) ~ N,* we get

$$c\beta_p > r \qquad \text{Equation 1.48}$$

Substituting *1/D* for *r* and rearranging we get

$$c\beta_p D > 1 \qquad \text{Equation 1.49}$$

So, for the number of infectious persons to increase following the introduction of an infectious person into a totally susceptible population, the quantity $c\beta_p D$ must be greater than one. During their infectious period of duration D, a single infectious person will have cD sexual partners. Of these, a proportion β_p will be infected and therefore $c\beta_p D$ equals the average number of secondary infectious persons resulting from a single infectious person in a totally susceptible population. As this is the verbal definition of R_0 and $c\beta_p D$ has the correct threshold behaviour, it is reasonable to define $R_0 = c\beta_p D$.

Panel 1-3 Maintaining overall partner change rate

The equation for the mean rate of partner change is:

$$c_{mean} = c_H \frac{N_H}{N} + c_L \frac{N_L}{N}$$ Equation 1.50

To keep the total number of partnerships formed per year equal in both models but increase the difference in the rate of partner change between the high and low-activity groups we rearrange Equation 1.50 to give:

$$c_H = \frac{c_{mean} N - c_L N_L}{N_H}$$ Equation 1.51

Equation 1.51 calculates the correct partner change rate for the high-activity group, c_H , if we reduce the partner change rate in the low-activity group, c_L so that the overall mean partner change rate, c_{mean} , remains at 2 partners per year,

Note equation Equation 1.51 can be re-written in terms of the n_H and n_L , the proportions of the population that are in the high and low-activity groups such that $n_H + n_L = 1$*:*

$$c_H = \frac{c_{mean} - c_L n_L}{n_H}$$ Equation 1.52

Panel 1-4 The probability that a partner will be a member of the high or low-activity group assuming proportionate mixing

The probability, g_H , that a partner chosen according to proportionate mixing will be a member of the high-activity group is equal to the number of partnerships generated by high-activity group members per year, $c_H \times N_H$, divided by the total number of partnerships generated in the population each year: $c_H \times N_H + c_L \times N_L$

$$g_H = \frac{c_H N_H}{c_H N_H + c_L N_L}$$ Equation 1.53

By substituting the proportion of the population in these two groups (n_H and n_L) for the sizes of the two groups (N_H and N_L) where $n_H = N_H / N$ and $n_L = N_L / N$, and cancelling the N's, we obtain:

$$g_H = \frac{c_H n_H}{c_H n_H + c_L n_L}$$ Equation 1.54

$$g_H = \frac{31.4 \times 0.02}{31.4 \times 0.02 + 1.4 \times 0.98}$$

$$g_H = 0.31$$

Similarly g_L is the probability that partner chosen according to proportionate mixing will be a member of the low-activity group. This can be calculated to be 0.69 using Equation 1.54 and substituting in the values of c_L and n_L in the numerator, or by noting that the probability of selecting either a high or low-activity group member must sum to one, ie, $g_H + g_L = 1$, and therefore $g_L = 1 - g_H = 1 - 0.31 = 0.69$.

Panel 1-5 Calculating R_0 in a population with heterogeneity in sexual activity and proportionate mixing

We can use any of the methods described in earlier sessions to calculate the R_0 for an STI in a population with heterogeneity in sexual activity. In this panel we show how to calculate R_0 for a two group model with heterogeneity in sexual activity and proportionate mixing by (a) solving the simultaneous equations of the average number of secondary infections caused by a single infection in a typical infectee and (b) based on the mean and variance of the partnership distribution. In Appendix 1.1 we show how to calculate R_0 using a matrix determinant approach or by calculating the dominant eigenvalue of the next generation matrix by simulation.

a) Calculate R_0 by solving the simultaneous equations of the average number of secondary infections caused by a single infection in a typical infectee

The basic reproduction number of gonorrhoea in this population is the average number of secondary infections caused by a single infection in a *typical infectee* introduced into a susceptible population. The typical infectee is some theoretical average of a high and low-activity individual. If we let x be the probability the typical infectee is a member of the high-activity group, and $1-x$ be the probability the typical infectee is a member of the low-activity group, R_0 is the maximum value that satisfies the following matrix equation

$$\begin{pmatrix} R_{HH} & R_{HL} \\ R_{LH} & R_{LL} \end{pmatrix} \begin{pmatrix} x \\ 1-x \end{pmatrix} = R_0 \begin{pmatrix} x \\ 1-x \end{pmatrix} \qquad \text{Equation 1.55}$$

Where R_{HH} is the number of secondary infections in high-activity members generated by an infected high-activity member, R_{LH} is the number of secondary infections in low-activity members generated by an infected high-activity member, R_{HL} is the number of secondary infections in high-activity members generated by an infected low-activity member, and R_{LL} is the number of secondary infections in low-activity members generated by an infected low-activity member.

Using Equation 1.11 and Equation 1.12 we know that in a completely susceptible population, a high-activity member generates R_H (3.93) infections and a low-activity member generates R_L (0.18) infections. As we assume proportional mixing, using Equation 1.54 we also know that 31% (g_H) of these infections will be transmitted to high-activity members and 69% (g_L) will be transmitted to low-activity members, so we can write:

$$\begin{pmatrix} R_H g_H & R_L g_H \\ R_H g_L & R_L g_L \end{pmatrix} \begin{pmatrix} x \\ 1-x \end{pmatrix} = R_0 \begin{pmatrix} x \\ 1-x \end{pmatrix} \qquad \text{Equation 1.56}$$

The two implicit equations can be solved for R_0, eliminating the unknown term, x:

$$R_H g_H x + R_L g_H (1-x) = R_0 x \qquad \text{Equation 1.57}$$

$$R_H g_L x + R_L g_L (1-x) = R_0 (1-x)$$ Equation 1.58

Rearranging Equation 1.57 we get

$$x = \frac{-R_L g_H}{g_H (R_H - R_L) - R_0}$$ Equation 1.59

Substituting Equation 1.59 into Equation 1.58 we get the rather untidy equation

$$R_H g_L \frac{-R_L g_H}{g_H (R_H - R_L) - R_0} + R_L g_L \left(1 - \frac{-R_L g_H}{g_H (R_H - R_L) - R_0}\right) = R_0 \left(1 - \frac{-R_L g_H}{g_H (R_H - R_L) - R_0}\right)$$

Equation 1.60

Multiplying all terms by $g_H (R_H - R_L) - R_0$ and expanding we get

$$\begin{aligned} &- R_H g_L R_L g_H + R_H g_L R_L g_H - R_L g_L R_L g_H - R_L g_L R_0 + R_L g_L R_L g_H \\ &\quad = R_H g_H R_0 - R_L g_H R_0 - R_0^{\,2} + R_L g_H R_0 \end{aligned}$$ Equation 1.61

Most of these terms cancel and after a little rearrangement this reduces to

$$- R_H g_H - R_L g_L = - R_0$$ Equation 1.62

And one further rearrangement yields

$$R_0 = g_H R_H + g_L R_L$$ Equation 1.63

From Panel 1-4 we know $g_H = 0.31$ and $g_L = 0.69$, so:

$R_0 = 0.31 \times 3.93 + 0.69 \times 0.18$
$= 1.36$

b) Calculate R_0 based on the mean and variance of the partnership distribution

Alternatively we can use an approximation based on the mean and variance of the partnership distribution to calculate R_0 (Anderson and May, 1991). In this method R_0 is calculated using the *effective partner change rate*, $\hat{c}$, where $\hat{c}$ is the 'epidemiologically relevant' average number of sexual partners and can be much larger than the simple mean if the variance in the number of partners is high.

In a population with heterogeneity in activity, $\hat{c}$ can be approximated by the sum of the arithmetic mean, c, and the variance, σ^2, divided by the arithmetic mean (Anderson and May, 1991).

$$\hat{c} = c + \sigma^2 / c$$ Equation 1.64

As 2% of our population have a partner change rate of 31.4 partners per year and 98% of our population have a partner change rate of 1.4 partners per year, the variance is 17.6 (calculated here using the Excel 'VARP' function). As the mean partner change rate is two partners per year, the *effective partner change* rate is:

$$\begin{aligned} \hat{c} &= 2 + 17.6 / 2 \\ &= 10.8 \end{aligned}$$

Substituting $\hat{c}$ into Equation 1.6 we get

$$\begin{aligned} R_0 &= \hat{c}\,\beta_p D \\ &= 10.8 \times 0.75 \times 0.167 \\ &= 1.36 \end{aligned}$$

Equation 1.65

Panel 1-6 Immunity and 'pre-emptive saturation'

An STI can invade a population if the R_0 of the STI in the population is above one, but after introduction the prevalence of infection cannot increase indefinitely as it is bounded by 100%. Therefore some mechanism must be responsible for reducing the number of secondary infections generated by infectious cases from the R_0 value.

For infections such as measles, mumps, rubella and others that induce immunity in the host, transmission is limited primarily by the increasing proportion of contacts that are between infected and immune individuals. These 'wasted' contacts do not cause new infections in the population and therefore do not increase the prevalence. Figure 1-18-left shows an example in which, for a hypothetical infection, 3 out of 4 potentially effective contacts are 'wasted' on immune individuals.

Most short-duration STIs do not induce immunity in the host. In fact, short-duration STIs could not persist if they induced sterilising immunity in the host because they rely on repeatedly re-infecting the same high-risk individuals. Inducing immunity in the host would greatly limit their reproductive success.

So, what limits the rise in prevalence of these STIs? This process was named 'pre-emptive saturation' by Yorke et al (Yorke, Hethcote et al, 1978). It acts in the same way as immunity, and arises because the contacts of infectious individuals are 'wasted' not on immune individuals, but on individuals who are already infected (Figure 1-18-right).

Figure 1-18 I – infectious individual; R – recovered and immune individual; S – susceptible individual.

This suggests that if two infections are identical in every way (duration of infectiousness, transmission probability per contact, etc) except that infection with one pathogen results in recovery and re-susceptibility in the host (Figure 1-18-right), and infection with the other pathogen results in recovery and induces immunity in the host (Figure 1-18-left), we would expect the prevalence of infected individuals to be higher for the non-immunising infection. This is because for non-immunising infections, only infected individuals can limit the number of secondary infections generated by infectious cases.

Panel 1-7 Modelling purely *with-unlike* mixing by both activity groups

As we saw, in our example purely *with-unlike* mixing by high-activity group members did not result in purely *with-unlike* mixing by low-activity group members (Equation 1.29). This is because the number of partnerships required by the two activity groups differs. The high-activity group requires $c_H N_H = 31.4 \times 20 = 628$ partnerships per year while the low-activity group requires $c_L N_L = 1.4 \times 980 = 1372$ partnerships per year. Therefore even when all of the available partnerships of the high-activity group are formed with low-activity group members, the low-activity group still requires 1372 - 628 = 744 partnerships. If partner change rates are not altered, these 744 remaining partnerships can only be supplied by other low-activity group members. As such, *Q* equals -0.46, not -1 as we would expect if purely *with-unlike* mixing was being modelled (Equation 1.19).

Overall, this means that 46% of the partnerships of low-activity group members are formed with the high-activity group, and 54% of the partnerships of low-activity group members are formed with the low-activity group (Equation 1.29). While this overall mixing pattern is clearly more *with-like* than was achieved by modelling proportionate mixing by the high-activity group (in which only 31% of the partnerships of low-activity group members were formed with the high-activity group, Equation 1.28), we were not able to model purely *with-unlike* mixing between the low and high-activity groups as represented by the mixing matrix shown in Equation 1.18.

To model purely *with-unlike* mixing by both activity groups, the partner change rates must be altered so that the numbers of partnerships required by both activity groups balance. In our example, if we maintain a mean partner change rate of 2 partners per year, the partner change rate in the high and low-activity groups must change to 50 and 1.02 per year. If this is done both groups require 1000 partnerships per year, ie, $c_H N_H = 50 \times 20 = 1000$ and $c_L N_L = 1.02 \times 980 = 1000$, and each group can supply all the partnerships the other group requires.

Mixing matrix (Using Equation 1.26)	**Q value** (Using Equation 1.19)

Purely *with-unlike* mixing by high-activity group and altered partner change rates

$$\begin{pmatrix} 0 & (1-0)\left(\frac{50\times 20}{1.02\times 980}\right) \\ 1-0 & 1-(1-0)\left(\frac{50\times 20}{1.02\times 980}\right) \end{pmatrix} = \begin{matrix} \\ H \\ L \end{matrix}\begin{matrix} \begin{matrix} H & L \end{matrix} \\ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \end{matrix}$$

Q = (0 + 0 - 1) / (2 -1)
= -1 / 1
= -1

Equation 1.66

So if we alter partner change rates we can model purely *with-unlike* mixing, and the value of *Q* would equal -1 as expected.

For other methods of modelling mixing between different groups in the population see (Boily and Anderson, 1991; Garnett and Anderson, 1994; Gupta, Anderson et al, 1989; Hallett, Gregson et al, 2007).

Panel 1-8 Calculating R_0 in a population with heterogeneity in sexual activity and any mixing pattern

We can use any of the methods described in earlier sessions to calculate the R_0 for an STI in a population with heterogeneity in sexual activity and mixing. Here we use only the simultaneous equations approach.

Here we show the derivation of R_0 for a two-activity group model with any mixing pattern. It is more complicated, but more useful, than the simpler derivations if mixing is assumed to be either proportionate (Panel 1-5) or purely *with-unlike* (section 1.6.1).

The basic reproduction number of the STI in this population is the average number of secondary infections caused by a single infection in a *typical infectee* introduced into a susceptible population. The typical infectee is some theoretical average of a high and low-activity individual. If we let *x* be the probability the typical infectee is a member of the high-activity group, and *1-x* be the probability the typical infectee is a member of the low-activity group, R_0 is the maximum value that satisfies the following matrix equation :

$$\begin{pmatrix} R_{HH} & R_{HL} \\ R_{LH} & R_{LL} \end{pmatrix} \begin{pmatrix} x \\ 1-x \end{pmatrix} = R_0 \begin{pmatrix} x \\ 1-x \end{pmatrix} \qquad \text{Equation 1.67}$$

Where R_{HH} is the number of secondary infections in high-activity group members generated by an infected high-activity group member, R_{LH} is the number of secondary infections in low-activity group members generated by an infected high-activity group member, R_{HL} is the number of secondary infections in high-activity group members generated by an infected low-activity group member, and R_{LL} is the number of secondary infections in low-activity group members generated by an infected low-activity group member.

The two implicit equations can be solved for R_0, eliminating the unknown term, *x:*

$$R_{HH}x + R_{HL}(1-x) = R_0 x \qquad \text{Equation 1.68}$$

$$R_{LH}x + R_{LL}(1-x) = R_0(1-x) \qquad \text{Equation 1.69}$$

Rearranging Equation 1.68 we get:

$$x = \frac{-R_{HL}}{R_{HH} - R_{HL} - R_0} \qquad \text{Equation 1.70}$$

Rearranging Equation 1.69 we get:

$$x = \frac{R_0 - R_{LL}}{R_{LH} - R_{LL} + R_0} \qquad \text{Equation 1.71}$$

Eliminating x by setting Equation 1.70 equal to Equation 1.71 and rearranging, we obtain:

$$-R_{HL}\left(R_{LH}-R_{LL}+R_0\right)=\left(R_0-R_{LL}\right)\left(R_{HH}-R_{HL}-R_0\right) \qquad \text{Equation 1.72}$$

By rearranging this equation we obtain a quadratic equation in R_0 (ie the R_0 term is squared):

$$R_0^{\,2}+R_0\left(-R_{HH}-R_{LL}\right)+\left(R_{LL}R_{HH}-R_{HL}R_{LH}\right)=0 \qquad \text{Equation 1.73}$$

This equation can be solved for R_0 using the standard solution for quadratic equation of the form:

$$aR_0^{\,2}+bR_0+c=0 \qquad \text{Equation 1.74}$$

Which is:

$$R_0=\frac{-b\pm\sqrt{b^2-4\times a\times c}}{2a} \qquad \text{Equation 1.75}$$

By comparing Equation 1.73 with Equation 1.74 we can see that, in this example, a = 1, b = $-R_{HH}-R_{LL}$, and c = $R_{LL}R_{HH}-R_{HL}R_{LH}$. (Note *b* and *c* represent different quantities elsewhere in this session).

So we can write down the formula for R_0 for gonorrhoea in this model population:

$$R_0=\frac{R_{HH}+R_{LL}+\sqrt{\left(-R_{HH}-R_{LL}\right)^2-4\times\left(R_{LL}R_{HH}-R_{HL}R_{LH}\right)}}{2} \qquad \text{Equation 1.76}$$

Note, the '±' symbol has been replaced with a '+' because R_0 is the maximum value that satisfies Equation 1.75 and therefore the smaller solution can be ignored.

We can see from this formula that the R_0 will depend on the number of secondary infections that are generated by high and low-activity group members in each of the two activity groups. These numbers will depend on the mixing pattern between the two groups,

Therefore the value of R_0 will depend on the overall number of secondary infections generated by each group, and also the mixing pattern of the population.

Appendix 1.1 Other methods to calculate R_0 in a population with heterogeneity in sexual activity and proportionate mixing

a) Calculate R_0 using the matrix determinant approach

For our model population with two sexual activity groups, R_0 is the maximum value that satisfies the following equation:

$$(R_{HH} - R_0)(R_{LL} - R_0) - R_{LH}R_{HL} = 0 \qquad \text{Equation 1.77}$$

where R_{HH} is the number of secondary infections generated by an infected high-activity member in high-activity members, R_{HL} is the number of secondary infections generated by an infected low-activity member in high-activity members, R_{LH} is the number of secondary infections generated by an infected high-activity member in low-activity members, and R_{LL} is the number of secondary infections generated by an infected low-activity member in low-activity members.

R_{HH}, R_{HL}, R_{LH} and R_{LL} together make up our next generation matrix $\begin{pmatrix} R_{HH} & R_{HL} \\ R_{LH} & R_{LL} \end{pmatrix}$

Using Equation 1.11 and Equation 1.12 we know that in a completely susceptible population, a high-activity member generates R_H (3.93) infections and a low-activity member generates R_L (0.18) infections. As we assume proportionate mixing, using Equation 1.54 we also know that 31% (g_H) of these infections will be transmitted to high-activity members and 69% (g_L) will be transmitted to low-activity members, so we can write:

$R_{HH} = R_H g_H$
$R_{HL} = R_L g_H$
$R_{LH} = R_H g_L$
$R_{LL} = R_L g_L$

Substituting these expressions into Equation 1.77 we obtain the following expression that R_0 has to satisfy:

$$(R_H g_H - R_0)(R_L g_L - R_0) - R_H g_L R_L g_H = 0 \qquad \text{Equation 1.78}$$

Multiplying out this expression and cancelling, we see that R_0 has to satisfy the following expression:

$$R_0^{\,2} - R_0 R_H g_H - R_0 R_L g_L = 0 \qquad \text{Equation 1.79}$$

that simplifies to:

$$R_0(R_0 - (R_H g_H + R_L g_L)) = 0$$ Equation 1.80

To satisfy this equation R_0 must equal 0, or

$$\begin{aligned} R_0 &= g_H R_H + g_L R_L \\ &= 0.31 \times 3.93 + 0.69 \times 0.18 \\ &= 1.36 \end{aligned}$$

b) Calculate R_0 using the dominant eigenvalue of the next generation matrix

Using the simulation approach outlined in the previous sessions to calculate R_0 we need to calculate the next generation matrix, or the number of secondary cases in the high and low-activity groups that result from an infected high or low-activity individual. Again we use the results of Equation 1.54, Equation 1.11 and Equation 1.12 to write down the next generation matrix:

Activity group of infectious individual

Activity group of susceptible individual		High	Low
	High	R_{HH}	R_{HL}
	Low	R_{LH}	R_{LL}

Equation 1.81

$$= \begin{pmatrix} R_H g_H & R_L g_H \\ R_H g_L & R_L g_L \end{pmatrix}$$

$$= \begin{pmatrix} 3.93 \times 0.31 & 0.18 \times 0.31 \\ 3.93 \times 0.69 & 0.18 \times 0.69 \end{pmatrix}$$

$$= \begin{pmatrix} 1.23 & 0.06 \\ 2.70 & 0.12 \end{pmatrix}$$

By simulation R_0 is calculated to be 1.36.

Reference List

Anderson, R. M. and May, R. M.,Spatial, temporal, and genetic heterogeneity in host populations and the design of immunization programmes,IMA J Math Appl Med Biol,1984

Anderson, RM and May, RM,Ch 11: Social heterogeneity and sexually transmitted diseases,Infectious Diseases of Humans: Dynamics and Control,1991

Boily, M-C and Masse, B.,Mathematical models of disease transmission: A precious tool of the study of sexually transmitted diseases,Canadian Journal of Public Health,1997

Boily, M.-C. and Anderson, R. M.,Sexual contact patterns betwen men and women and the spread of HIV-1 in urban centres in Africa,IMA J.Math.Appl.Med.Biol.,1991

Coffin, J., Haase, A., Levy, J. A., Montagnier, L., Oroszlan, S., Teich, N., Temin, H., Toyoshima, K., Varmus, H., Vogt, P., and et al.,Human immunodeficiency viruses,Science,1986

Cooke, K. L and Yorke, J. A.,Some equations modelling growth processes and gonorrhea epidemics,Math.Biosc.,1973

Dietz, K.,The estimation of the basic reproduction number for infectious diseases,Stat Methods Med Res,1993

Fenton, K. A., Johnson, A. M., McManus, S., and Erens, B.,Measuring sexual behaviour: methodological challenges in survey research,Sex Transm Infect,2001

Fleming, D. T. and Wasserheit, J. N.,From epidemiological synergy to public health policy and practice: the contribution of other sexually transmitted diseases to sexual transmission of HIV infection,Sex Transm Inf,1999

Garnett, G. P. and Anderson, R. M.,Contact tracing and the estimation of sexual mixing patterns: the epidemiology of gonococcal infections,Sex Transm Dis,1993

Garnett, G. P. and Anderson, R. M.,Balancing sexual partnerships in an age and activity stratified model of HIV transmission in heterosexual populations,IMA J Math Appl Med Biol,1994

Garnett, G. P. and Anderson, R. M.,Strategies for limiting the spread of HIV in developing countries: conclusions based on studies of the transmission dynamics of the virus,J Acquir Immune Defic Syndr Hum Retrovirol,1995

Garnett, G. P., Hughes, J. P., Anderson, R. M., Stoner, B. P., Aral, S. O., Whittington, W. L., Handsfield, H. H., and Holmes, K. K.,Sexual mixing patterns of patients attending sexually transmitted diseases clinics,Sex Transm Dis,1996

Garnett, G. P., Swinton, J., Brunham, R. C., and Anderson, R. M.,Gonococcal infection, infertility, and population growth: II. The influence of heterogeneity in sexual behaviour,IMA J Math Appl Med Biol,1992

Gupta, S., Anderson, R. M., and May, R. M.,Networks of sexual contacts: implications for the pattern of spread of HIV,AIDS,1989

Hallett, T. B., Gregson, S., Lewis, J. J., Lopman, B. A., and Garnett, G. P.,Behaviour change in generalised HIV epidemics: The impact of reducing cross-generational sex and delaying age at sexual debut,Sex Transm Infect,2007

Hethcote, H. and Yorke, J.,Lecture notes in biomathematics: Gonorrhea transmission and control (vol 56),1984

Hethcote, H. W.,Qualitative analysis for communicable disease models,Math.Biosci.,1976

Holmes, K. K., Sparling, P. F., Stamm, W., Piot, P., Wasserheit, J. N., Corey, L, and Cohen, M.,Sexually Transmitted Diseases,2008

Kaplan, E. H.,Modeling HIV infectivity: must sex acts be counted?,J Acquir Immune Defic Syndr,1990

Keeling, M. J. and Rohani, P.,Chapter 3: Host Heterogeneities,Modeling infectious diseases in humans and animals,2008

Laga, M., Nzila, N., and Goeman, J.,The interrelationship of sexually transmitted diseases and HIV infection: implications for the control of both epidemics in Africa. [Review],AIDS,1991

McPherson, Miller, Smith-Lovin, Lynn, and Cook, James M.,Birds of a Feather: Homophily in Social Networks,Annual Review of Sociology,2001

Pinkerton, S. D. and Abramson, P. R.,Occasional condom use and HIV risk reduction,Journal Of Acquired Immune Deficiency Syndromes And Human Retrovirology,1996

Ross, R.,The prevention of malaria. Second edition with the addendum on the theory of happening.,1911

Rottingen, J. A., Cameron, D. W., and Garnett, G. P.,A systematic review of the epidemiologic interactions between classic sexually transmitted diseases and HIV: how much really is known?,Sex Transm Dis,2001

Schneeberger, A., Mercer, C. H., Gregson, S. A., Ferguson, N. M., Nyamukapa, C. A., Anderson, R. M., Johnson, A. M., and Garnett, G. P.,Scale-free networks and sexually transmitted diseases: a description of observed patterns of sexual contacts in Britain and Zimbabwe,Sex Transm Dis,2004

Turner, K. M., Garnett, G. P., Ghani, A. C., Sterne, J. A., and Low, N.,Investigating ethnic inequalities in the incidence of sexually transmitted infections: mathematical modelling study,Sex Transm Infect,2004

Wasserheit, J. N.,Epidemiological synergy. Interrelationships between human immunodeficiency virus infection and other sexually transmitted diseases. [Review],Sex Transm Dis,1992

Yorke, J., Hethcote, H., and Nod, A.,Dynamics and control of the transmission of gonorrhea,Sex.Transm.Dis.,1978

Introduction to Infectious Disease Modelling and its Applications – 2018

Session 29: The transmission dynamics and control of malaria

Lecture

Objectives

This lecture gives an introduction to the epidemiology and control of malaria using simple mathematical models that represent the transmission of the parasite between human hosts and mosquito vectors.

By the end of the lecture you should:

- Be familiar with the Ross-MacDonald model for malaria transmission and how it captures key stages of the malaria parasite lifecycle
- Appreciate how key features of malaria epidemiology affect transmission dynamics and how these can be incorporated into models
- Understand the expression for R_0 for malaria and how it relates to expressions for R_0 for directly transmitted infections
- Be familiar with the main control measures against malaria

Introduction

In this lecture we look into the epidemiology of malaria from the point of view of the human host and community, and consider models for malaria transmission and control. Malaria is an ancient disease which has existed alongside humans since their earliest history. As a cause of death it has been so important as to drive selection of multiple genes in humans which provide some protection against malaria mortality. The parasite is the protozoa Plasmodium, of which there are five species that can infect humans. *P. falciparum* is the focus of this lecture because it is responsible for the great majority of malaria morbidity and mortality.

In order to develop models of the impact of control interventions on malaria at community level we need first to be able to capture key features of endemic human-malaria interaction, such as superinfection and acquired immunity. Subsequently, we look into how to include interventions such as drug treatment or vector control into malaria transmission models.

Malaria has a long history of mathematical modelling initiated by Ross (1911) and further developed by Macdonald (1957). The insights gained from this work impacted on the first malaria eradication campaign in the mid 1900's. Although full eradication was not achieved, the parasite was successfully eliminated in parts of Asia and more temperate regions such as Europe and North America.

The modelling framework presented here – the Ross-Macdonald model – represents dynamics of *prevalence* of infection in humans and vectors, which is suitable for micro-parasitic vector-borne infections caused by protozoa or viruses. Indeed, extensions of the

Ross-Macdonald model have been developed for the epidemiology of trypanosomiasis, leishmaniasis and dengue. On the other hand, macro-parasitic infections, such as filarial infections, require the use of models describing the dynamics of the *number* of parasites per hosts rather than the number of infected hosts.

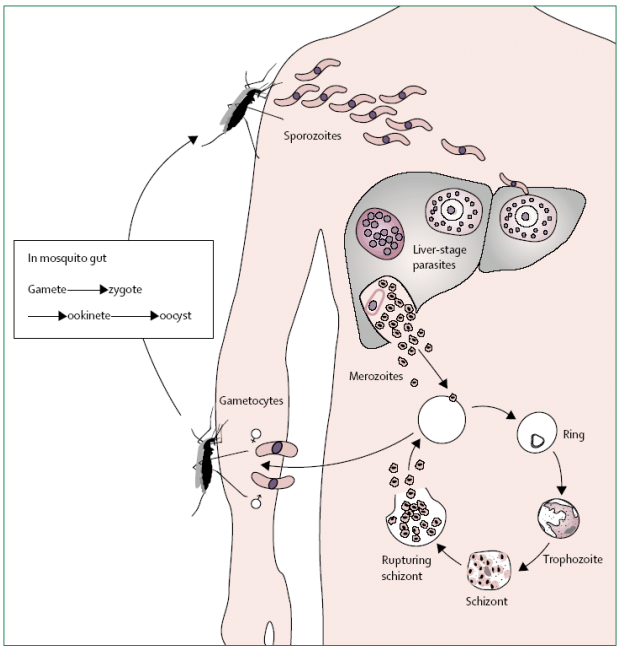


Figure 1: **Life cycle of the malaria parasite**
Adapted from Good MF. Vaccine-induced immunity to malaria parasites and the need for novel strategies. *Trends Parasitol* 2005; 21: 29–34.

Lifecycle of the *Plasmodium* parasite.
Reproduced from Greenwood et al (2005). *The Lancet* **365**, 1487.

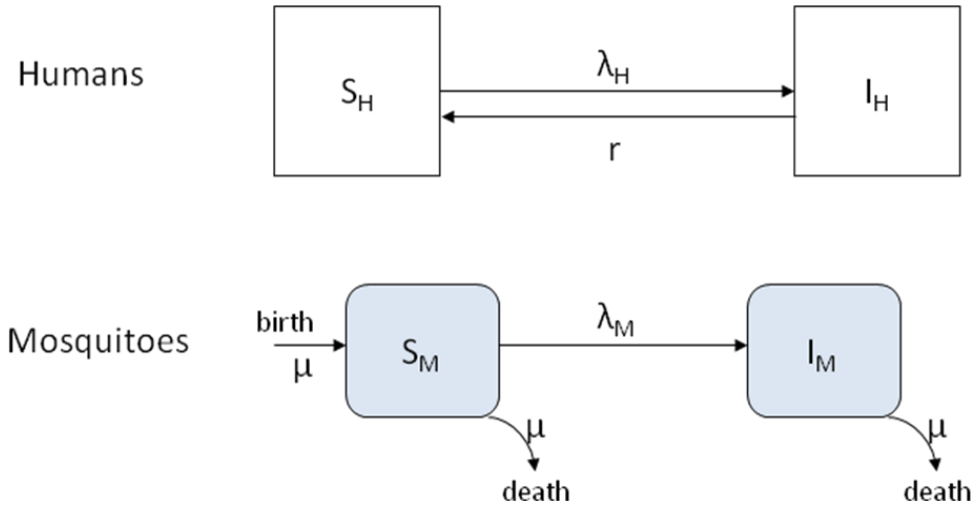
Parasite life cycle

The protozoan parasite *Plasmodium* has a complex life-cycle with obligatory asexual stages in the human and mosquito hosts and sexual reproduction within the female mosquito. Some of the main developmental stages are:

- **The sporozoite**, are transmitted to humans by infectious mosquitoes
- Asexual stages (e.g. **merozoites**), develop from inoculated sporozoites and are the cause of pathology in humans
- **The gametocyte** (sexual stages), are transmitted to mosquitoes by infectious humans; only a few of the merozoites transform into gametocytes.

The Ross-Macdonald model

In its simplest form, the Ross-MacDonald model assumes the human and mosquito populations are subdivided into susceptible and infected (and infectious) groups each. The structure is as shown overleaf.



The change in the prevalence of infection in humans (I_H) over time is

$$\frac{dI_H}{dt} = \lambda_H S_H - rI_H$$

where:
I_H = prevalence of infection in humans
r = rate of recovery of humans
λ_H = force of infection in humans.

The force of infection in humans depends on the prevalence of infection in mosquitoes, which is a key distinctive feature of vector-borne infections relative to directly-transmitted infections; it is given by

$$\lambda_H = mabI_M$$

where:
m = density of female mosquitoes per person
a = biting rate on humans per female mosquito
b = probability of infection of a human upon being bitten by an infected mosquito
I_M = prevalence of infection in female mosquitoes

A measure of transmission intensity which is more often assessed in field studies is the entomological inoculation rate (EIR), which is the number of bites received from infectious mosquitoes per person per unit time (this ignores the probability of successful transmission of the parasite upon biting):

$$EIR = maI_M$$

In this model, the change in the prevalence of infection in mosquitoes over time is

$$\frac{dI_M}{dt} = \lambda_M S_M - \mu I_M$$

where:
I_M = prevalence of infection in mosquitoes
μ_M = rate of mortality of mosquitoes
λ_M = force of infection in mosquitoes.

The force of infection in mosquitoes depends on the prevalence of infection in humans, and is given by

$$\lambda_M = acI_H$$

where:
a = biting rate on humans per female mosquito
c = probability of infection of a mosquito upon biting an infected human
I_H = prevalence of infection in humans

Latent period in the mosquito

In fact this simple model typically overestimates prevalence of infectious mosquitoes relative to field observations (Anderson & May 1991). An improved model is obtained by incorporating a latent period in the mosquito part of the model (where the mosquito is infected but not infectious). This means that infected mosquitoes can die before becoming infectious, and therefore the predicted prevalence of infectious mosquitoes is reduced and more realistic. The probability of an infected mosquito surviving the latent period is given by

$$p^n$$

where:
p = probability of a mosquito surviving 1 day
n = duration of the mosquito latent period (days)

Model structure

The dynamics in the human population are analogous to that of a **SIS model**, as humans can recover from infection, but with an infection rate that depends on the infection status of the vector population. The dynamics in the vector population are analogous to that of an **SI model** (in the simplest case) or of an **SEI model** (for a model with a mosquito latent period), as vectors don't recover from infection, although it is important to account for vector mortality.

Control interventions and R_0

A number of control interventions are available to counteract the malaria parasite: these target either the vector (such as insecticide-treated bednets), or the parasite in humans (e.g. symptomatic case treatment). Recently there has been progress in increasing coverage of control interventions in endemic areas, and a second eradication attempt is being planned, largely as a result of the funding and policies of the Bill & Melinda Gates Foundation. Malaria elimination may be feasible in areas with low levels of transmission such as Asia, but will require much greater effort in the highly endemic areas of sub-Saharan Africa.

The malaria basic reproduction number, R_0, is given by

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

This expression contains the same key components as for other diseases: effective contact rates and duration of infectious periods. It also incorporates the probability of an infected mosquito surviving the latent period to become infectious. The difference with a vector borne disease compared to a directly transmitted disease is that we need to multiply together the human R_0 and the vector R_0 to obtain the overall R_0 expression for the infection.

The formula gives insights into which control interventions may have greatest impact in the value of R_0 and therefore are more likely to lead to substantial reductions in prevalence and possibly local elimination of infection. MacDonald (1957) used this expression for R_0 to predict that targeting adult mosquitoes by insecticide spraying would be more effective than controlling mosquito larvae, with self-protection from biting as the second potentially most effective control measure.

Malaria epidemiology in models

A number of features of malaria epidemiology can have an important effect on transmission dynamics. In the lecture we will consider the following key factors and how these could affect the predicted impact of control interventions.

1. Superinfection – the existence of an infection in a human host does not prevent further infection, or 'superinfection' when bitten by an infectious mosquito.

2. Acquired immunity is a distinctive feature of malaria epidemiology, as humans gradually gain clinical and parasite immunity with repeated exposure to bites from infectious mosquitoes. However protection against the parasite is only ever partial, and immune individuals continue to be infected and transmit disease, contrary to many of the viral infections covered in previous lectures. There are several facets of immunity which develop at different speeds, although these are not fully understood. The chance of developing a severe life-threatening illness reduces swiftly with exposure, while more exposures are required before individuals can be infected without suffering any symptoms at all. Immunity creates a pool of infected, asymptomatic parasite carriers who transmit onwards to mosquitoes, but do not seek treatment, and so has important implications for the epidemiology and control of malaria.

3. Heterogeneity of exposure – this is a feature of almost all infectious diseases (Woolhouse et al.), but is particularly important with regards to malaria. Individual exposure to infection can vary greatly even within a small area due to factors such as proximity to vector breeding sites and housing quality.

References

Malaria background

Greenwood BM (2008). Control to elimination: implications for malaria research. *Trends Parasitol.*24(10): 449-54.

Guerra CA et al (2008). The limits and intensity of Plasmodium falciparum transmission: implications for malaria control and elimination worldwide. *PLoS Med*. 5(2):e38.

Warrell D and Gilles HM (2002). Essential Malariology, Oxford: Arnold, 4th edition

Modelling references

Macdonald G (1957). *The epidemiology and control of malaria*. London: Oxford University Press.

Dietz K (1988) *Mathematical models for transmission and control of malaria* (book chapter). Malaria: principles and practice of malariology. Volume 2. Eds Wernsdorfer, W. H., McGregor I. Edinburgh, Churchill Livingstone.

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

Smith DL, McKenzie FL, Snow RW, Hay SI (2007). Revisiting the basic reproductive number for malaria and its implications for control. *PloS Biology,* 5(3) e42.

Woolhouse, M. E., C. Dye, et al. (1997). *Heterogeneities in the transmission of infectious agents: implications for the design of control programs.* Proc Natl Acad Sci U S A 94(1): 338-42.

Introduction to Infectious Disease Modelling and its Applications – 2018

Session 30: Network modelling: definition, properties and application of networks to the spread of epidemics

Lecture

Objectives

- To give an introduction to the use of networks in epidemiology.
- To present simple network definitions and concepts.
- To become familiar with examples of commonly used network types.
- To introduce to methods of modelling diseases spreading through networks.
- To understand the difference between network and mass-action models.

Introduction to networks

Why use networks?

The mass action models you have seen so far are the simplest model of human behaviour. They assume that everyone meets everyone else, and that meetings are instantaneous and not repeated. This is not realistic: in large populations, we will only meet a small fraction of people, and many of our contacts will be repeated. Using network models allows more realistic patterns of contact between people to be simulated.

What are networks?

From a theoretical point of view a *network* (or *graph*) is any system that can be described in terms of two types of element:

- *Nodes* (or *vertices*, *points*, *actors*, or *individuals*): these represent the constituents of the system. They can be individual people in a contact network; cities in a transportation network; gene proteins in a regulatory network. Nodes may have associated properties, e.g. age, gender.
- *Links* (or *edges*, *lines*, or *relations*): these represent interactions between nodes. They can be conversations between people; roads between two cities; chemical reactions between proteins.

The size of the network, *N*, is the number of nodes of the network (Albert & Barabási, 2001; Wasserman & Faust, 1994).

The definition of a network is quite general and encompasses a huge range of phenomena. A network's characteristics depend on the function for which that network has been created.
There are many examples of epidemiologically relevant networks. Figure 1 shows:

A. A contact network in a school, showing interactions between children. Nodes are coloured according to the respondent's gender. The network contains social clusters and segregation based on gender (Conlan et al, 2011).
B. A British livestock network showing movements between premises on a single day. The network shows clear spatial structure and contains some nodes with many links – generally markets and slaughterhouses (Keeling et al, 2010).

C. Sexual contacts linked with Chlamydia infection. The network shows the pattern of sexual encounters in Manitoba, Canada. The network has been measured using a snowball sample: during contact tracing, cases were asked about their partners, who were followed-up (Wylie & Jolly, 2001). The network shows a large degree of heterogeneity: most individuals have few encounters, but node S is particularly active.

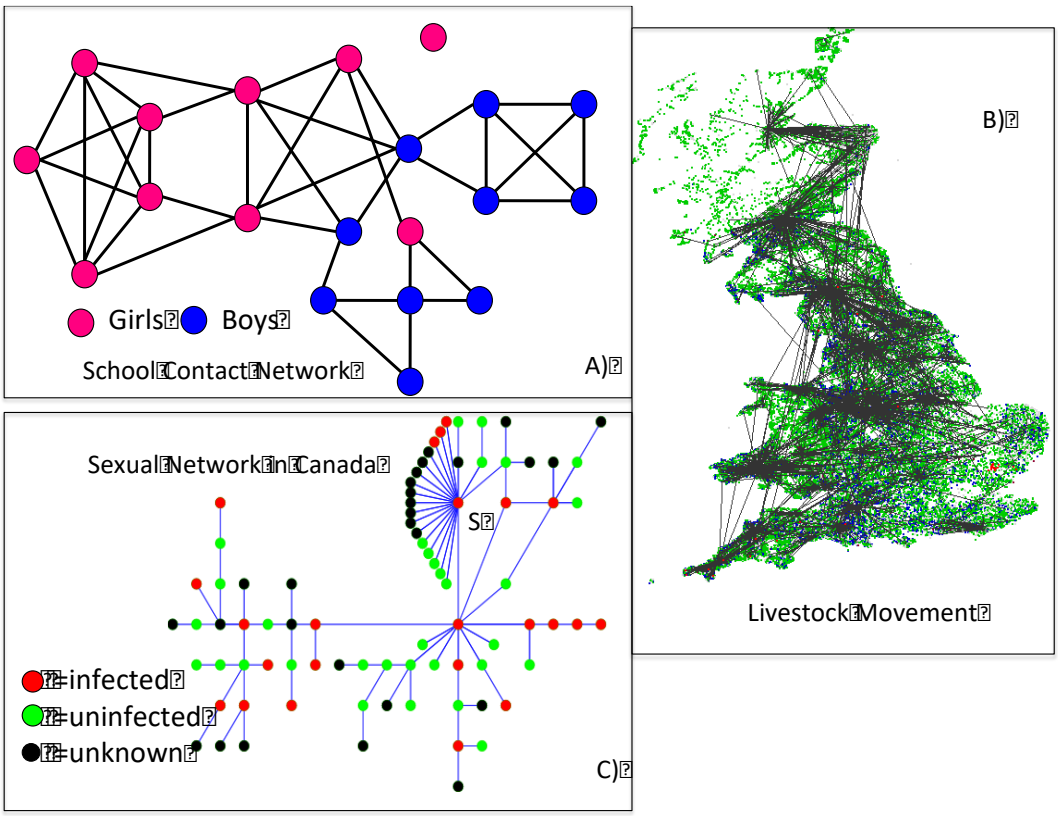


Figure 1: Three examples of epidemiologically relevant networks

Network measures

In this section we define some topological characteristics of the nodes in a network (Wasserman & Faust, 1994). We will illustrate them using the example networks in figure 2.

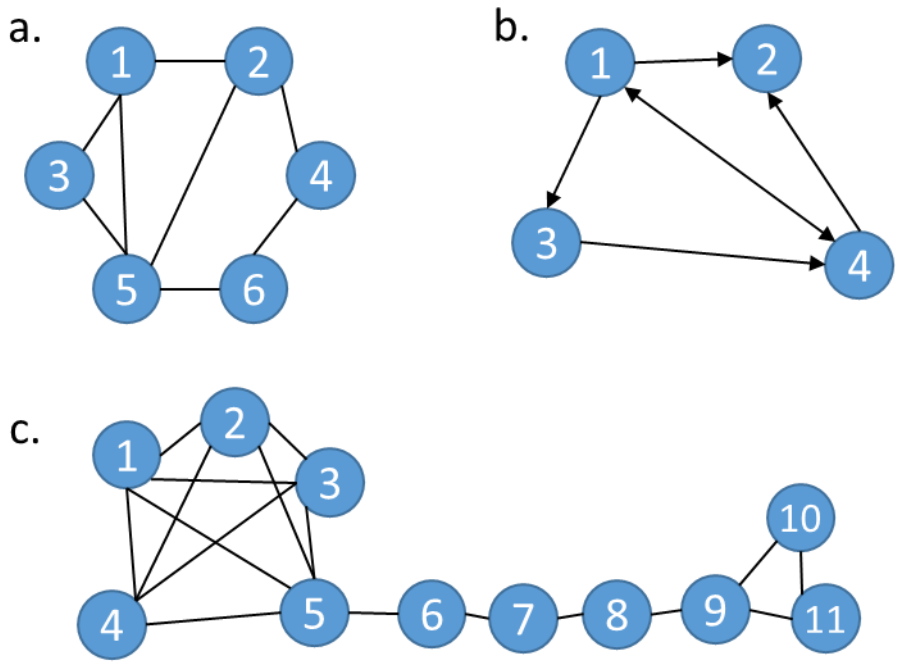


Figure 2: Example networks

Distance and diameter

The *distance* between two nodes on a network is defined as the length of the shortest path connecting them. If we consider that at each time step an infected individual can spread a disease to just the individuals with whom she is in contact, evaluating the distance between a seed (index case) and a given node can provide an estimate of the time that node can get infected. In figure 2c, the distance between nodes 1 and 4 is 1, and between nodes 1 and 8 is 5. The *diameter* of a network is the largest distance contained within the network – i.e. the longest shortest path. The diameter of the network in figure 2 is 6.

Undirected and directed networks

In undirected networks, if *i* is linked to *j*, then *j* is also linked to *i*. In directed networks, *i* can be linked to *j*, without *j* necessarily being linked to *i*. Graphically, the link from *i* to *j* is indicated by arrow starting from *i* and pointing to *j* (Figure 2b). One example of a directed network would be twitter, with directed links indicating that a person is following another person.

Weighted networks

In weighted networks, the information about the link includes not only the nodes where it begins and ends, but also some information about the strength of the relation itself, quantified as a positive number. E.g. a contact network that takes account of the duration of an interaction between individuals. A weighted matrix may be directed or undirected.

Degree

The *degree* of a node is the number of contacts (or *neighbours*, or *partners*) a node has. We denote by k_i the degree of node *i*. In Figure 2a, node 4 has degree 2, and node 5 has degree 4.

The *neighbourhood* of node *i*, *V(i)*, is the set contacts of node *i*. The neighbourhood of node 2 in Figure 2a is nodes 1, 4 and 5.

In a directed network we can extend this definition to specify a node's in-degree, k^{in}, and out-degree, k^{out}, as the number of links from and to a node respectively. In Figure 2b, node 1 has in-degree 1 and out-degree 2. In an un-directed network, $k_i^{out} = k_i^{in}$.

In a weighted network, we can define a node's degree by ignoring the link weights. We can also define the *strength s* of a node as the sum of the weights of its links.

The importance of the presence of weak and strong links has been discussed by Granovetter (Granovetter, 1983): strong ties are often friendship ties, important in conveying ideas and behaviour; weak ties, however, can be important in transmitting information (or infection) beyond immediate social groups.

Clustering and transitivity

In many social networks, 2 friends of a person *i* are likely to know each other. This property is called *transitivity*. Transitivity is not an egocentric property like node degree but depends on the characteristics of the neighbourhood of a node. We define the *clustering coefficient* of a network, ϕ, as the probability that two randomly chosen contacts of a node will be connected. The clustering coefficient can be approximated as 3 times the number of triangles, divided by the number of *triplets* (single node with edges to two other nodes). In Figure 2a, the clustering coefficient is 6/15.

Betweenness and Centrality

Depending on their position within a network, some nodes are more important than others. We introduce two measures of a node's importance, or centrality.

The *betweenness* of a node is the fraction of shortest paths in the network that pass through that node. Nodes with high betweenness are central to efficient communication between other nodes and have a large influence on the flow of information through the network. The *s-centrality* of a node is the number of nodes that can be reached in *s* or fewer steps from that node.

Whilst informative, measures of a node's relative importance do not always agree. For example, if we compare nodes 1 and 7 in Figure 2c, node 1 has a higher 2-centrality (5 vs 4), whereas node 6 has a higher betweenness (the shortest path between all nodes <7 and all nodes >7 passes through node 7).

Statistical description of networks

So far, we have worked with an example network that we have full knowledge about. In real life, in the majority of cases, we have only partial information available. Furthermore, we may not be interested in the exact details of the network.

Alternatively, we introduce the *degree distribution* of a network, in which *P(k)* is the probability that a node has degree *k*. The average degree $\bar{k}$ can be calculated from the degree distribution *P(k):*

$$\bar{k} = \frac{1}{N}\sum_i k_i = \sum_k P(k)k = \frac{2\,E}{N}$$

where E is the number of links in the network.

We can construct a network of size *N* with a desired degree distribution *P(k),* as follows (Albert & Barabási, 2001):

1. Generate a sequence of *N* numbers from the degree distribution *P(k)*. These numbers will give the degree of each node. If $\sum_i k_i = 2E$ is odd, increase or decrease one of the numbers by 1 (Figure 3a).
2. Associate each node with one of the numbers from the degree distribution computed above. At this point the network looks like a family of hedgehogs: each node has k_i sticking-out quills (Figure 3b).
3. Connect at random the ends of pairs belonging to different vertices (Figure 3c).

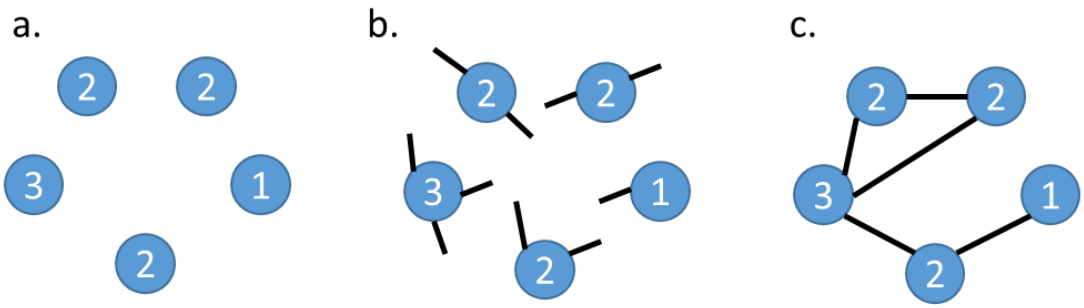


Figure 3. Construction of a network with average degree = 2. Numbers in nodes indicate node degree size.

Common network types

Erdős-Rényi networks

The Erdős-Rényi network is a common example of a random network. Amongst a set of *N* nodes, there is a link between any pair of nodes with probability *p*. The probability that a node has *k* contacts follows a binomial distribution:

$$P(k,N) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

Using the properties of the binomial distribution it follows that the mean degree is given by: $\bar{k} = p(N-1)$. For instance, in Figure Xc, $p = 0.5$, $N = 5$, and $\bar{k} = 2$.

As *N* increases with $\bar{k}$ held constant (i.e. decreasing *p* accordingly), the binomial distribution can be approximated by a Poisson distribution with

$$P(k) = \frac{\bar{k}^k e^{-\bar{k}}}{k!}$$

The clustering coefficient of an Erdős-Rényi network approaches zero as the network becomes large, since the chance that any two given nodes are connected, p, becomes

vanishingly small ($p \approx \frac{\bar{k}}{N}$). Thus, locally the network looks like a tree. In Figure Xc, if N was increased to 10 or 100, keeping $\bar{k} = 2$, then p would fall to 0.22 and 0.02 respectively.

Small world networks

Social networks are often found to have low diameter (any person can be reached from any person using few intermediaries) and high clustering (the friend of my friend is likely to also be my friend). Random networks have a short diameter, since the random nature of connections introduces many short-cuts across the network. However, the Erdős-Rényi random network has low levels of clustering. Therefore, to describe a social network, we seek a network that has high clustering but short average path length.

In 1998 Watts and Strogatz (Watts & Strogatz, 1998) developed a network model that included these 2 characteristics of realistic social networks: a shorter diameter than a regular lattice (a regular network with edges between nearby nodes only, e.g. the first network in Figure 4), and a higher clustering coefficient than an Erdős-Rényi network.

The model starts with a ring lattice of *N* nodes as in Figure 4 where each node is connected to its nearest neighbours. Links are then rewired: for each link, with probability *p* one of its ends is rewired at random, creating *long range links*.

If *p=0* the lattice remains unperturbed. If *p=1* the resulting network is completely random. When the *p* is small the short-cuts substantially reduce the diameter but make little difference to the clustering coefficient of the network. In this intermediate region we have a *small world network*, with high clustering but low diameter.

The rationale behind the Watts-Strogatz model is sociological: we have a lot of friends in our physical neighbourhood, with a high probability that they know each other, but occasionally we create links with individuals far away. This idea was tested experimentally by Stanley Milgram (Travers & Milgram, 1969). Randomly chosen people from Omaha were asked to transfer a package via personal contacts to a named individual in Boston. The average number of steps to cross half a continent was found to be around 6.

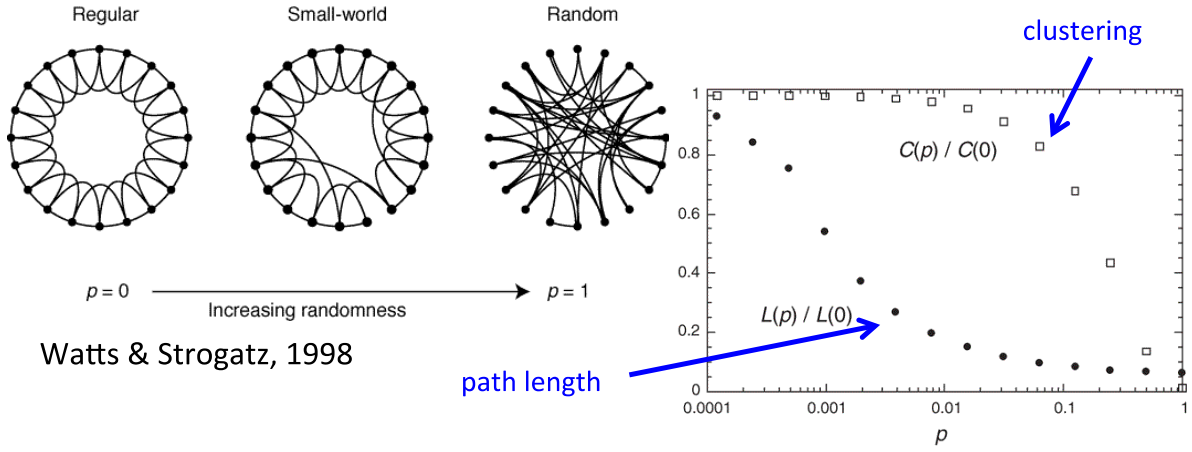


Figure 4: Watts-Strogatz small world network model. In the plot on the right the clustering coefficient and the diameter are shown relative to the unperturbed (lattice) case

Scale-free networks

So far we have considered networks whose size does not change over time. In contrast, the Barabasi-Albert model (Barabási & Albert, 1999) used to study growing networks, such as the WWW. In this model the number of nodes is not constant: new nodes appear and new links are created. There are two mechanisms at the basis of the model:

1. *Growth*: at each time step a node is added to the network. The node forms *m* new links with existing nodes.
2. *Preferential attachment:* the probability of linking to node *i* at time *t, P(i,t)* is not uniform, but is proportional to the current degree that node:

$$P(i,t) = \frac{k_i(t)}{\sum_i k_i\,(t)}$$

This *preferential attachment* increases the probability of linking to nodes that are already well-connected (in this way, the "rich get richer"). The degree distribution approaches

$$P(k) = k^{-\gamma}$$

where γ is a positive constant that depends on the number of new links created at each time. In the case of *WWW* $\gamma \approx 3$. For measured sexual partnership networks $\gamma \approx 2$ (Liljeros et al, 2001). Most measured scale-free networks have exponent $2 < \gamma < 3$.

Network with such "power law" degree distributions contain some highly connected nodes (called *hubs*) and a large fraction of nodes with few links. The network is said to be "heavy tailed", meaning that the possible existence of highly connected nodes cannot be ignored (as it could in the case of a bell shaped degree distribution); for such networks, the average degree is not a measure that by itself can be used to describe network characteristics, but large *k* fluctuations should be taken in account. When $\gamma > 3$, the network is characterised by a set of small sub-graphs highly internally connected. For $\gamma < 2$, the network is dense, the fraction of hubs is large, and the clustering coefficient is high (Albert & Barabási, 2001).

Epidemics on networks

In order to understand the effect of network structure on the spread of infection through a population, we compare the case of homogeneous mass-action mixing (the normal assumption for SIR models) with a network in which each individual has degree *k.*

The assumption for homogenous mixing is that in each time step each individual has a small and equal probability of making contact with everybody else (Keeling & Eames, 2005). Alternatively we can say that although the *number* of contacts is fixed and equal for everybody, at each time step the *identities* of the contacts change. Under this assumption, with τ as the per-contact transmission rate, the force of infection for individual *i* in the mass-action model is given by:

$$\lambda_{i,mean-field} = \tau \times k \times \frac{I(t)}{N}$$

A network model of the same population considers that not only is the *number* of contacts fixed, but so is their *identity*. The force of infection for a network model is given by:

$$\lambda_{i,network} = \tau \times (\text{number of infected neighbours of node } i)$$

We note that in the mass-action model, each susceptible individual experiences the same force of infection whereas in the network model each susceptible individual's force of infection depends on their neighbourhood.

Because of the local nature of transmission within networks, epidemics in networks frequently cause local depletion of susceptibles. Furthermore, infected individuals are connected to other infected individuals, and are thus obstructing each other. Therefore, epidemics spread slower in a network model than in a mass-action model. This is illustrated by the fact that in a network an epidemic has a lower value of R_0 than it would in the equivalent mass-action model.

Epidemics behave differently on different network topologies. We here summarise some of the characteristics of epidemics spreading in different types of network.

- *Lattice:* In a lattice model, nodes are positioned on a regular grid. The network is highly clustered. Epidemic spread is local, and the initial growth is reduced compared to a mass-action model due to the local depletion of suceptibles. The epidemic proceeds as a wave. The absence of long-range links makes the spread slower than other network models: its takes many steps to pass between two randomly selected nodes.
- *Erdős-Rényi:* This network is characterized by the absence of clustering. The growth rate and final size are smaller than mass-action equivalents.
- *Small world:* This kind of network shows high clustering and small diameter. The transmission remains prevalently local, but long-range links connect distant regions of the network thus increasing synchronisation of the epidemic.
- *Scale-free:* The extreme heterogeneity in a scale free network has a striking effect in the spread of epidemics. Hubs act as super spreaders, and are important targets for vaccination or other interventions. The small diameter indicates that an epidemic needs only a few steps to spread throughout the network.

Pair approximation model

When infection spreads through a network, the infection status of connected nodes becomes correlated. The pair approximation model of the spread of infection through a generic network takes into account the presence of these correlations (Keeling, 1999; Keeling & Eames, 2005). The model keeps track not only of the number of susceptible, infected and recovered individuals, but also of the number of links between individuals of different types.

We indicate by [S], [I], and [R] the number of individuals of each type, and we suppose that all the nodes have the same degree, *k*. The number of new infections per unit time depends on both on the transmission rate τ and on the number of contacts between susceptible and infected individuals, written as [SI]. The SIR model on a network can be written as:

$$\frac{d[S]}{dt} = -\tau\,[SI]$$

$$\frac{d[I]}{dt} = \tau\,[SI] - r[I]$$

$$\frac{d[R]}{dt} = r[I]$$

where *r* is the recovery rate.

To run the model, we need to know how the number of susceptible-infected pairs, [SI], changes over time. We can identify four events that modify the number of S-I pairs (Figure 5). Focusing on a given pair of individuals:

1. *Within pair recovery:* The infected member of the pair recovers with rate r; an S-I pair becomes an S-R pair, reducing [SI] by 1.
2. *Within pair transmission:* The susceptible member of the pair gets infected by the infected member with rate τ; an S-I pair becomes an I-I pair, reducing [SI] by 1
3. *Infection from outside the pair:* a S-S pair can become an S-I pair if it is part of an S-S-I triple; this takes place at rate τ, increasing [SI] by 1.
4. *Infection from outside the pair:* an S-I pair can become an I-I pair if it is part of an I-S-I triple; this takes place at rate τ, reducing [SI] by 1.

Taking account of all possible mechanisms, the evolution of the number of [SI] pairs is given by:

$$\frac{d[SI]}{dt} = \tau\left([SSI] - [ISI] - [SI]\right) - r[SI]$$

We have now expressed the behaviour of the number of S-I pairs in terms of various triples. We could continue this process – modelling triples in terms of quadruples, and so on - but this quickly becomes messy. Instead, we approximate the number of triples of type [ABC] by:

$$[ABC] = \frac{[AB][BC](k-1)}{k[B]}$$

Here, we have assumed that all individuals have the same degree, and have ignored clustering. The model can be extended to include degree heterogeneity and clustering (Keeling & Eames, 2005).

We can calculate the basic reproduction number R_0 using the pair-approximation model. We do this by evaluating the initial exponential growth rate of the epidemic. We must take into account the fact that the first few infected individuals are likely to be connected to each other, thus reducing their potential for onward transmission. We find that:

$$R_0 = (k-2)\frac{\tau}{r}$$

Compared to the mass-action model we see that growth is slower and the final size of the epidemic is smaller on a network (Figure 5).

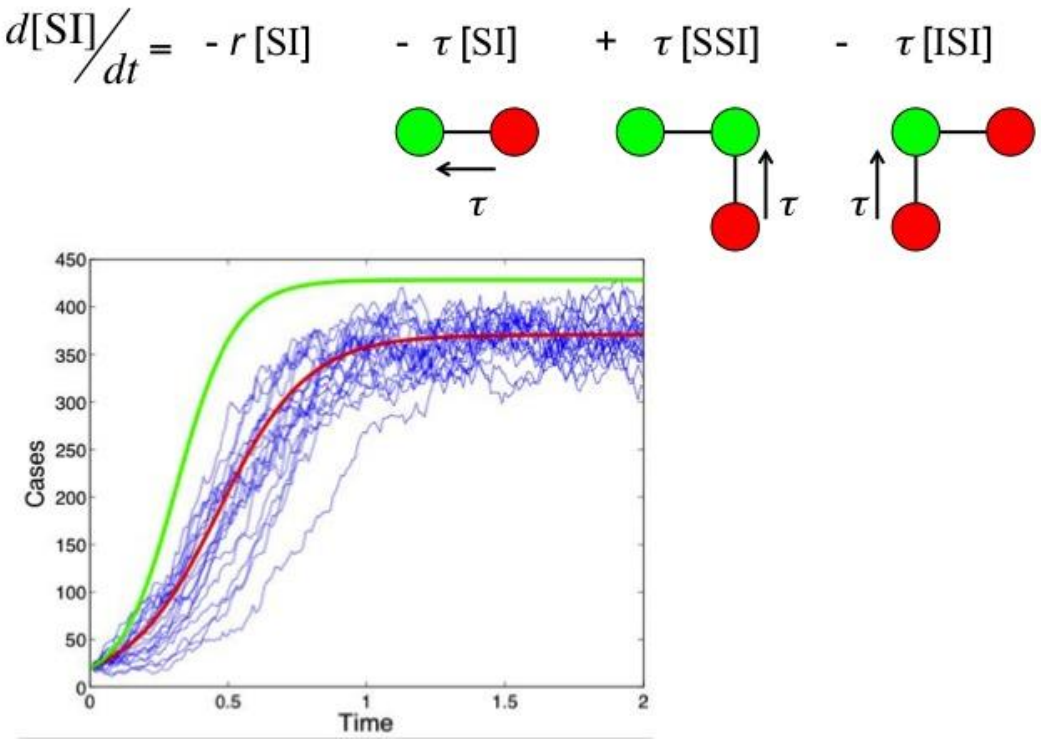


Figure 5 Top: events leading to a change in the number of S-I pairs. Bottom: model comparison for an SIS epidemic on a network, showing mass-action model (green), pair-approximation model (red), stochastic simulation (blue) for the same values of $\boldsymbol{\tau}$, *r*, and *k*

Optional reading: The adjacency matrix

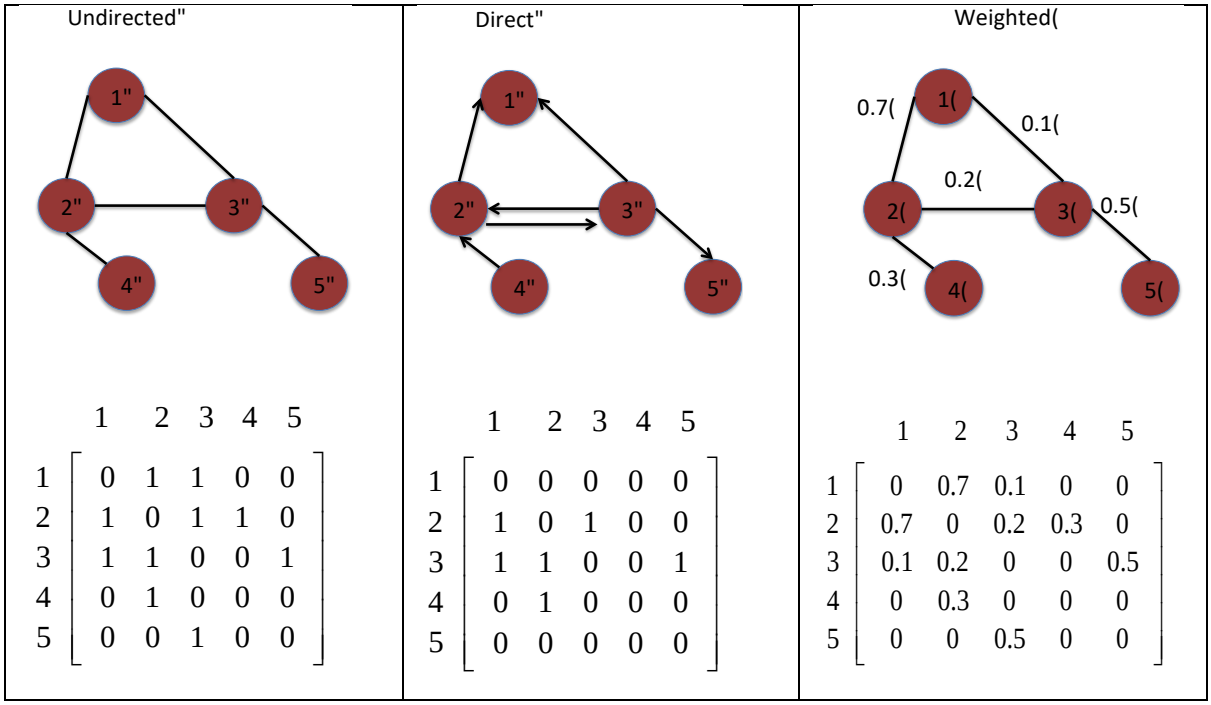
A useful way of representing a network is by using an *adjacency matrix*. In a network *G* of size *N*, we associate each node with an index *i* between 1 and N. In the simplest case, the adjacency matrix, *A,* is a non-negative matrix with elements $\{A_{ij}\}$:

$$A_{ij} = \begin{cases} 1 \text{ if there is a link between } i \text{ and } j \\ 0 \text{ otherwise} \end{cases}$$

We assume that a node cannot be connected to itself, i.e. $A_{ii}=0$ for all *i*.

In Figure 6 we show how the adjacency matrix can be adapted to deal with different assumptions about link properties (Wasserman & Faust 1994):

- *Undirected:* if *i* is linked to *j*, then *j* is also linked to *i*. The adjacency matrix is symmetric, i.e. $A_{ij} = A_{ji}$ for all *i* and *j*. E.g. Facebook friendships.
- *Directed: i* can be linked to *j*, without *j* necessarily being linked to *i*. In this case the adjacency matrix is asymmetric, i.e. A_{ij} might not equal A_{ji}. Graphically, the link from *i* to *j* is indicated by arrow starting from *i* and pointing to *j*. E.g. twitter followers.
- *Weighted:* the information about the link includes not only the nodes where it begins and ends, but also some information about the strength of the relation itself, quantified as a positive number. A weighted matrix may be directed or undirected. The entries in the adjacency matrix are non-negative real numbers. E.g. a contact network that takes account of the duration of an interaction between individuals.



$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \end{array} \qquad \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array} \qquad \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & 0.7 & 0.1 & 0 & 0 \\ 0.7 & 0 & 0.2 & 0.3 & 0 \\ 0.1 & 0.2 & 0 & 0 & 0.5 \\ 0 & 0.3 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0 \end{bmatrix} \end{array}$$

Figure 6: Different types of network involving the same set of nodes

In most epidemiological situations, any one of these three network types can be usefully applied: in the case of a sexually transmitted infection (STI), an undirected unweighted network can show sexual partnerships; a directed network can be used to show chains of infection within the network; weights can be added to represent the frequency of intercourse.

The adjacency matrix shows whether a link between the nodes *i* and *j* exists ($A_{ij} \neq 0$) or not ($A_{ij} = 0$). In an unweighted network, the adjacency matrix can also be interpreted in another way: non-zero entries indicate the number of paths of length 1 connecting nodes. This interpretation is useful when studying the connectivity and reachability of a network. A network is *connected* if it cannot be split into multiple sub-networks with no links between them. A node is said to be *reachable* if there exists at least one chain of links connecting the node to all the other nodes in the network.

Before using the adjacency matrix to study the connectivity properties of the network we recap matrix multiplication:

Matrix multiplication

Consider two $N \times N$ matrices *A* and *B* with elements $\{a_{ij}\}$ and $\{b_{ij}\}$. Their product $C = A \times B$ is an $N \times N$ matrix whose elements are $c_{ij} = \sum_{k=1}^{N} a_{ik} b_{kj}$. For example, using 2×2 matrices:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$$

$$C = A \times B = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \times \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{pmatrix}$$

We can multiply more than 2 matrices together by first multiplying the first two matrices and then multiplying the product by the following matrix and so on. As with normal multiplication, $(A \times B) \times C = A \times (B \times C)$ [1].

The matrix product allows us to evaluate powers of the adjacency matrix. The elements of matrix $A^2 = A \times A$ represent the number of paths of length 2 between nodes. If we consider the undirected matrix of Figure 6, the matrix A^2 is:

$$A^2 = A \times A = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \times \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 1 & 1 & 1 \\ 1 & 3 & 1 & 0 & 1 \\ 1 & 1 & 3 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix}$$

To understand the interpretation of this, consider the third row of A^2: the elements indicate the number of path of length 2 between node 3 and the rest of the nodes. So a path exists between node 1 and 3 (passing through node 2); between node 2 and 3 (passing through node 1); there are 3 paths from node 3 back to itself; there is a path of length 2 from node 3 to node 4; there is no such path between nodes 3 and 5.

The diagonal matrix elements are all non-zero. These elements give the number of closed paths (loops) from each node back to itself.

In the matrix A^3, diagonal elements indicate the number of triangles (closed loops of length 3) whilst non-diagonal elements indicate paths of length 3.

$$A^3 = A \times A^2 = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \times \begin{pmatrix} 2 & 1 & 1 & 1 & 1 \\ 1 & 3 & 1 & 0 & 1 \\ 1 & 1 & 3 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 4 & 4 & 1 & 1 \\ 4 & 2 & 5 & 3 & 1 \\ 4 & 5 & 2 & 1 & 3 \\ 1 & 3 & 1 & 0 & 1 \\ 1 & 1 & 3 & 1 & 0 \end{pmatrix}$$

The first 3 diagonal elements are equal to 2, indicating the existence of the triangle 1-2-3-1. The value 2 indicates that the triangle can be traversed either “clockwise” or “anti-clockwise”. Nodes 4 and 5 are not part of any triangles but are connected by a path of length three.

If a network is connected it means that there is at least one path connecting any 2 possible nodes. Since the longest path in network is composed by *N-1* links, we can say that a network is connected if the matrix

$$R = \sum_{l=1}^{N-1} A^l = A + A^2 + \cdots + A^{N-1}$$

has all elements non-zero. A zero element, R_{ij}=0, indicates that there is no path connecting nodes *i* and *j,* and therefore they are members of different sub-networks. These definitions and properties can also be extended to directed and weighted networks.

Studying the connectivity of the network can give some idea of the possible extent of disease propagation. If a network is unconnected, epidemics cannot spread between sub-networks, so will be confined to just a portion of the population.

[1] But note that, unless *A* and *B* are symmetric, generally $A \times B \neq B \times A$. This is important in the case of directed networks since the multiplication order has an implication in the direction of the paths.

Degree

We denote by k_i the degree of node *i*. k_i can be evaluated as the sum of the *i*-th row in the adjacency matrix, i.e. $k_i = \sum_j A_{ij}$. Considering the first matrix in Figure 6, $k_1 = 2; k_2 = k_3 = 3; k_4 = k_5 = 1$.

The *neighbourhood* of node *i*, *V(i)*, is the set contacts of node *i*: in the case of the first network in Figure 6, $V(2) = \{1,3,4\}, V(5) = \{3\}$.

In a directed network we can extend this definition to specify a node's in-degree, k^{in}, and out-degree, k^{out}, as the number of links from and to a node respectively. k^{out} corresponds to the row sum of the adjacency matrix and k^{in} to the column sum:

$$k_i^{out} = \sum_j A_{ij}$$

$$k_i^{in} = \sum_j A_{ji}$$

In an un-directed network, $k_i^{out} = k_i^{in}$.

In a weighted network, we can define a node's degree by ignoring the link weights. We can also define the *strength s* of a node as the sum of the weights of its links:

$$s_i = \sum_j A_{ij}$$

Clustering

Recalling the properties of adjacency matrix powers we can express the clustering coefficient as

$$\phi = \frac{\sum_i A_{ii}^3}{\sum_{i \neq j} A_{ij}^2}$$

References

Albert, R., & Barabási, A.-L. (2001). Statistical mechanics of complex networks. *Rev. Mod. Phys.*, *74*, 47–97. doi:10.1103/RevModPhys.74.47

Barabási, A.-L., & Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, *286*, 509–512. doi:10.1126/science.286.5439.509

Conlan, A.J.K, Eames, K.T.D. et al (2011) Measuring social networks in British primary schools through scientific engagement *Proc. Roy. Soc. B* doi: 10.1098/rspb.2010.1807

Granovetter, M. (1983). The strength of weak ties: A network theory revisited. *Sociol. theory*, *1*, 201–233.

Keeling, M.J. (1999). The effects of local spatial structure on epidemiological invasions. *Proc. Roy. Soc. B*, *266*, 859–867. doi:10.1098/rspb.1999.0716

Keeling, M. J., & Eames, K. T. (2005). Networks and epidemic models. *J. R. Soc. Interface*, *2*, 295–307.

Keeling, M.J., Danon, L. et al (2010). Individual identity and movement networks for disease metapopulations. *Proc. Natl. Acad. Sci USA 107*, 8866–8870. doi:10.1073/pnas.1000416107

Liljeros, F., Edling, C.R. et al (2001). The Web of Human Sexual Contacts. *arXiv.org*. doi:10.1038/35082140

Travers, J. & Milgram, S. (1969). An Experimental study of the small world problem.

Sociometry. Retrieved from http://www.uvm.edu/~pdodds/files/papers/others/1969/travers1969.pdf

Wasserman, S. & Faust, K. (1994) Social network analysis. *Cambridge University Press.*

Watts, D.J., & Strogatz, S.H. (1998). Collective dynamics of “small-world” networks. *Nature*, *393*, 440–442. doi:10.1038/30918

Wylie, J.L. & Jolly, A.M. (2001). Patterns of chlamydia and gonorrhea infection in sexual networks in Manitoba, *Sex. Transm. Dis. 28*, 14-24.

Introduction to Infectious Disease Modelling and its Applications - 2018

Session 31: Modelling HIV Transmission and Control

Lecture

Objectives

By the end of this session you should have:

- Appreciated the heterogeneity in the distribution of HIV worldwide
- Appreciated how modelling studies of HIV have evolved with the pandemic
- Seen how a simple HIV model can be used to estimate R_0 and to predict the shape and timing of the epidemic
- Seen how simple models of HIV/STI co-infection can be used to predict the potential impact of cofactor STIs on the HIV epidemic
- Seen how modelling has been used to explore changing role of STI treatment for HIV prevention
- Seen a couple more examples … male circumcision and HAART

The heterogeneity in the distribution of HIV worldwide

One of the most important questions in the epidemiology of HIV/AIDS is why there is so much heterogeneity in the prevalence of HIV worldwide. Around 33 million people were living with HIV in 2008 and most live in sub-Saharan Africa (UNAIDS, 2009). Even within Africa there are large differences in prevalence between countries and by age and gender (Figure 1). Prevalence is particularly high in Eastern and Southern Africa and incidence is particularly high in young women (UNAIDS, 2008).

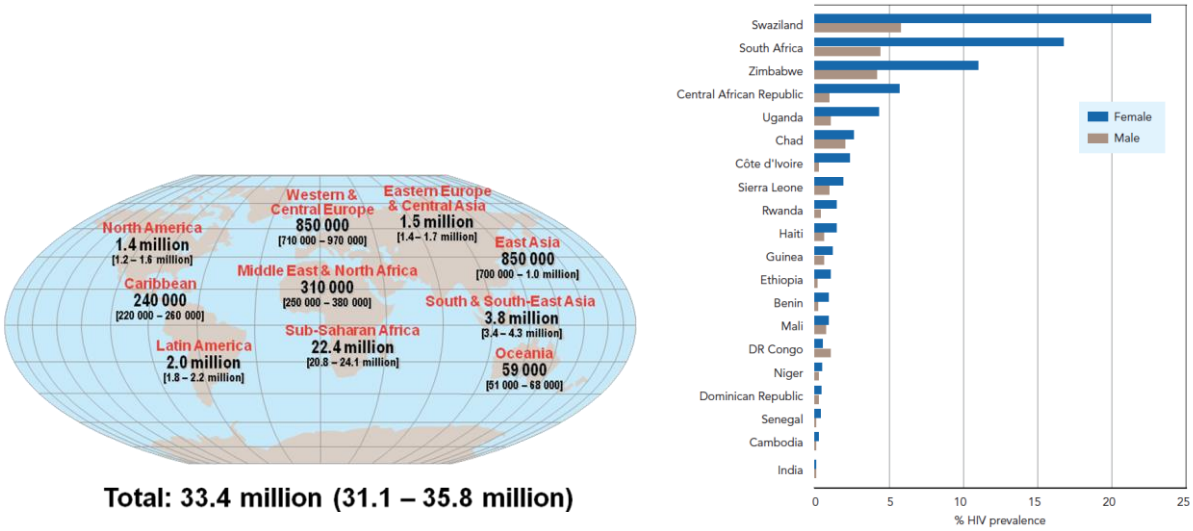


Figure 1 Left: Number of adults and children living with HIV, 2008 (UNAIDS, 2009). Right: HIV prevalence (%) among 15-24 years old, by sex, selected countries, 2005-2007 (UNAIDS, 2008)

How modelling studies of HIV have evolved with the pandemic

The evolution of mathematical modelling studies of the sexual transmission of HIV has developed in line with our increasing understanding of its spread and control. Studies in the late 1980s and early 1990s tended to be more theoretical, identifying the importance of the

various risk factors for the transmission of HIV, such as partner change rates, heterogeneity in risk behaviour and mixing between risk groups (Anderson, Gupta et al, 1991; Hyman and Stanley, 1988; Van Druten, Reintjes et al, 1990).

These early studies were crucial in showing how little was known about the identified risk factors and in guiding the subsequent data collection. Models in the early literature were also presented with the intention that they could be used in the future to estimate the impact of control strategies, initially primarily as a tool for HIV researchers (Bailey, 1994; Hyman and Stanley, 1988), but later for use by policy makers (Rehle, Saidel et al, 1998; Van der Ploeg, Van Vliet et al, 1998). The early modelling studies showed that the impact of interventions on HIV incidence would be non-linear (as for most infectious disease control) and that the stage of the HIV epidemic would be important. Impact on HIV incidence would be larger when the basic reproduction number (R_0) of HIV was close to unity, than at higher reproductive potentials (Morris and Dean, 1994; Rowley and Anderson, 1994). It also showed that intervention early in the epidemic and targeted at higher risk individuals would be most effective (Garnett and Anderson, 1995; Over and Piot, 1996; Stover and O'Way, 1995).

The focus of many of these modelling studies was on sub-Saharan Africa as it became apparent that this was the most severely affected region. However as early as 1990, modelling studies had suggested that HIV was already so widespread in parts of sub-Saharan Africa, that large scale behaviour change in the general population and greatly increased resources, would be required to control the HIV epidemic (Auvert, Moore et al, 1990; Stover and O'Way, 1995).

As scientific understanding of the risk factors for HIV transmission improved and the data required to parameterise and validate these models became more readily available, modelling was increasingly used to predict the impact of more realistic prevention strategies, to interpret the results of empirical HIV control trials, and to warn of potentially perverse outcomes of HIV prevention activities. We will see examples of some of these modelling studies later in the lecture.

As most new infections occur in sub-Saharan Africa via heterosexual transmission, we will focus on this today.

Simple transmission models of Human Immunodeficiency Virus/AIDS

AIDS is the disease caused by the retrovirus human immunodeficiency virus (HIV) (Coffin, Haase et al, 1986). HIV primarily targets CD4 cells and without treatment this leads to the collapse of the host immune system and ultimately death. The clinical syndrome was called `AIDS) in 1982 and four years later the causative virus was named HIV-1 (Coffin, Haase et al, 1986). HIV has been the focus of a large number of modelling studies since its discovery, some of which are described below. Many of the insights gained from modelling the transmission of short-duration curable STIs also apply to HIV, but there are also important differences between HIV/AIDS and the STIs considered so far that we will need to address.

Most importantly there is currently no cure for HIV so that in the absence of treatment, once the AIDS stage of HIV infection is reached, death quickly follows. Also, in the absence of other 'cofactors' for HIV transmission, HIV is typically much less infectious (Boily, Baggaley et al, 2009; Powers, Poole et al, 2008; Wawer, Gray et al, 2005) than short-duration bacterial STIs, potentially lowering its R_0 compared to short-duration bacterial STIs. However, HIV is infectious for far longer than the short-duration STIs (CASCADE, 2000; Morgan, Mahe et al, 2002; Todd, Glynn et al, 2007), increasing its R_0 relative to short-duration bacterial STIs. Depending on the research question, we may also need to consider

that the infectiousness of HIV-infected individuals varies markedly with time since infection (Wawer, Gray et al, 2005).

Let's start by adapting the simple two-activity group SIS model for gonorrhoea we used in the STI lecture by removing the possibility of cure, adding AIDS compartments, and an AIDS-related mortality rate, μ (Figure 2).

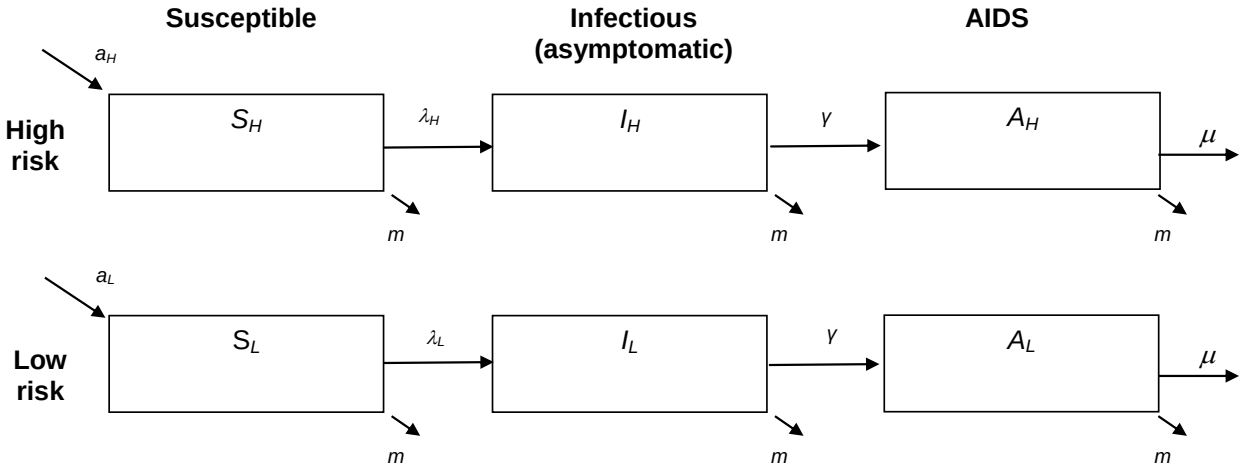


Figure 2 A simple model of HIV transmission and progression.

As we are modelling the long term dynamics of a persistent infection that leads to death we also need to model the rate of recruitment into sexually active age groups, '*a'* otherwise we will run out of susceptibles, and also the rate of non-AIDS mortality among sexually active individuals, *m*. Different assumptions can be made, but here we assume that the proportion of newly sexually active individuals recruited into each of the two activity groups (high and low) remains constant over time and is equal to the proportion of the population in the high and low-activity groups at *t=0*. For simplicity, we'll ignore all other population heterogeneities such as age and gender, assume proportionate mixing, assume that the period between infection and infectiousness is very short in comparison with the period of infectiousness, assume no sexual activity takes place in the AIDS stage, and assume that individuals do not change activity group over their lifetime.

The model equations for the high-activity group are shown below. Note that, except for $n_H(0)$, the proportion of individuals in the high-activity group at t = 0, the explicit time dependence '*(t)'* has been omitted to shorten the equations.

$$\frac{dS_H}{dt} = a_H N - \lambda_H S_H - mS_H$$

$$\frac{dI_H}{dt} = \lambda_H S_H - \gamma I_H - mI_H$$

$$\frac{dA_H}{dt} = \gamma I_H - \mu A_H - mA_H$$

$$a_H = a\,n_H(0)$$

$$\lambda_H = c_H \beta_p p \qquad \text{Equation 1}$$

$$p = g_H i_H + g_L i_L \qquad \text{Equation 2}$$

$$i_H = \frac{I_H}{U_H} \text{ and } i_L = \frac{I_L}{U_L}$$

$$U_H = S_H + I_H \text{, and } U_L = S_L + I_L$$

Where

S_H is the number of individuals susceptible to HIV at time *t* in high-activity group
I_H is the number of individuals infectious with HIV at time *t* in high-activity group
A_H is the number of individuals with AIDS at time *t* in high-activity group
N is the total population size
λ_H is the force of HIV infection on the high-activity group at time *t*

a is the recruitment rate, per year
$n_H(0)$ is the proportion of individuals in the high-activity group at $t = 0$
a_H is the recruitment rate into the high-activity group, per year
m is the non-HIV mortality rate among sexually active individuals, per year
γ is progression rate to AIDS, per year
μ is the death rate due to AIDS, per year
c_H is the partner change rate per year in the high-activity group
β_p is the probability of HIV transmission probability per partnership
p is the probability a new partner is HIV infectious
g_H and g_L are the probability that a sexual partner will be a member of the high and low-activity group, respectively
$i_H(t)$ and $i_L(t)$ are the prevalences of infectious individuals in the high and low-activity groups, respectively
U_H and U_L are the numbers of sexually active individuals in the high and low-activity groups, respectively

The equations for the low-activity group are obtained by replacing the subscript $_H$ with $_L$.

The R_0 of HIV infection

We can calculate R_0 for HIV in this model using the equation we derived for R_0 for gonorrhoea in a proportionately mixing population with two activity groups in the STI lecture. However we need to adjust the duration of HIV infectiousness because we are explicitly modelling non-AIDS mortality (Anderson and May, 1991). Let's assume the average life expectancy in the absence of HIV is 50 years and the average age of debut is 15 years. Therefore the average life expectancy at sexual debut is 35 years. Let's also assume the median time from HIV infection to death is 10 years (CASCADE, 2000; Morgan, Mahe et al, 2002; Todd, Glynn et al, 2007), and therefore the duration of infectiousness is 9 years if we allow one year for the (assumed) sexually-inactive AIDS stage. Therefore, assuming these events are distributed exponentially, HIV infected individuals will leave the infectious stage at a rate which is the sum of the progression rate to AIDS (γ , 1/9 per year) and the non-AIDS mortality rate among sexually active individuals (*m* , 1/35 per year). The duration of the infectious stage will be $\frac{1}{(1/9 + 1/35)}$ = 7.16 years.

The mean per-sex act HIV transmission probability over the entire period of HIV infection in low-risk HIV-discordant partnerships in rural Uganda (Wawer, Gray et al, 2005) was around 0.0016. If we assume that on average a partnership lasts around 30 sex-acts (averaging over one-off encounters, casual and regular partnerships), the per-partnership transmission probability will be around 0.05, ie $1-(1-0.0016)^{30}$ using the same methods shown in the STI

lecture notes. So if we assume the average per partner transmission probability is 0.05 and the partner change rates in the high and low-activity groups are 8 and 0.2 per year, respectively, we get the following predictions for the number of secondary infections from a high and low-activity individual in a totally susceptible population:

$$R_H = c_H \beta_p D$$
$$= 8 \times 0.05 \times 7.16$$
$$= 2.86$$

$$R_L = c_L \beta_p D$$
$$= 0.2 \times 0.05 \times 7.16$$
$$= 0.07$$

Again, for simplicity, we assume that the per-partnership transmission probability is the same in high and low-activity groups when, in reality, the probabilities may differ. To calculate R_0 we also need to re-calculate the probability that a new sexual partner will be a member of the high and low-activity group. Using the same logic we used in the STI lecture, the partner change rates above and assuming 15% of the population are in the high-activity group we calculate $g_H = 0.88$ and $g_L = 0.12$. Therefore:

$$R_0 = g_H R_H + g_L R_L$$
$$= 0.88 \times 2.86 + 0.12 \times 0.07$$
$$= 2.52$$

Predictions for HIV epidemic

Using this model we get the following predictions for the prevalence and incidence of infection over time and the cumulative number of AIDS deaths. In contrast to the steady rise to an equilibrium prevalence and incidence predicted for short-duration curable STIs, the prevalence and incidence of HIV is predicted to rise and then fall (Figure 3A). This is because HIV kills, and preferentially kills higher-risk individuals in the population. Therefore, unless new higher-risk individuals are recruited at the same rate they die, the average partner-change rate in the population will fall over time (Figure 3B). This reduction in partner change rates can occur even in the absence of explicit safer-sex interventions. If nothing else changes, such as the impact of intervention efforts, HIV prevalence will level-off when the annual number of HIV deaths and new HIV infections come into equilibrium (Figure 3C).

Falls in HIV prevalence and incidence may also be the result of intervention, but as we see HIV prevalence and incidence are also expected to change due to the natural dynamics of infection making interpretation of HIV trends difficult (Garnett, Gregson et al, 2006). Further, because incidence falls earlier than prevalence (Figure 3A), if the early years of the HIV epidemic are missed then it is also possible to see falls in the prevalence of HIV without falls in incidence as was observed in Rakai, Uganda (Wawer, Serwadda et al, 1997).

Figure 3 also shows that in our model HIV prevalence took around 45 years to reach its peak, which is relatively slow compared to the rate at which HIV epidemics have risen in Southern Africa and probably rose in East Africa (Figure 4). The HIV epidemic spread earlier in East Africa so HIV surveillance systems tended to miss the rise in prevalence (Figure 4, right panel).

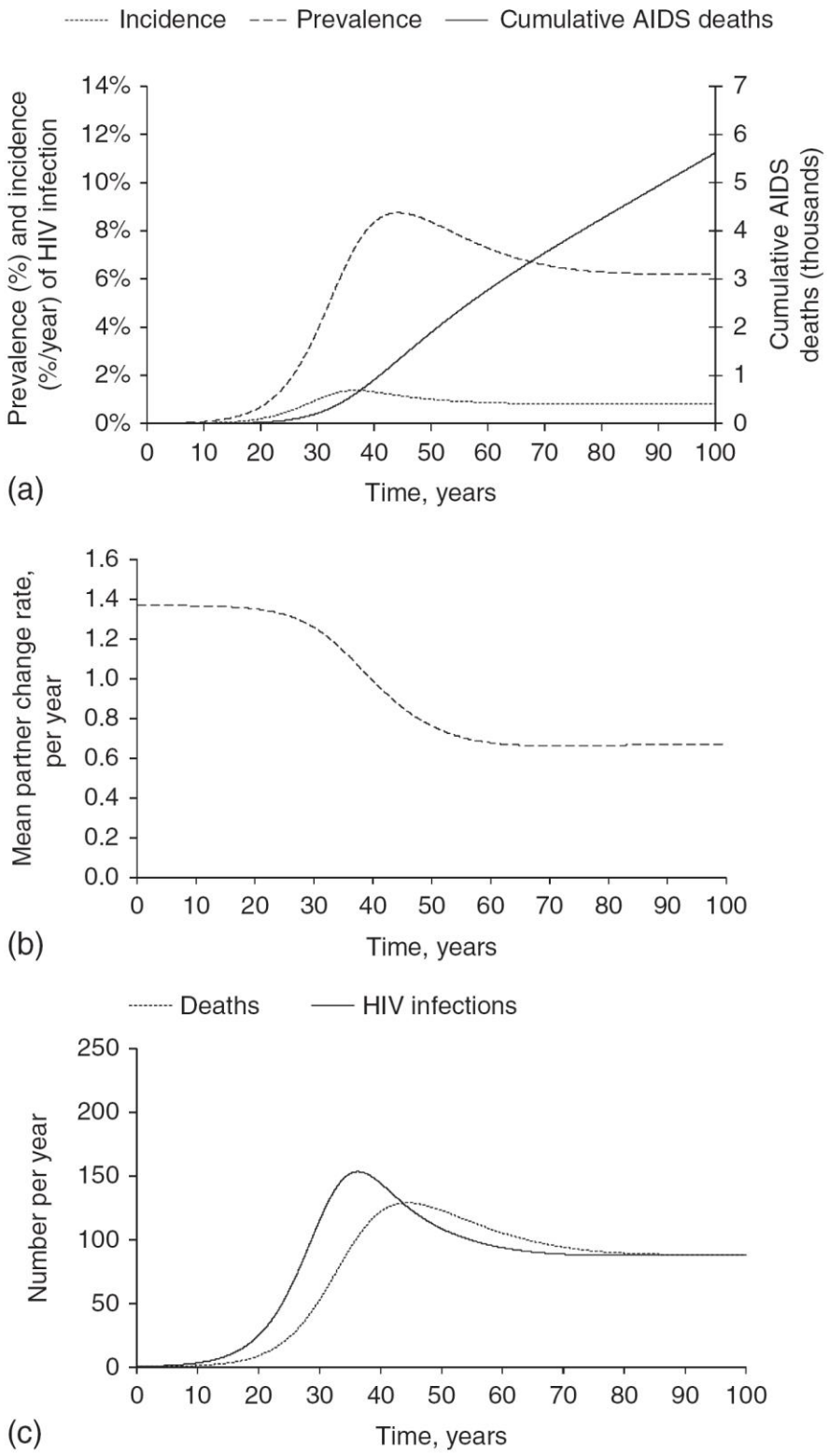


Figure 3 Predictions of (A) the incidence and prevalence of HIV and cumulative AIDS deaths, (B) mean partner change rate in the population, and (C) numbers of new HIV infections and deaths of HIV infecteds. The equations and parameter values for this model are shown above.

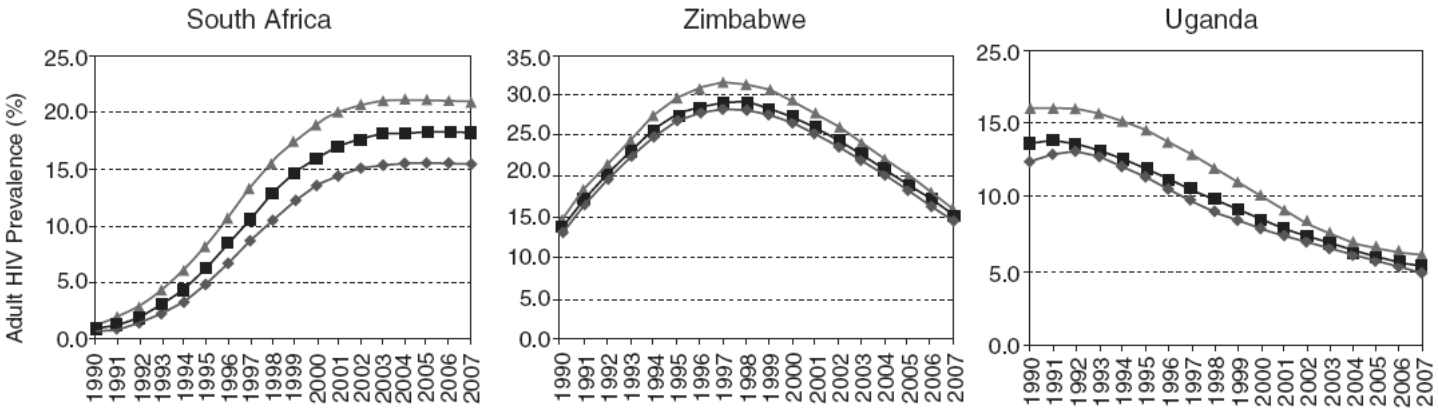


Figure 4 Estimated trend in adult HIV (with high and low estimate) prevalence in Southern Africa (15-49 years). From (UNAIDS/ WHO, 2008a; UNAIDS/ WHO, 2008b; UNAIDS/ WHO, 2008c). Note different scales on y-axis.

The model prediction (the 'numerical solution') for the doubling time of this epidemic in our model can be obtained by examining the time it takes for the number of HIV infectious individuals to increase from 1 to 2. It is about 2.8 years. We can also make a cruder analytic estimate for the doubling time of the HIV epidemic by assuming there is no heterogeneity in sexual activity. The derivation is shown in Appendix 1 and the estimated doubling time, t_d , is reasonably close to the numerical estimate:

$$T_d \approx \frac{\ln(2)}{c\ \beta_p - \gamma\ - m} = 2.7 \text{ years.}$$

The reason for the very rapid spread of HIV in Southern and Eastern Africa is not totally understood, but is likely to have been due to a combination of various biological and behaviour factors, such as low rates of male circumcision combined with high rates of other STIs that are cofactors for HIV transmission.

STI cofactors are likely to have been more important for HIV epidemics in resource-poor, than in resource-rich settings, because STI treatment services are more limited and consequently the prevalence and duration of STI infections were higher (WHO, 2001). In the next section we use a simple model of HIV/STI co-infection to explore the possible impact of cofactor STIs in increasing rate of spread of HIV, and also the likely impact of the HIV epidemic on the prevalence of cofactor STIs.

In Southern Africa, the effect of population mobility, specifically labour migration, is also thought to have been important in explaining the rapid spread of HIV (Coffee, Lurie et al, 2007). One of the effects labour migration is likely to have had is to increase the prevalence of concurrency, or the overlap of sexual partnerships.

Models of HIV/STI co-infection

In this section we use a simple model of HIV/STI co-infection to explore the possible impact of cofactor STIs (cSTIs) in increasing rate of spread of sexually transmitted HIV and describe some of the other cSTI/HIV coinfection modelling studies that have been carried out to help understand the epidemiology and control of sexually transmitted HIV in sub-Saharan Africa.

The interactions between classical STIs and HIV have been the subject of repeated review over the past couple of decades (Fleming and Wasserheit, 1999; Laga, Nzila et al, 1991; Rottingen, Cameron et al, 2001; Wasserheit, 1992). There is now a large body of evidence

from laboratory, clinical and epidemiological studies supporting the hypothesis, first voiced in 1984, that classical STIs may facilitate the spread of HIV (Piot, Quinn et al, 1984).

For the HIV-uninfected individual, cofactor STIs increase susceptibility to HIV infection owing to breaks in the skin caused by ulcers or increased presence of T-lymphocytes and macrophages in the genital tract which are targets for HIV. For the HIV-infected individual, cofactor STIs increase transmission of HIV as STIs frequently cause increased shedding of HIV in the genital tract. These effects are collectively known as STI cofactor effects. The magnitude of an STI cofactor effect is defined as the multiple by which HIV transmission will be increased in the presence of the cofactor STI.

However, translating the relative risks of cofactor STIs for HIV transmission that are measured in epidemiological studies into *per-partnership* or *per-act* transmission probabilities, and that are needed to predict the impact of cSTIs on HIV transmission using mathematical models, is not straightforward (Boily and Anderson, 1996; Hayes, Schulz et al, 1995; Korenromp, De Vlas et al, 2001). Typically the actual exposure period with ulcers or inflammation is not measured in epidemiological studies and the exposure period is frequently much longer than the actual duration of ulcers or inflammation caused by the STI (measured using survey questions such as '*Have you had genital ulcers in the last X months?'*). This differential misclassification bias will lead to an underestimate of the true per-contact relative risk (Hayes, Schulz et al, 1995). Conversely, even though most epidemiological studies adjust for confounding as best they can, the estimated effects are likely to be subject to residual confounding by poorly measured aspects of sexual risk behaviour in the study subjects, or unmeasured risk behaviours of their sexual partners (Korenromp, De Vlas et al, 2001). Residual confounding would lead to an overestimate of the true relative risk.

The *per-act* cofactor effect for genital ulcer disease (GUD) has been estimated from empirical data to be in the range of 50-300 for female-to-male transmission and 10-50 for male-to-female transmission for ulcerative STI (Hayes, Schulz et al, 1995). Recognising that these estimates may suffer from the biases listed above, if we assume the per-act cofactor effect for female-to-male and male-to-female transmission combined is 10 and that GUD was only present for 25% of the three month period in which an ulcer was reported, then a *per-contact* cofactor of 10 is equivalent to a *per-partnership* cofactor of around 3.1 (see Appendix 2)

We can predict the impact of cSTIs on the spread of HIV by adapting our simple HIV/AIDS model to explicitly model the transmission of a cofactor STI as shown in Figure 5. The model equations and parameter values are shown in Appendix 3.

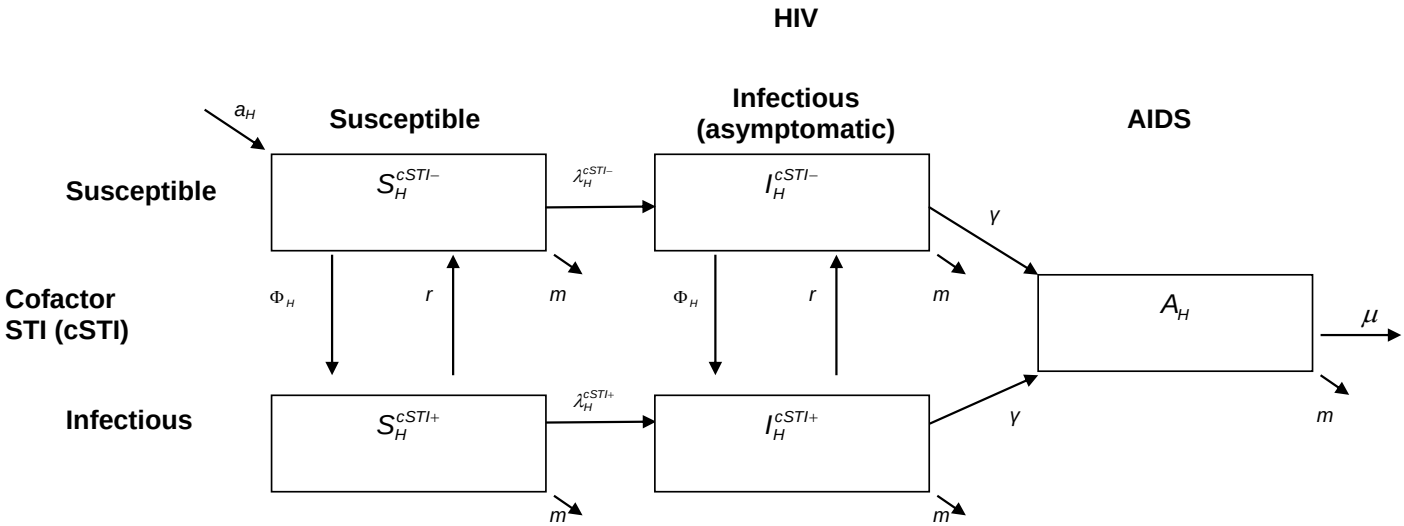


Figure 5 A simple model of HIV transmission and progression with a cofactor STI for HIV transmission. The figure illustrates the high-activity group in the model only. See Appendix 3 for model equations and parameter values.

Predictions for the impact of cofactor STIs on the HIV epidemic

The relatively slow rise in the prevalence and incidence of HIV we predicted above using the HIV/AIDS model without a cofactor STI is reproduced in Figure 6A for comparison. The addition of a cofactor STI greatly increases the rate of spread of HIV (Figure 6B) such that the HIV epidemic now peaks at around 20 years after its introduction, rather than around 45 years after introduction.

The model predicts (the numerical solution) that the addition of the cofactor STI reduced the doubling time of the HIV epidemic from 2.9 years to 1.1 years.

The interaction between HIV and other sexually transmitted infections is complex. Note that our model also predicted that the prevalence of the cofactor STI may fall during an HIV epidemic (Figure 6C) in line with the predictions of other modelling studies (Boily and Masse, 1997). This is because HIV will tend to preferentially kill the more sexually active members of the population, and these individuals are also more likely to be infected with other STIs because of the common risk-behaviour. In Uganda, a historically high HIV prevalence population, an empirical analysis suggests cSTI rates may have fallen during the HIV epidemic (Orroth, Korenromp et al, 2003). However it is difficult to disentangle the effects of HIV-attributable mortality, reduction in social disruption caused by the civil war in Uganda, volitional behaviour change due to safer-sex campaigns and improvements in STI treatment services, on STI trends.

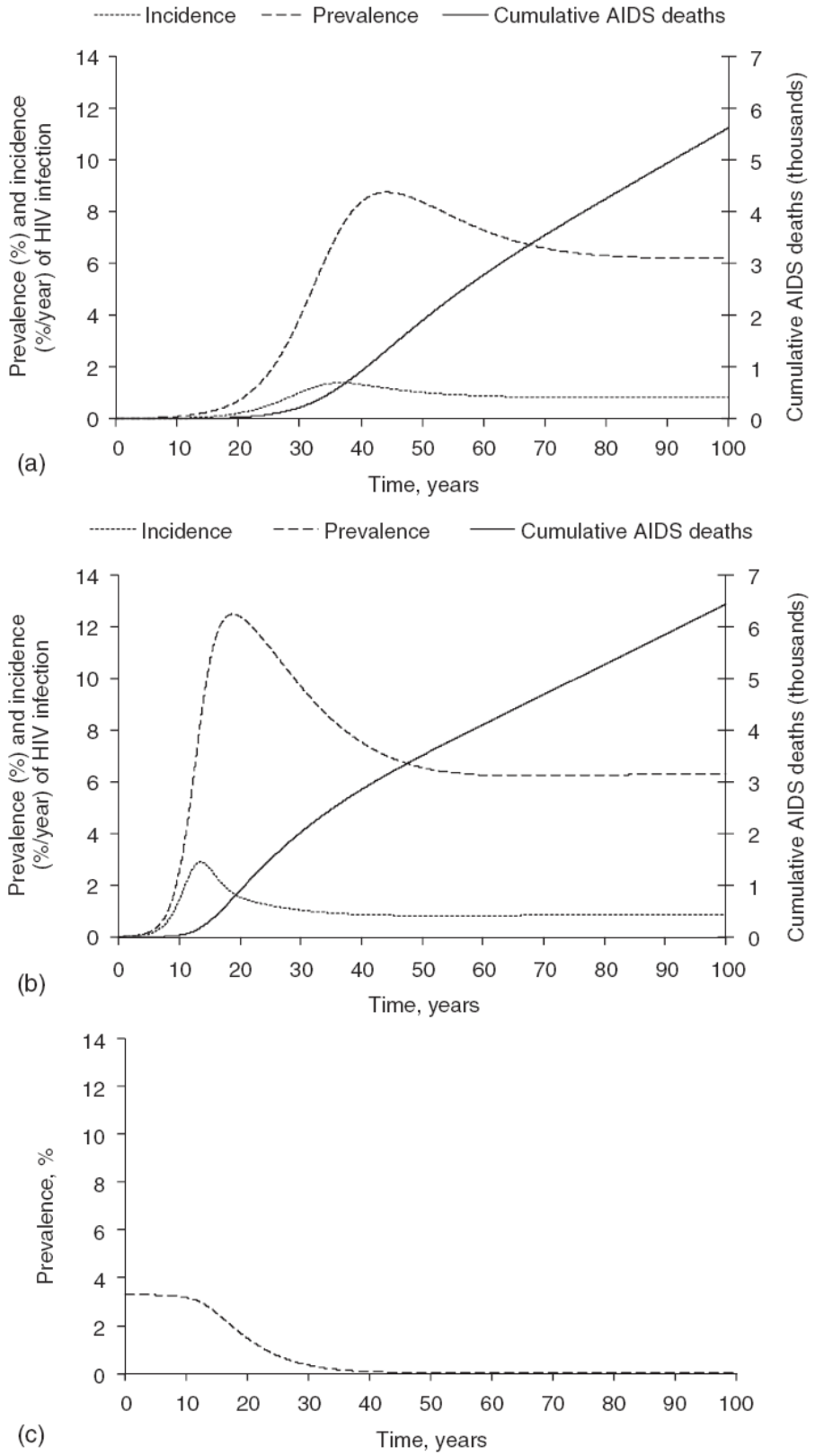


Figure 6 HIV epidemic with and without STI cofactor. Predictions of (A) the incidence and prevalence of HIV and cumulative AIDS deaths without STI cofactor, (B) the incidence and prevalence of HIV and cumulative AIDS deaths with STI cofactor, (C) the prevalence of the cofactor STI in model with STI cofactor. The model is run for 50 years before introducing HIV to allow the cSTI to reach equilibrium (ie t= -50 to t= 0), and then for another 100 years. See Appendix 3 for model equations and parameter values.

The changing role of STI treatment for HIV prevention

Many cSTI/HIV coinfection modelling studies that have been carried out to help understand the utility of the control of cofactor sexually transmitted infections on HIV transmission in sub-Saharan Africa.

Modelling has helped better understand the seemingly contrasting results of three first randomised controlled trials of STI treatment for HIV prevention in sub-Saharan Africa. Early modelling studies of the interaction between HIV and other STIs showed that the role of STIs, particularly STIs causing genital ulcers, was likely to be critical in the rapid spread of HIV in sub-Saharan Africa (Robinson, Mulder et al, 1997; Van der Ploeg, Van Vliet et al, 1998) and a community randomised-controlled trial in Mwanza, Tanzania showed that improved clinic based syndromic STI treatment reduced the incidence of HIV infection by around 38% (95% confidence interval= 15% to 55%) among the general population (Grosskurth, Mosha et al, 1995; Habbema and De Vlas, 1995; Hayes, Grosskurth et al, 1995). Soon after however, trials of STI mass treatment in Rakai, Uganda, and of an information, education, and communication intervention with and without improved syndromic STI treatment in Masaka, Uganda, showed no significant effect on the incidence of HIV infection (Kamali, Quigley et al, 2003; Wawer, Sewankambo et al, 1999), despite similar reductions in the prevalence of curable STIs in all three sites. Detailed data analysis and modelling of the interventions and the populations in which the trials took place showed that population differences in sexual behaviour, curable STI rates, and HIV epidemic stage could explain most of the contrast in HIV impact observed between the three trials, and concluded that STI management is likely to be an effective HIV prevention strategy in populations with a high prevalence of curable STIs, particularly in an early HIV epidemic (Orroth, Korenromp et al, 2003; White, Orroth et al, 2004).

Figure 7 illustrates this by showing the model prediction that the impact of either syndromic or mass STI treatment on HIV incidence would have been larger in the early HIV epidemic in Mwanza, than in the later HIV epidemics in Rakai or Masaka in Uganda (White, Orroth et al, 2004).

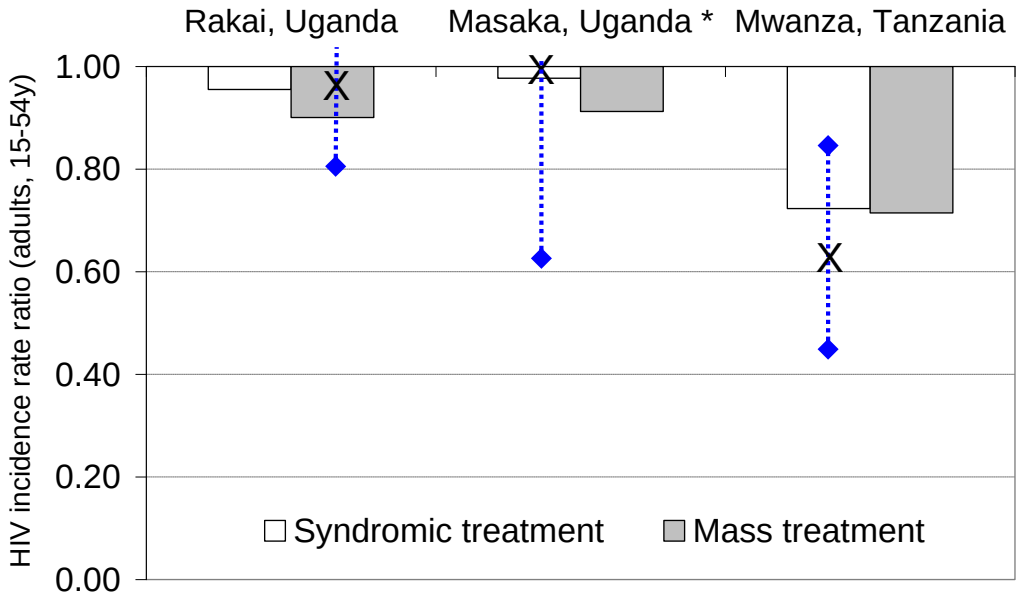


Figure 7 Simulated (bars) and observed ('X's with 95% confidence intervals) impact of syndromic and mass STI treatment interventions on HIV incidence in three populations in East Africa. Adapted from (White, Orroth et al, 2004).

Modelling studies have suggested that the impact of STI treatment on HIV incidence will depend on the characteristics of the population (Korenromp, Van Vliet et al, 2000; Korenromp, White et al, 2005), the characteristics of those reached (Boily, Lowndes et al, 2000), and the stage of the HIV epidemic (Korenromp, Bakker et al, 2002b; Korenromp, Bakker et al, 2002a).

Modelling was also helpful in highlighting that the proportion of HIV incidence due to curable STIs (the population attributable fraction, or 'PAF') was likely to fall as the HIV epidemic matures. This is because AIDS mortality and behaviour change (if it had occurred) reduced curable STI rates, and because, as HIV prevalence increases, a larger proportion of HIV transmission occurs outside higher-risk groups, in groups with lower rates of curable STIs (Korenromp, Bakker et al, 2002b; Korenromp, Bakker et al, 2002a). Conversely, later in the HIV epidemic, the proportion of HIV incidence due to the incurable STI *Herpes simplex* Virus type-2 (HSV-2) would rise, due to the fall in classical STI rates and increased HSV-2 ulceration among immuno-compromised HIV-infected individuals (Abu-Raddad, Magaret et al, 2008; Blower and Ma, 2004; Freeman, Orroth et al, 2007; Korenromp, Bakker et al, 2002c). This is illustrated using the predicted trends in the PAFs of herpes simplex virus type 2 and chancroid on HIV incidence in four cities in sub-Saharan Africa in Figure 8 (Freeman, Orroth et al, 2007).

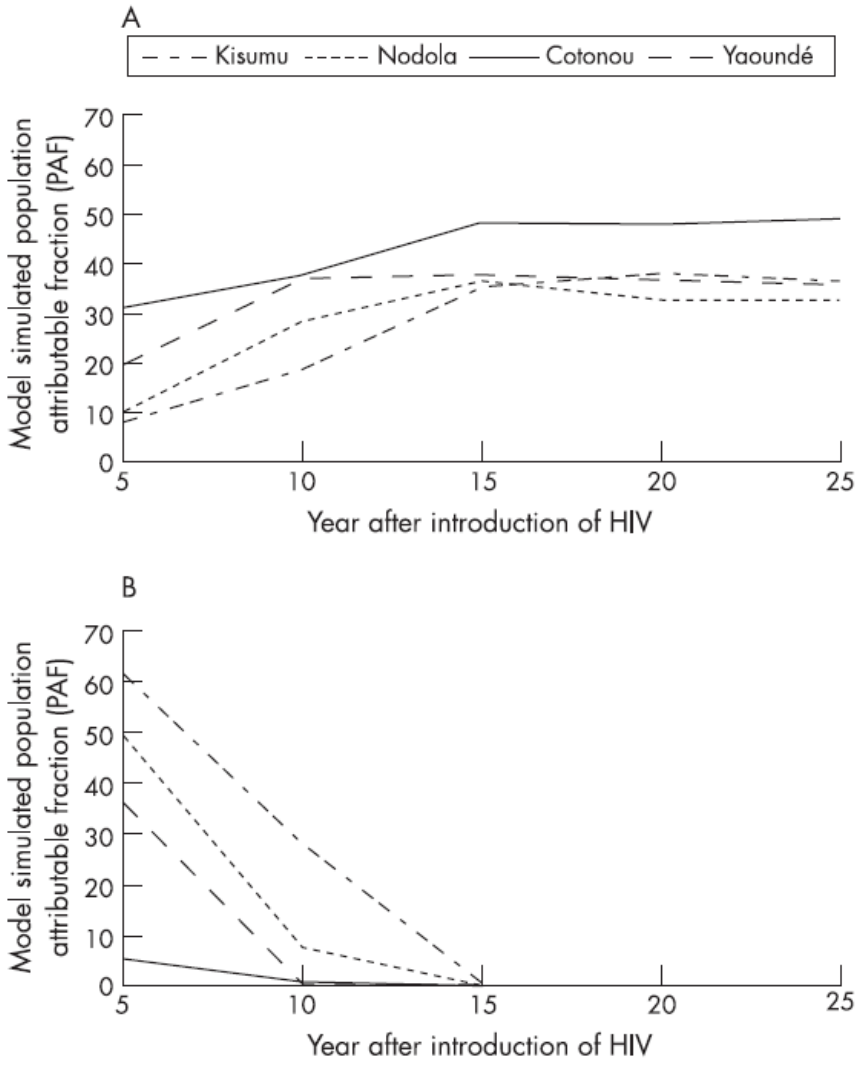


Figure 8 Simulated proportion of new adult (15–49 years) HIV infections due to (A) herpes simplex virus type 2 (B) chancroid in four cities in sub-Saharan Africa, by year after the introduction of HIV. Model: *STDSIM* (Freeman, Orroth et al, 2007).

Although these studies suggest that the relative impact of curable STI treatment will be lower later in generalized HIV epidemics, other work has shown that curable STI treatment may remain cost-saving in generalised epidemics because the absolute impact remains high (White, Orroth et al, 2008).

Recently, the potential of HSV-2 treatment as an HIV prevention strategy has been explored. However, in general, the modelling work has suggested that a substantial impact of HSV-2 therapy on population HIV incidence is unlikely because it would require high coverage and long-duration therapy, or very high symptom recognition and treatment seeking behaviour (Baggaley, Griffin et al, 2009; White, Freeman et al, 2008), and all three of randomised controlled trials of herpes suppressive therapy failed to show an impact on the HIV transmission (Celum, Wald et al, 2008; Watson-Jones, Weiss et al, 2008). This lack of impact may be because any STI cofactor effect is small, or because there was insufficient herpes suppression due to inadequate drug dosage or adherence, or because the mechanism of action of acyclovir does not adequately control the effect of HSV-2 on HIV transmission (Celum, Wald et al, 2008; Watson-Jones, Weiss et al, 2008).

More encouragingly perhaps, Figure 9 shows some initial modelling results that suggest that an effective prophylactic HSV-2 vaccine, if developed, could have a substantial impact on HIV incidence (Freeman, White et al, 2009).

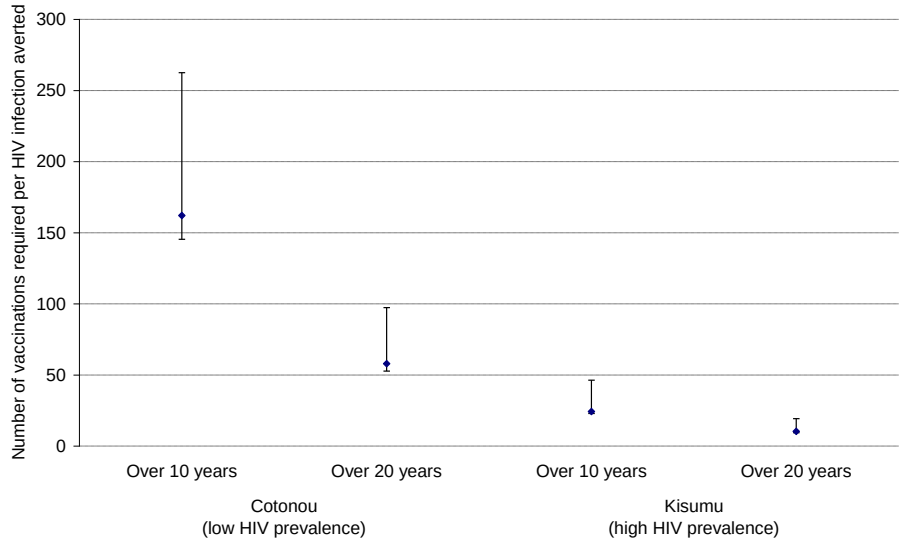


Figure 9 Simulated number of vaccinations per HIV infection averted over 10 and 20 years using a hypothetical prophylactic HSV-2 vaccine, in a low and high HIV prevalence African city (all ages). The modelled scenario assumes an increase from 0% to 70% in vaccine coverage over 5 years in 14-29 year olds using a vaccine with 10 year duration of effect on susceptibility and reactivation in males and females. Three scenarios are shown corresponding to the 'weak' (30% , ⊤), 'moderate' (75% , ♦) and 'strong' (90% , ⊥) vaccine efficacy. See (Freeman, White et al, 2009) for full details.

A couple more examples … male circumcision and HAART

With the advent of treatment for HIV, mathematical modelling has been used to explore the impact of antiretroviral treatment (ART) on HIV incidence. Early work showed that without reduction in infectiousness and in risk behaviour, ART would tend to increase HIV incidence (Garnett and Anderson, 1996; Gupta, Anderson et al, 1993) as HIV infected individuals would remain healthy and sexually active for longer. However as ART has been shown to reduce viral load considerably and therefore should reduce the infectiousness of HIV-

infected individuals, more recent work has focussed on the critical importance of risk-compensation (the possible increase in risk behaviour due to reduction in the perceived severity of the consequences of HIV infection) (Blower, Gershengorn et al, 2000; Gray, Xianbin Li et al, 2003), the benefits of integrating HIV prevention and treatment activities (Baggaley, Garnett et al, 2006; Salomon, Hogan et al, 2005), and the potential of universal voluntary HIV testing and ART (Granich, Gilks et al, 2008). The work by *Baggaley* and colleagues illustrated an obvious, but potentially worrying, effect of HIV treatment programmes, which is that population HIV prevalence will increase due to falling mortality rates (Figure 10, left). As such, a rapid rise in prevalence when treatment programmes are initiated should be regarded as an indictor of the likely success of a treatment programme, rather than a failure. More pessimistically, the study by *Baggaley* and colleagues suggested that HIV treatment was unlikely to have much effect on the prevention of new HIV infections, because of increases in the sexual activity of treated patients and the development of resistance (Figure 10, right). Other modelling studies have been more optimistic, but these differences highlight that we need more data to constrain the models.

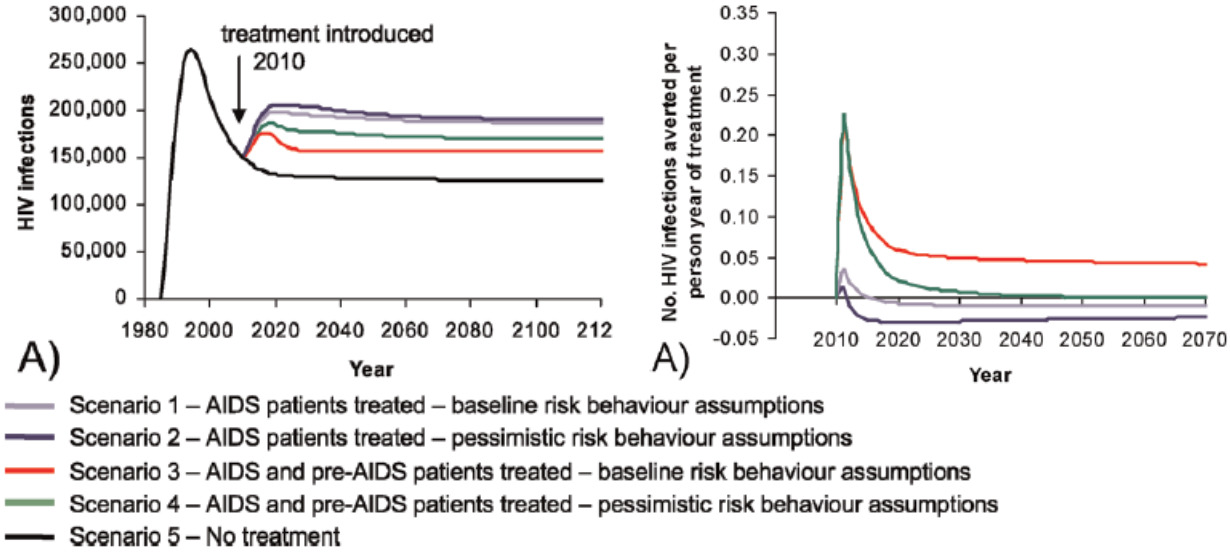


Figure 10 Predicted total number of prevalent HIV infections (left) and incident HIV infections averted per person-year of treatment (right) over time in a high prevalence mature sub-Saharan Africa epidemic for various assumptions of behaviour change of treated patients and treatment targeting (Baggaley, Garnett et al, 2006)

Male circumcision is likely to have been very important in explaining the heterogeneous spread of HIV in sub-Saharan Africa (Buve, Carael et al, 2001; Moses, Bradley et al, 1990; Orroth, Freeman et al, 2007; Weiss, Quigley et al, 2000). Three individual randomised controlled trials in Africa have shown male circumcision to reduce HIV-incidence in males by 50-60% (Auvert, Taljaard et al, 2005; Bailey, Moses et al, 2007; Gray, Kigozi et al, 2007). Modelling work has been very useful in projecting the potential population-level impact implied by these individual-level trial findings. The predictions of these modelling studies have been very consistent (Gray, Azire et al, 2004; Hallett, Singh et al, 2008; Nagelkerke, Moses et al, 2007; White, Glynn et al, 2008; Williams, Lloyd-Smith et al, 2006), primarily because the impacts of the empirical trials were so similar. Overall, modelling studies suggest that circumcision could prevent 6 million HIV infections over 20 years in sub-Saharan Africa and will be cost-saving because of the reduction of future antiretroviral treatment costs (Auvert, Kahn et al, 2007; Expert Group on Modelling the Impact and Cost of Male Circumcision for HIV Risk Reduction, 32767; Williams, Lloyd-Smith et al, 2006) and Figure 11.

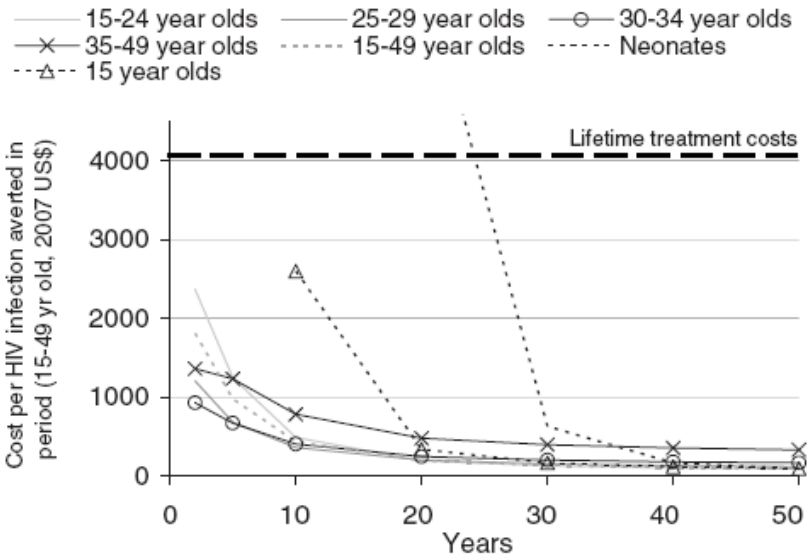


Figure 11 Simulated cost-effectiveness of male-circumcision for HIV prevention in sub-Saharan Africa, by target age group. Cost required to avert one HIV infection in adults aged 15–49 years, over time in 2007US$. For comparison, an adjusted estimate of the lifetime treatment costs of an HIV infection in sub-Saharan Africa (Stover, Bertozzi et al, 2006), is shown as a horizontal dashed line. From (White, Glynn et al, 2008).

Summary

This lecture has highlighted several topics in modelling HIV transmission and control namely the heterogeneity in the distribution of HIV worldwide, how modelling studies have evolved with the HIV pandemic, how a simple HIV model can be used to estimate R_0 and to predict the shape and timing of the epidemic, how simple models of HIV/STI co-infection can be used to predict the potential impact of cofactor STIs on the HIV epidemic, how modelling has been used to explore changing role of STI treatment for HIV prevention and the role of male circumcision and HAART on HIV spread and control.

Unfortunately, with 33 million men, women and children infected with HIV (UNAIDS, 2009), the mathematical modelling of HIV control will continue to be an active area of research for the foreseeable future.

Appendix 1 Derivation of HIV epidemic doubling time

If we assume that the partner change rates in the two groups are identical and equal to 8 partners per year (the doubling time of the epidemic is primarily determined by the partner change rate in the high-activity group) and that at the start of the epidemic all individuals are essentially susceptible, then from Equation 2, $p = \frac{I}{U}$ and from Equation 1 $\lambda = c\ \beta_p \frac{I}{U}$, so the expression for the rate of change in the number of infectious individuals is:

$$\begin{aligned}
\frac{dI}{dt} &= \lambda S - \gamma I - mI \\
&= c\beta_p \frac{I}{U} S - \gamma I - mI \\
&\approx c\beta_p \frac{I}{S} S - \gamma I - mI \\
\frac{dI}{dt} &\approx \left(c\beta_p - \gamma - m\right) I
\end{aligned}$$

Equation 3

Integrating Equation 3 to get the number of infectious individuals at time t:

$$I(t) \approx I(0) e^{\Lambda t}$$

Where $\Lambda = c\ \beta_p - \gamma - m$ = the initial growth rate of the epidemic. By definition, at the initial doubling time of the epidemic, T_d, the number of infectious individuals will have doubled. Hence $I(T_d) = 2 \times I(0)$ and therefore:

$$\begin{aligned}
\frac{2I(0)}{I(0)} &\approx e^{\Lambda T_d} \\
2 &\approx e^{\Lambda T_d}
\end{aligned}$$

Taking the natural logarithm of both sides of the equation yields

$$\ln(2) \approx \Lambda T_d$$

and rearranging

$$\begin{aligned}
T_d &\approx \frac{\ln(2)}{\Lambda} \\
T_d &\approx \frac{0.69}{8 \times 0.05 - 1/9 - 1/35} \\
T_d &\approx 2.7 \text{ years}
\end{aligned}$$

Appendix 2 Calculation of the per-partnership STI cofactor effects on HIV transmission

HIV transmission probability per act without GUD	0.0016	
Per-act GUD cofactor magnitude	10	
Observation period	3	months
Number of acts in observation period	30	(10/mth)
% of observation period with GUD	25%	
Transmission probability in observation period without GUD	0.047	
Transmission probability in observation period with GUD	0.145	
Relative risk	3.1	

Where:

Transmission probability in observation period without GUD
$= 1 - (1 - 0.0016)^{30}$ = 0.047
Transmission probability in observation period with GUD
$= 1 - (1 - 0.0016)^{30*75\%}(1 - 0.0016*10)^{30*25\%}$ = 0.145

Appendix 3 Equations for model of HIV/STI co-infection

The model equations for the high-activity group are shown below. Note that, except for $n_H(-50)$, the proportion of individuals in the high-activity group at *t* = -50, the explicit time dependence '*(t)'* has been omitted to shorten the equations. The model is started at t= -50 to allow the cofactor STI (cSTI) to reach equilibrium before HIV is introduced at t=0.

For cSTI uninfected individuals:

$$\frac{dS_H^{cSTI-}}{dt} = + \text{Recruited} + \text{cSTIRecovered} - \text{HIVInfected} - \text{cSTIInfected} - \text{OtherDeaths}$$

$$= +a_H N + rS_H^{cSTI+} - \lambda_H^{cSTI-} S_H^{cSTI-} - \Phi_H S_H^{cSTI-} - mS_H^{cSTI-}$$

$$\frac{dI_H^{cSTI-}}{dt} = + \text{HIVIntro} + \text{HIVInfected} + \text{cSTIRecovered} - \text{cSTIInfected} - \text{cSTIUninfectAIDScases} - \text{OtherDeaths}$$

$$= +HIVIntro + \lambda_H^{cSTI-} S_H^{cSTI-} + rI_H^{cSTI+} - \Phi_H I_H^{cSTI-} - \gamma I_H^{cSTI-} - mI_H^{cSTI-}$$

For cSTI infected individuals:

$$\frac{dS_H^{cSTI+}}{dt} = + \text{cSTIInfected} - \text{cSTIRecovered} - \text{HIVInfected} - \text{OtherDeaths}$$

$$= +\Phi_H S_H^{cSTI-} - rS_H^{cSTI+} - \lambda_H^{cSTI+} S_H^{cSTI+} - mS_H^{cSTI+}$$

$$\frac{dI_H^{cSTI+}}{dt} = + \text{cSTIInfected} + \text{HIVInfected} - \text{cSTIRecovered} - \text{cSTIInfectedAIDScases} - \text{OtherDeaths}$$

$$= +\Phi_H I_H^{cSTI-} + \lambda_H^{cSTI+} S_H^{cSTI+} - rI_H^{cSTI+} - \gamma I_H^{cSTI+} - mI_H^{cSTI+}$$

For individuals with AIDS:

$$\frac{dA_H}{dt} = + \text{cSTIUninfectedAIDScases} + \text{cSTIInfectedAIDScases} - \text{AIDSdeaths} - \text{OtherDeaths}$$

$$= +\gamma I_H^{cSTI-} + \gamma I_H^{cSTI+} - \mu A_H - mA_H$$

Where

S_H^{cSTI-} is the number of individuals HIV-uninfected and cSTI-uninfected at time *t* in high-activity group

S_H^{cSTI+} is the number of individuals HIV-uninfected and cSTI-infected at time *t* in high-activity group

I_H^{cSTI-} is the number of individuals who are HIV-infectious and cSTI-uninfected at time *t* in high-activity group

I_H^{cSTI+} is the number of individuals who are HIV-infectious and cSTI-infected at time *t* in high-activity group

A_H is the number of individuals with AIDS at time *t* in high- activity group

N is the total population size

λ_H^{cSTI-} is the force of HIV infection on cSTI-uninfected individuals at time *t* in high-activity group

λ_H^{cSTI+} is the force of HIV infection on cSTI-infected individuals at time *t* in high-activity group

Φ_H is the force of cSTI infection on individuals at time *t* in high-activity group

a is the rate of recruitment into sexually active age groups, per year
$n_H(-50)$ is the proportion of individuals in the high-activity group at *t* = -50
$a_H = a\,n_H(-50)$, is the recruitment rate into the high-activity group, per year
m is the non-HIV mortality rate among sexually active individuals, per year
γ is progression rate to AIDS, per year
μ is the death rate due to AIDS, per year
r is recovery rate from cSTI infection, per year
HIVIntro introduces an HIV infected person at t=0.

The force of cofactor STI infection

The force of cofactor STI infection on individuals in the high-activity group Φ_H will depend on the partner change rate, c_H, the transmission probability per partnership, β_{cSTI}, and the probability a partner is infected, p_{cSTI}. For simplicity, because we assume no sexual activity during AIDS and because this stage is relatively short, we don't track cSTI status in the AIDS stage. Note that we now make a distinction between the transmission probability of the cofactor STI, β_{cSTI} and the transmission probability of HIV, β_{HIV}.

$$\Phi_H = c_H \beta_{cSTI} p_{cSTI}$$

Where

c_H is the partner change rate per year in high-activity group
β_{cSTI} is the cSTI transmission probability
p_{cSTI} is the probability partners are infected with cSTI and this is the weighted average of the cSTI prevalence among sexually actives individuals in the high, f_H, and low, f_L activity groups (see STI chapter).

$$p_{cSTI} = g_H f_H + g_L f_L$$

Where

g_H and g_L are the probability that a sexual partner will be a member of the high and

low-activity group, respectively,

$$f_H = \frac{S_H^{cSTI+} + I_H^{cSTI+}}{U_H}$$

$$U_H = S_H^{cSTI-} + I_H^{cSTI-} + S_H^{cSTI+} + I_H^{cSTI+}$$

Where U_H is the number of sexually active individuals in the population in the high-activity group at time *t*.

The equations for the low-activity group can be written by replacing the $_H$ subscript with $_L$.

The force of HIV infection

In our model the force of HIV infection will depend on whether one or both of the partners are infected with the cofactor-STI. Alternative assumptions can be made, but here we assume the per-partnership STI cofactor for HIV acquisition and separately HIV transmission, *x,* is = 3.1. β_{HIV} is the probability of transmission of HIV infection during a sexual partnership if neither partner is cSTI-infected. Therefore if one partner is cSTI-infected, the transmission probability in an HIV-discordant partnership is $x\beta_{HIV}$ and if both partners are cSTI-infected the transmission probability in an HIV-discordant partnership is $x^2\beta_{HIV}$.

There are four equations for the force of HIV infection to consider. One for each activity group and cSTI-status of HIV-susceptible partner, ie λ_H^{cSTI-}, λ_H^{cSTI+}, λ_L^{cSTI-} and λ_L^{cSTI+}.

Focussing first on the equation for the force of HIV infection on cSTI-uninfected individuals in the high-activity group, λ_H^{cSTI-} (Equation 4). This is equal to their partner change rate, c_H, the HIV transmission probability (without STI-cofactors), β_{HIV} and the probability that a partner is HIV-infected multiplied by the magnitude of the STI-cofactor effect if a HIV-infected partner is also co-infected with the cSTI.

This probability is calculated by the terms between the square brackets in Equation 4. It depends on whether the partner is selected from the high or low-activity group (but as we assume proportionate mixing this is just g_H or g_L), and whether the selected partner is HIV-infectious and cSTI-uninfected or HIV-infectious and cSTI-infected. These probabilities will vary between the two activity groups. If the selected partner is HIV-infectious and cSTI-infected then the HIV transmission probability is multiplied by the magnitude of the STI-cofactor effect, *x* .

$$\lambda_H^{cSTI-} = c_H \beta_{HIV} [g_H (q_H^{HIV+cSTI-} + q_H^{HIV+cSTI+} x) + g_L (q_L^{HIV+cSTI-} + q_L^{HIV+cSTI+} x)]$$

Equation 4

Where

$$q_H^{HIV+cSTI-} = \frac{I_H^{cSTI-}}{U_H}, \text{ and} \qquad q_H^{HIV+cSTI+} = \frac{I_H^{cSTI+}}{U_H}$$

$$q_L^{HIV+cSTI-} = \frac{I_L^{cSTI-}}{U_L}, \text{ and} \qquad q_L^{HIV+cSTI+} = \frac{I_L^{cSTI+}}{U_L}$$

The equation for the force of HIV infection on a cSTI-uninfected individuals in the low-activity group, λ_L^{cSTI-}, can be derived in the same way and is identical to Equation 4 except the partner change rate is c_L.

By slightly adapting Equation 4 we can write down the force of HIV infection on cSTI-infected individuals in the high-activity group, λ_H^{cSTI+} (Equation 5). We can use the same logic as above but need to account for the fact that the HIV-uninfected partner is already infected with the cSTI. λ_H^{cSTI+} is equal to the product of

a) the partner change rate, c_H,
b) the HIV transmission probability (with the STI-cofactor), β_{HIV} x , and
c) the probability that the selected partner is HIV infected multiplied by the magnitude of the STI-cofactor effect if the HIV-infected partner is also co-infected with the cSTI. Note that because the HIV-uninfected partner is already cSTI-infected, the HIV transmission probability will be $\beta_{HIV}\, x$ if their partner is HIV-infectious and STI-uninfected and $\beta_{HIV}\, x^2$ if the partner is HIV-infectious and STI-infected.

$$\lambda_H^{cSTI+} = c_H \beta_{HIV} x[g_H(q_H^{HIV+cSTI-} + q_H^{HIV+cSTI+} x) + g_L(q_L^{HIV+cSTI-} + q_L^{HIV+cSTI+} x)]$$

Equation 5

Similarly, the equation for the force of infection on a cSTI-infected individual in the low-activity group, λ_L^{cSTI+}, can be derived in the same way and is identical to Equation 5 except the partner change rate is c_L.

References

Abu-Raddad, L. J., Magaret, A. S., Celum, C, Wald, A, Longini, I. M. Jr., Self, S. G., and Corey, L.,Genital herpes has played a more important role than any other sexually transmitted infection in driving HIV prevalence in Africa,PLoS ONE,2008

Anderson, R. M., Gupta, S., and May, R. M.,Potential of community-wide chemotherapy or immunotherapy to control the spread of HIV-1,Nature,1991

Anderson, RM and May, RM,Ch 11: Social heterogeneity and sexually transmitted diseases,Infectious Diseases of Humans: Dynamics and Control,1991

Auvert, B., Kahn, J., Korenromp, E., Lloyd-Smith, J., Y, Mahabeer., Taljaard, D., Sitta, R., Hargrove, J., Williams, B. G., and Marseille, E.,Cost of the roll-out of male circumcision in sub-Saharan Africa,4th International AIDS Society Conference,2007

Auvert, B., Moore, M., Bertrand, W. E., Beauchet, A., Aegerter, P., Lusamba, D., Diong, K. T., and Klinowski, J.,Dynamics of HIV infection and AIDS in central African cities,Int J Epidemiol,1990

Auvert, B., Taljaard, D., Lagarde, E., Sobngwi-Tambekou, J., Sitta, R., and Puren, A.,Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: the ANRS 1265 Trial,PLoS Med,2005

Baggaley, R. F., Garnett, G. P., and Ferguson, N. M.,Modelling the Impact of Antiretroviral Use in Resource-Poor Settings,PLoS Med,2006

Baggaley, R. F., Griffin, J. T., Chapman, R., Hollingsworth, T. D., Nagot, N., Delany, S., Mayaud, P., de Wolf, F., Fraser, C., Ghani, A. C., and Weiss, H. A.,Estimating the public health impact of the effect of herpes simplex virus suppressive therapy on plasma HIV-1 viral load,AIDS,2009

Bailey, N.,Core-group dynamics and public health action,Modelling the AIDS epidemic: Planning, policy and prediction,1994

Bailey, R. C., Moses, S., Parker, C. B., Agot, K., Maclean, I., Krieger, J. N., Williams, C. F., Campbell, R. T., and Ndinya-Achola, J. O.,Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomised controlled trial,Lancet,2007

Blower, S, Gershengorn, H, and Grant, R. M.,A Tale of Two futures: HIV and Antiretroviral Therapy in San Francisco,Science,2000

Blower, S. and Ma, L.,Calculating the contribution of herpes simplex virus type 2 epidemics to increasing HIV incidence: treatment implications,Clin Infect Dis,2004

Boily, M-C and Masse, B.,Mathematical models of disease transmission: A precious tool of the study of sexually transmitted diseases,Canadian Journal of Public Health,1997

Boily, M. C., Baggaley, R. F., Wang, L., Masse, B., White, R. G., Hayes, R. J., and Alary, M.,Heterosexual risk of HIV-1 infection per sexual act: a systematic review and meta-analysis of observational studies,Lancet Infectious Diseases,2009

Boily, Marie-Claude, Lowndes, Catherine M., and Alary, Michel,Complementary hypothesis concerning the community sexually transmitted disease mass treatment puzzle in Rakai, Uganda,AIDS,2000

Boily, MC and Anderson, M,Human immunodeficiency virus transmission and the role of other sexually transmitted diseases: Measures of Association and Study Design,Sexually Transmitted Diseases,1996

Buve, A., Carael, M., Hayes, R. J., Auvert, B., Ferry, B., Robinson, N. J., Anagonou, S., Kanhonou, L., Laourou, M., Abega, S., Akam, E., Zekeng, L., Chege, J., Kahindo, M., Rutenberg, N., Kaona, F., Musonda, R., Sukwa, T., Morison, L., Weiss, H. A., and Laga, M.,The multicentre study on factors determining the differential spread of HIV in four African cities: summary and conclusions,AIDS,2001

CASCADE,Time from HIV-1 seroconversion to AIDS and death before widespread use of highly-active antiretroviral therapy: a collaborative re-analysis. Collaborative Group on AIDS Incubation and HIV Survival including the CASCADE EU Concerted Action. Concerted Action on SeroConversion to AIDS and Death in Europe,Lancet,2000

Celum, C, Wald, A., Hughes, J., Sanchez, J, Reid, S, Delany-Moretlwe, S, Cowan, F, Casapia, M, Ortiz, O, Fuchs, J, Buchbinder, S, Koblin, B., Zwerski, S, Rose, S, Wang, J., Corey, L, and the HPTN 039 Protocol Team,Effect of aciclovir on HIV-1 acquisition in herpes simplex virus 2 seropositive women and men who have sex with men: a randomised, double-blind, placebo-controlled trial,Lancet,2008

Coffee, M. P., Lurie, M., and Garnett, G. P.,Modelling the impact of migration on the HIV epidemic in South Africa,AIDS,2007
Coffin, J., Haase, A., Levy, J. A., Montagnier, L., Oroszlan, S., Teich, N., Temin, H., Toyoshima, K., Varmus, H., Vogt, P., and et al.,Human immunodeficiency viruses,Science,1986
Expert Group on Modelling the Impact and Cost of Male Circumcision for HIV Risk Reduction,Informing decision making on male circumcision for HIV prevention in high HIV prevalence settings: what mathematical modelling can contribute,PLoS Med,32767
Fleming, D. T. and Wasserheit, J. N.,From epidemiological synergy to public health policy and practice: the contribution of other sexually transmitted diseases to sexual transmission of HIV infection,Sex Transm Inf,1999
Freeman, E., Orroth, K., White, R. G., Glynn, J. R, Bakker, R., Boily, M-C, Habbema, D, Buve, A, and Hayes, R. J.,The proportion of new HIV infections attributable to HSV-2 increases over time: simulations of the changing role of sexually transmitted infections in sub-Saharan African HIV epidemics,STI,2007
Freeman, E. E., White, R. G., Bakker, R., Orroth, K. K., Weiss, H. A., Buve, A, Hayes, R. J., and Glynn, J. R.,Population-level effect of potential HSV2 prophylactic vaccines on HIV incidence in sub-Saharan Africa. http://dx.doi.org/10.1016/j.vaccine.2008.11.074,Vaccine,2009
Garnett, G P, Gregson, S, and Stanecki, K A,Criteria for detecting and understanding changes in the risk of HIV infection at a national level in generalised epidemics10.1136/sti.2005.016022,Sex Transm Infect,2006
Garnett, G. P. and Anderson, R. M.,Strategies for limiting the spread of HIV in developing countries: conclusions based on studies of the transmission dynamics of the virus,J Acquir Immune Defic Syndr Hum Retrovirol,1995
Garnett, G. P. and Anderson, R. M.,Antiviral therapy and the transmission dynamics of HIV-1,J.Antimicrob.Chemother.,1996
Granich, R. M., Gilks, C. F., Dye, C., De Cock, K. M., and Williams, B. G.,Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model,Lancet,2008
Gray, R., Azire, J., Serwadda, D., Kiwanuka, N., Kigozi, G., Kiddugavu, M., Nalugoda, F., Li, X., and Wawer, M.,Male circumcision and the risk of sexually transmitted infections and HIV in Rakai, Uganda,AIDS,2004
Gray, R. H., Kigozi, G., Serwadda, D., Makumbi, F., Watya, S., Nalugoda, F., Kiwanuka, N., Moulton, L. H., Chaudhary, M. A., Chen, M. Z., Sewankambo, N. K., Wabwire-Mangen, F., Bacon, M. C., Williams, C. F., Opendi, P., Reynolds, S. J., Laeyendecker, O., Quinn, T. C., and Wawer, M. J.,Male circumcision for HIV prevention in men in Rakai, Uganda: a randomised trial,Lancet,2007
Gray, Ronald H, Xianbin Li, Wawer, Maria J, Gange, Stephen J, Serwadda, David, Sewankambo, Nelson K, Moore, R., Wabwire-Mangen, Fred, Lutalo, Tom, Quinn, Thomas C, Kiwanuka, G., and For the Rakai Project Group,Stochastic simulation of the impact of antiretroviral therapy and HIV vaccines on HIV transmission; Rakai Uganda,AIDS,2003
Grosskurth, H., Mosha, F., Todd, J., Mwijarubi, E., Klokke, A., Senkoro, K., Mayaud, P., Changalucha, J., Nicoll, A., ka-Gina, G., and et al.,Impact of improved treatment of sexually transmitted diseases on HIV infection in rural Tanzania: randomised controlled trial,Lancet,1995
Gupta, S., Anderson, R., and May, R. M.,Mathematical Models and the Design of Public Health Policy: HIV and Antiviral Therapy,SIAM Review,1993
Habbema, J. D. F. and De Vlas, S. J.,Impact of improved treatment of sexually transmitted disease on HIV infection [letter; comment],Lancet,1995
Hallett, T. B., Singh, K. J., Smith, J. A., White, R. G., Abu-Raddad, L., and Garnett, G. P.,Understanding the Impact of Male Circumcision Interventions on the Spread of HIV in Southern Africa,PLoS ONE,2008
Hayes, R., Grosskurth, H., and ka-Gina, G.,Impact of improved treatment of sexually transmitted diseases on HIV infection in rural Tanzania: randomised controlled trial [Letter to Lancet],Lancet,1995
Hayes, R. J., Schulz, K. F., and Plummer, F. A.,The cofactor effect of genital ulcers on the per-exposure risk of HIV transmission in sub-Saharan Africa. [Review],J Trop Med Hyg,1995
Hyman, J. and Stanley, E.,Using mathematical models to understand the AIDS epidemic,Math Biosc,1988
Kamali, A, Quigley, M, Nakiyingi, J, Kinsman, J, Kengeya-Kayondo, J, Gopal, R, Ojwiya, A, Hughes, P, Carpenter, LM, and Whitworth, J,Syndromic management of sexually-transmitted infections and behaviour change interventions on transmission of HIV-1 in rural Uganda: a community randomised trial,Lancet,2003
Korenromp, E. L., De Vlas, S., Nagelkerke, N., and Habbema, J.,Estimating the magnitude of STD cofactor effects on HIV transmission - how well can it be done?,STD,2001
Korenromp, E. L., Van Vliet, C., Grosskurth, H., Gavyole, A., Van der Ploeg, C. P. B., Fransen, L., Hayes, R. J., and Habbema, J. D. F.,Model-based evaluation of single-round mass STD treatment for HIV control in a rural African population,AIDS,2000
Korenromp, E. L., White, R. G., Orroth, K. K., Bakker, R., Kamali, A., Serwadda, D., Gray, R. H., Grosskurth, H., Habbema, J. D., and Hayes, R. J.,Determinants of the Impact of Sexually Transmitted Infection Treatment on Prevention of HIV Infection: A Synthesis of Evidence from the Mwanza, Rakai, and Masaka Intervention Trials,J Infect Dis,2005
Korenromp, EL, Bakker, R, De Vlas, SJ, Gray, RH, Wawer, MJ, Serwadda, D, and Habbema, JDF,The effect of HIV, behaviour change, and STD syndromic management on STD epidemiology in sub-Saharan Africa: simulations of Uganda,STI,2002a
Korenromp, EL, Bakker, R, De Vlas, SJ, Gray, RH, Wawer, MJ, Serwadda, D, Sewankambo, NK., and Habbema, JDF,HIV dynamics and behaviour change as determinants of the impact of sexually transmitted disease treatment on HIV transmission in the context of the Rakai trial,AIDS,2002b
Korenromp, EL, Bakker, R, De Vlas, SJ, Robinson, NJ, Hayes, R, and Habbema, JDF,Can behaviour change explain increases in the proportion of genital ulcers attributable to herpes in sub-Saharan Africa? A simulation modelling study,Sex Transm Dis,2002c
Laga, M., Nzila, N., and Goeman, J.,The interrelationship of sexually transmitted diseases and HIV infection: implications for the control of both epidemics in Africa. [Review],AIDS,1991
Morgan, D., Mahe, B., Okongo, M. J., Lubega, R., and Whitworth, J. A.,HIV-1 infection in rural Africa: is there a difference in median time to AIDS and survival compared with that in industrialized countries?,AIDS,2002
Morris, M. and Dean, L.,Effect of sexual behavior change on long-term human immunodeficiency virus prevalence among homosexual men,Am J Epidemiol,1994
Moses, S., Bradley, J. E., Nagelkerke, N. J., Ronald, A. R., Ndinya-Achola, J. O., and Plummer, F. A.,Geographical patterns of male circumcision practices in Africa: association with HIV seroprevalence,Int J Epidemiol,1990
Nagelkerke, N. J., Moses, S., De Vlas, S. J., and Bailey, R. C.,Modelling the public health impact of male circumcision for HIV prevention in high prevalence areas in Africa,BMC Infect Dis,2007
Orroth, K. K., Freeman, E., Bakker, R., Buve, A, Glynn, J., Boily, M-C, White, R. G., Habbema, J. D. F., and Hayes, R. J.,Understanding differences across the contrasting epidemics in East and West Africa: results from a simulation model of the Four Cities Study,STI,2007
Orroth, KK, Korenromp, EL, White, RG, Gavyole, A, Gray, R, Muhangi, L, Sewankambo, NK, Quigley, M, Wawer, MJ, Whitworth, JAG, Grosskurth, H, Habbema, JDF, and Hayes, RJ,Higher risk behaviour and rates of sexually transmitted diseases in Mwanza compared to Uganda may help explain HIV prevention trial outcomes,AIDS,2003
Over, M and Piot, P,Human immunodeficiency virus infection and other sexually transmitted diseases in developing countries: Public health importance and priorities for resource allocation,The Journal of Infectious Diseases,1996
Piot, P., Quinn, T. C., Taelman, H., Feinsod, F. M., Minlangu, K. B., Wobin, O., Mbendi, N., Mazebo, P., Ndangi, K., Stevens, W., and et al.,Acquired immunodeficiency syndrome in a heterosexual population in Zaire,Lancet,1984
Powers, K. A., Poole, C., Pettifor, A. E., and Cohen, M. S.,Rethinking the heterosexual infectivity of HIV-1: a systematic review and meta-analysis,Lancet Infect Dis,2008
Rehle, T. M., Saidel, T. J., Hassig, S. E., Bouey, P. D., Gaillard, E. M., and Sokal, D. C.,AVERT: a user-friendly model to estimate the impact of HIV/sexually transmitted disease prevention interventions on HIV transmission,AIDS,1998
Robinson, N. J., Mulder, D. W., Auvert, B., and Hayes, R. J.,Proportion of HIV infections attributable to other sexually transmitted diseases in a rural Ugandan population: simulation model estimates,International Journal Of Epidemiology,1997
Rottingen, J. A., Cameron, D. W., and Garnett, G. P.,A systematic review of the epidemiologic interactions between classic sexually transmitted diseases and HIV: how much really is known?,Sex Transm Dis,2001
Rowley, J. T. and Anderson, R. M.,Modeling the impact and cost-effectiveness of HIV prevention efforts,AIDS,1994

Salomon, Joshua A., Hogan, Daniel R., Stover, John, Stanecki, Karen A., Walker, Neff, Ghys, Peter D., and Schwartlander, Bernhard,Integrating HIV Prevention and Treatment: From Slogans to Impact,PLoS Medicine,2005

Stover, J and O'Way, P,Inpact of interventions on reducing the spread of HIV in developing countries: computer simulation applications,African journal of medical practice,1995

Stover, J., Bertozzi, S., Gutierrez, J. P., Walker, N., Stanecki, K. A., Greener, R., Gouws, E., Hankins, C., Garnett, G. P., Salomon, J. A., Boerma, J. T., De, Lay P., and Ghys, P. D.,The global impact of scaling up HIV/AIDS prevention programs in low- and middle-income countries,Science,10-3-2006

Todd, J., Glynn, J. R., Marston, M., Lutalo, T., Biraro, S., Mwita, W., Suriyanon, V., Rangsin, R., Nelson, K. E., Sonnenberg, P., Fitzgerald, D., Karita, E., and Zaba, B.,Time from HIV seroconversion to death: a collaborative analysis of eight studies in six low and middle-income countries before highly active antiretroviral therapy,AIDS,2007

UNAIDS,2008 Report on the global HIV/AIDS epidemic,2008

UNAIDS,AIDS epidemic update 2009,2009

UNAIDS/ WHO,UNAIDS/ WHO Epidemiological Fact Sheet, 2008 Update, South Africa,2008a

UNAIDS/ WHO,UNAIDS/ WHO Epidemiological Fact Sheet, 2008 Update, Uganda,2008b

UNAIDS/ WHO,UNAIDS/ WHO Epidemiological Fact Sheet, 2008 Update, Zimbabwe,2008c

Van der Ploeg, Catharina P B, Van Vliet, Carina, De Vlas, Sake J., Ndinya-Achola Jeckoniah O., Fransen Lieve, Van Oortmarssen Gerrit J, and Habbema, J Dik F.,STDSIM: A microsimulation model for decision support in STD control,Interfaces,1998

Van Druten, J. A., Reintjes, A. G., Jager, J. C., Heisterkamp, S. H., Poos, M. J., Coutinho, R. A., Dijkgraaf, M. G., and Ruitenberg, E. J.,HIV infection dynamics and intervention experiments in linked risk groups,Stat Med,1990

Wasserheit, J. N.,Epidemiological synergy. Interrelationships between human immunodeficiency virus infection and other sexually transmitted diseases. [Review],Sex Transm Dis,1992

Watson-Jones, D., Weiss, H. A., Rusizoka, M., Changalucha, J., Baisley, K., Mugeye, K., Tanton, C., Ross, D., Everett, D., Clayton, T., Balira, R., Knight, L., Hambleton, I., Le Goff, J., Belec, L., and Hayes, R.,Effect of Herpes Simplex Suppression on Incidence of HIV among Women in Tanzania,N Engl J Med,2008

Wawer, M. J., Gray, R. H., Sewankambo, N. K., Serwadda, D., Li, X., Laeyendecker, O., Kiwanuka, N., Kigozi, G., Kiddugavu, M., Lutalo, T., Nalugoda, F., Wabwire-Mangen, F., Meehan, M. P., and Quinn, T. C.,Rates of HIV-1 Transmission per Coital Act, by Stage of HIV-1 Infection, in Rakai, Uganda,J Infect Dis,2005

Wawer, M. J., Serwadda, D., Gray, R. H., Sewankambo, N. K., Li, C., Nalugoda, F., Lutalo, T., and Konde Lule, J. K.,Trends in HIV-1 prevalence may not reflect trends in incidence in mature epidemics: data from the Rakai population-based cohort, Uganda,AIDS,1997

Wawer, M. J., Sewankambo, N. K., Serwadda, D., Quinn, T. C., Paxton, L. A., Kiwanuka, N., Wabwire-Mangen, F., Li, C., Lutalo, Th., Nalugoda, F., Gaydos, C. A., Moulton, L. H., Meehan, M. O., Ahmed, S., the Rakai Project Group, and Gray, R. H.,Control of Sexually Transmitted diseases for AIDS prevention in Uganda: a randomised community trial,Lancet,1999

Weiss, HA, Quigley, MA, and HAYES, RJ,Male circumcision and risk of HIV infection in sub-Saharan Africa: a systematic review and meta-analysis,AIDS,2000

White, R. G., Freeman, E. E., Orroth, K. K., Bakker, R., Weiss, H. A., O'Farrell, N., Buve, A, Hayes, R. J., and Glynn, J. R.,Population-level effect of HSV-2 therapy on the incidence of HIV in sub-Saharan Africa,Sex Trans Inf,2008

White, R. G., Glynn, J. R., Orroth, K. K., Freeman, E. E., Bakker, R., Weiss, H. A., Kumaranayake, L, Habbema, J. D. F., Buve, A, and Hayes, R. J.,Male circumcision for HIV prevention in sub-Saharan Africa: who, what and when?,AIDS,2008

White, R. G., Orroth, K. K., Glynn, J. R., Freeman, E. E., Bakker, R., Habbema, J. Dik. F., Terris-Prestholt , F., Kumaranayake, L., Buve, A., and Hayes, R. J.,Treating curable Sexually Transmitted Infections to Prevent HIV in Africa: Still an Effective Control Strategy?,J Acquir Immune Defic Syndr,2008

White, R. G., Orroth, K. K., Korenromp, E. L., Bakker, R., Wambura, M., Sewankambo, N. K., Gray, R. H., Kamali, A., Whitworth, J. A., Grosskurth, H., Habbema, J. D., and Hayes, R. J.,Can Population Differences Explain the Contrasting Results of the Mwanza, Rakai, and Masaka HIV/Sexually Transmitted Disease Intervention Trials?: A Modeling Study,J Acquir Immune Defic Syndr,2004

WHO,Global Prevalence And Incidence Of Selected Curable Sexually Transmitted Infections: Overview And Estimates,2001

Williams, B. G., Lloyd-Smith, J. O., Gouws, E., Hankins, C., Getz, W. M., Hargrove, J., De Zoysa, I., Dye, C., and Auvert, B.,The Potential Impact of Male Circumcision on HIV in Sub-Saharan Africa,PLoS Med,2006

Introduction to Infectious Disease Modelling and its Applications – 2018

Session 32: Real-time modelling: some methods

Lecture

Objectives

By the end of this lecture, you should:

- Be able to reconstruct infection trees from case data;
- Understand how the growth rate of an outbreak relates to the reproductive number;
- Understand the challenges of model fitting for real-time analysis;

Introduction

One of the complications associated with modelling emerging diseases is that parameters required for the models are typically unknown and have to be estimated at the same time as the outbreak of the infection unfolds. Analysis of the data for an emerging infection is usually conducted in two stages: a first stage, during which it is possible to both count and confirm the cases, using individual-level data; a second stage, during which only aggregated reports are available.

Different methods have been designed, tailored for these two phases. In this lecture, we will discuss:

1. Reconstruction of infection trees (Phase 1)
2. Estimation from the growth rate of an epidemic (Phase 2)
3. Fitting transmission models (Phase 2)

It is important to note that the analyses conducted during these phases are usually not independent. In particular, the early analysis - done in the early stage of phase 1 – is particularly important for estimating key epidemiological quantities which are used during the second stage.

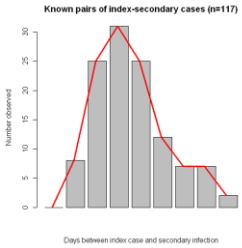
Reconstruction of infection trees

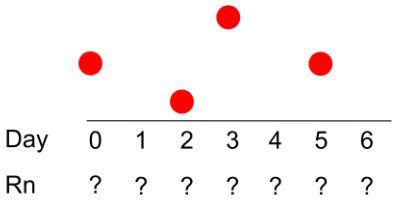
This method, described by Wallinga & Teunis 2004[1], is based on two strong assumptions. First it is assumed that all cases are known and then that the serial interval is known. This might have been true for SARS, for which Wallinga and Teunis designed the method, but is complicated for infections such as influenza, for which some infections may be asymptomatic.

The aim of the Wallinga and Teunis algorithm is to reconstruct infection trees and give their likelihood. It relies on knowledge of the distribution of the serial interval (time interval between onset of symptoms in a primary and secondary case). Estimates of the serial interval would usually come from known pairs of transmission-linked infections, gathered during the early stage of the epidemiological analysis.

For a given outbreak, a large number of infection trees linking cases can be calculated. We want to find the probability of each possible infection tree to estimate the instantaneous reproduction number R_n.

We are going to use a working example to explain the Wallinga and Teunis algorithm. Let us consider the following outbreak of four cases, with a known serial interval (see graph and table).





Time in days	0	1	2	3	4	5	6	7	8
Probability w(t)	0	0.07	0.21	0.26	0.21	0.10	0.06	0.06	0.017

From this, we can write down a "matrix of time weights" taken from the distribution of serial interval:

Source of infection

Secondary case	1	2	3	4
1	0	0	0	0
2	0.21	0	0	0
3	0.26	0.07	0	0
4	0.10	0.26	0.21	0

Time weights

Each cell in the matrix reflects the likelihood that a given case was the source of infection for another case.

For example on row 3 and column 1, the matrix is saying that the likelihood that case 1 infected case 3 is 0.26. It is thus much more likely that case 1 rather than case 2 infected case 3 (likelihood of 0.26 vs 0.07). To transform the likelihood into the probability that a given case had been infected by another specific case, the rows need to be renormalized to sum up to 1. After doing this for the above matrix, we then obtain the following matrix:

Source of infection

Secondary case	1	2	3	4
1	0	0	0	0
2	1.0	0	0	0
3	0.79	0.21	0	0
4	0.17	0.46	0.37	0

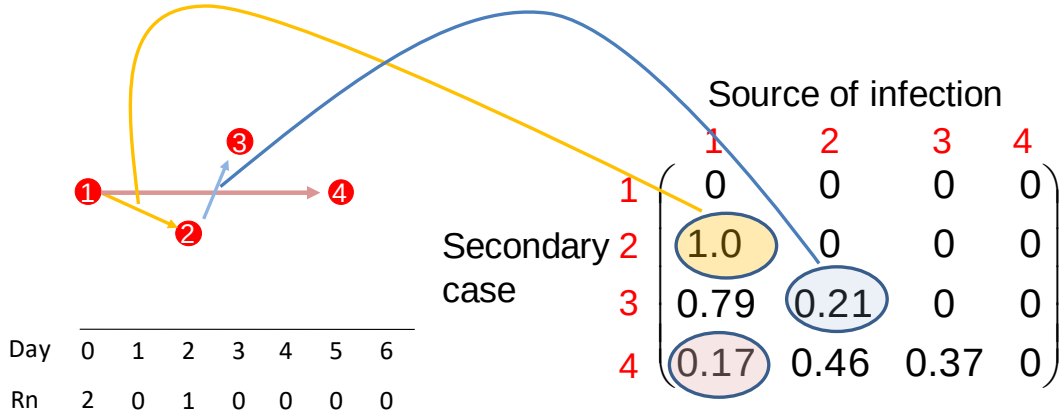
By summing up the values in each column, we can compute the expected number of transmissions resulting from case *j*, which we refer to using the symbol R_j. For example, the expected number of transmissions from case 1 (R_1) equals 0+1+0.79+0.17=1.96. Similarly, the expected numbers of transmissions from cases 2, 3 and 4 are given by R_2=0.67, R_3=0.37 and R_4=0 respectively.

It is possible also to work out what is the probability associated with each of the possible trees by multiplying the probability of each corresponding infection events in the tree. To calculate in the example shown below, we would need to multiply the probabilities of each of the following 3 events having occurred:

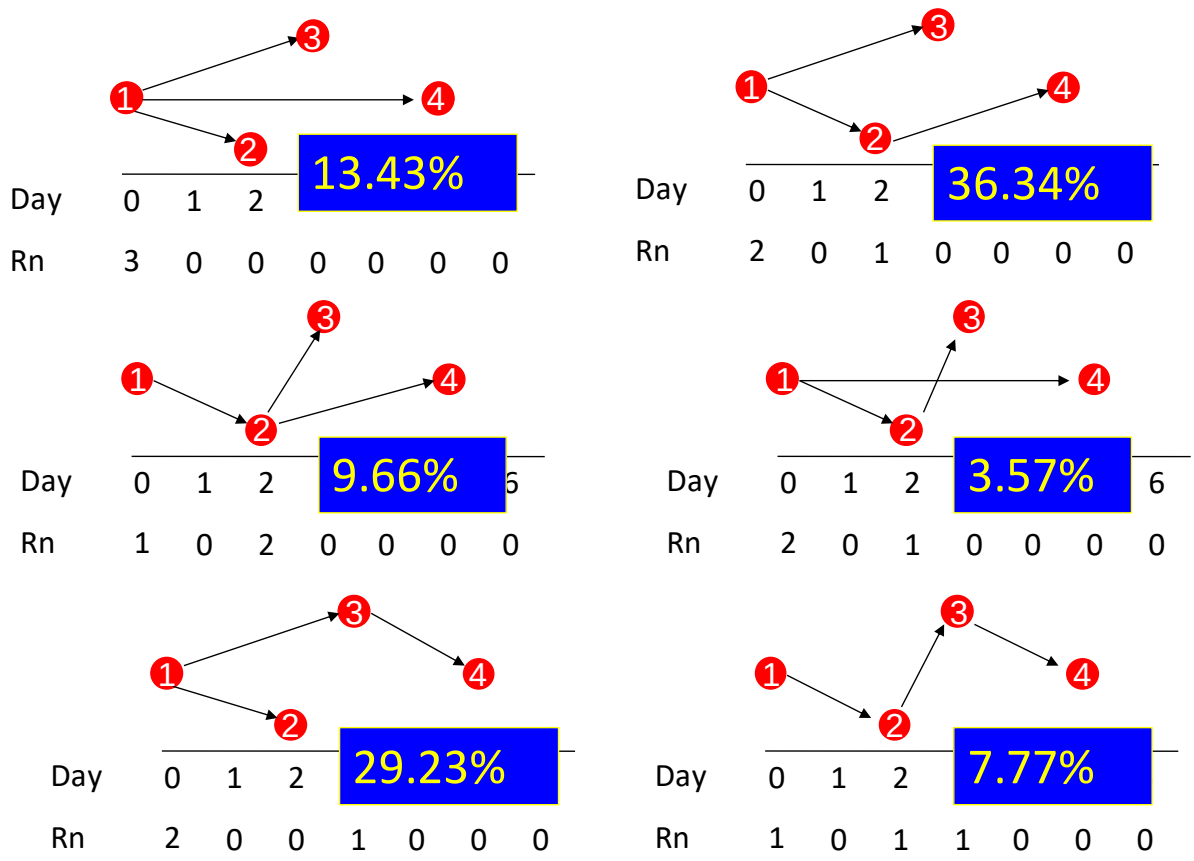
a) case 2 was infected by case 1, which occurs with a probability of 1
b) case 4 was infected by case 1, which occurs with a probability of 0.17
c) case 3 was infected by case 2, which occurs with a probability of 0.21.

The overall probability that this transmission tree occurred is then given by:

P(tree) = 1*0.21 * 0.17 = 0.0357



This can be done for each of the possible trees:



This method allows the value of the individual reproduction ratio R_n to be tracked in real-time, with infection trees being reconstructed and their probabilities assessed.

As the method operates at the individual level, it is possible to stratify the data to gain insight into the reproduction number associated with different subgroups of the population (e.g. adults/children, treated/non treated etc.). It should be stressed, however, that the method assumes that all cases are reported. Possible delays might bias the algorithm. Finally, the algorithm can be extended to include more information (by basing the likelihood of individual transmission on other data).

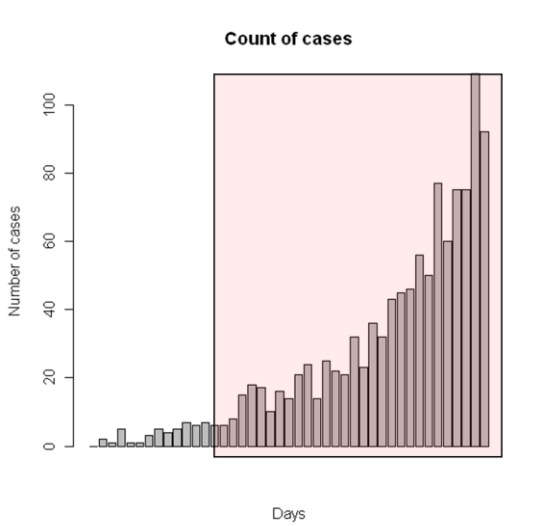
Estimation from growth rate

For this method, we assume that by now, transmission of the infection is widespread and that the growth in the number of cases is exponential, with plenty of susceptibles available. Assuming that the proportion of cases that are not reported is constant over time, the growth rate will be independent of the extent of underreporting. Indeed, the observed number of cases at a given time *t* (*N(t)*) can be written as *N(t)= N(0)*exp(r*t)* with *r* being easy to calculate as the slope of *log(N(t)). (*Note that the "r" used in this equation is equivalent to the Greek capital letter Λ that was used in the session on the natural dynamics in block 1 and should not be confused with the rate at which individuals recover from being infectious in the models described elsewhere in the course). Like R, the rate *r* is also a threshold value; it needs to be positive for the epidemic to be sustained.

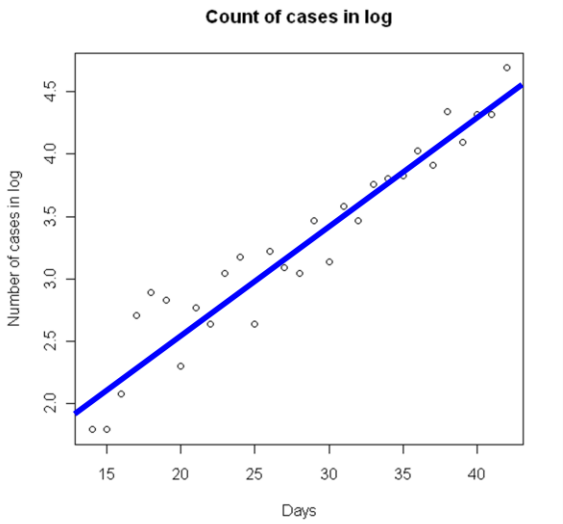
The main objective of the method is to link the growth rate r with the reproduction number *R* during the period considered. Note that *R* is an average value here (while R_n changes with time, and R_0 assumes that the population was fully susceptible at the start of the outbreak).

Let us consider the following epidemic:

First 6 weeks of the epidemic



Last 4 weeks, exponential growth



where the right panel presents the log of the number of cases taken in the weeks for which exponential growth is considered. The slope of the best fitting line is measured as

$$r \sim 0.0865 \text{ /day}$$

which can also be expressed as time needed for the epidemic to double in size

$$t(\text{doubling}) = \ln(2)/r \sim 8 \text{ days}.$$

The method for relating the value for the growth rate with the value for the reproduction number is not obvious and different equations can be used:

- $R = 1 + rT_c$ is the formula used usually in infectious diseases epidemiology.
- $R = exp(rT_c)$ is the formula used traditionally in ecological modelling.

where T_c is the mean of the serial interval. These equations and others are discussed in the recommended course text in chapter 4[4]. •Assuming an average serial interval (T_c) of 4 days and the estimated growth rate of 0.0865 /day, we obtain values for the reproduction number of 1.35 and 1.42 for these two methods respectively.

The growth rate will depend on how many secondary infections a typical infectious individual generates and will also be influenced by how they are spread temporally. The mean is important, as is the way that the serial interval is distributed. Wallinga and Lipsitch demonstrated in [2] that we have the following relationship between *r* and *R:*

$$\underbrace{M(z) = \int_{a=0}^{\infty} \mathrm{e}^{za} g(a) \mathrm{d}a.}_{\text{Moment generating function}} \qquad R = \frac{1}{M(-r)}$$

For common distributions about the serial interval, this equation translates into the following equations for R:

- Exponential distribution: $R = 1 + r*T_c$
- Delta[1] distribution: $R = \exp(r*T_c)$
- Empirical distributions:

$$R = \frac{r}{\sum_{i=1}^{n} y_i(\mathrm{e}^{-ra_{i-1}} - \mathrm{e}^{-ra_i})/(a_i - a_{i-1})}$$

where a_i are the bounds of the categories of the histogram describing the distribution of the serial interval and y_i are the relative frequencies of different durations for the serial interval. Using the distribution of the serial interval shown earlier in this session, we find that *R*=1.41.

Fitting transmission models

With this method, an assumed model (or a collection of models) is fitted it to the available data. This is usually done by minimizing or maximizing a criteria (e.g. least squares, maximum likelihood, but also Bayesian where you keep the information on the distributions). The fitting process can be informed by available information ('priors' from early analysis). The goal of fitting transmission models is twofold: 1) to refine the parameter estimates obtained from the early analysis 2) to facilitate the development of a predictive tool which can be used for forecasting.

The concepts will be developed in the accompanying practical. These methods allow a lot of the complexity of 'real life' to be included. For example for one real-time economical model of the transmission dynamics of swine flu [3], numerous parameters were fitted in a real-life complex model: the incubation period, the infectious period, the initial reproduction number R (not R_0 because of pre-existing immunity), the age-susceptibility profile, the date of school closure, and the contact structure during term and holiday periods. Usually the parameters are fitted using ***priors*** obtained from the early phase analysis.

As discussed in the previous lecture, to be able to fit a transmission model, the number of cases picked up by surveillance needs to be scaled up. This is difficult before the peak (see practical). Models fitted in real time can be used to forecast the future course of the epidemic. They have to be adapted to embed the complexity of real life (social contacts, changing behaviours, prior immunity…)

Conclusions

The main problem with developing real time models is that the situation, especially regarding delays and under-reporting, is not well known. Sero-epidemiology may be able to help with this issue in future, potentially allowing the 'tracking down' of the antibodies in the population in real time which would allow the issue relating to "scale" to be addressed. Models remain an essential tool for informing policy makers and assessing the uncertainty around the possible decisions.

[1] delta distribution means that all cases happen after the same duration

References

1. Wallinga J, Teunis P (2004) Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. American journal of epidemiology 160: 509–516. Available: http://www.ncbi.nlm.nih.gov/pubmed/15353409.

2. Wallinga J, Lipsitch M (2007) How generation intervals shape the relationship between growth rates and reproductive numbers. Proceedings Biological sciences / The Royal Society 274: 599–604. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1766383&tool=pmcentrez&rendertype=abstract. Accessed 1 August 2011.

3. Baguelin M, Hoek AJV, Jit M, Flasche S, White PJ, et al. (2010) Vaccination against pandemic influenza A/H1N1v in England: a real-time economic evaluation. Vaccine 28: 2370–2384. Available: http://www.ncbi.nlm.nih.gov/pubmed/20096762. Accessed 5 July 2011.

4. Vynnycky E, White RG. An introduction to infectious disease modelling. Oxford University Press. Oxford 2010

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.

Introduction to Infectious Disease Modelling and its Applications – 2018

Session 34: Groupwork

In these sessions, you are asked to work in your allocated groups to address the problem that you have been assigned.

Introduction to Infectious Disease Modelling and its Applications – 2018

Sessions 35-6: Groupwork

In this session, you are asked, as a group, to present the poster relating to the problem that you have been assigned.