## Population heterogeneity, structure, and mixing

**Jamie Lloyd-Smith**

Center for Infectious Disease Dynamics
Pennsylvania State University

## Outline

Introduction: Heterogeneity and population structure

Models for population structure

Structure example: rabies in space

Models for heterogeneity

- individual heterogeneity and superspreaders
- group-level heterogeneity

Population structure and mixing mechanisms

Pair formation and STD transmission

## Heterogeneity and structure – what's the difference?

Tough to define, but *roughly…*

Heterogeneity describes differences among individuals or groups in a population.

Population structure describes deviations from random mixing in a population, due to spatial or social factors.

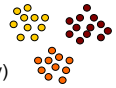
The language gets confusing:

- models that include heterogeneity in host age are called "age-structured".
- models that include spatial structure where model parameters differ through space are called "spatially heterogeneous".

## Modelling heterogeneity

*Group-level heterogeneity and multi-group models*

Break population into sub-groups, each of which is homogeneous.
(often assume that all groups mix randomly)



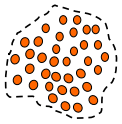
However, epidemiological traits of each host individual are due to a complex blend of host, pathogen, and environmental factors, and often can't be neatly divided into groups (or predicted in advance).

*Individual-level heterogeneity*

Allow continuous variation among individuals.

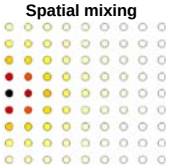
## Models for population structure

**Random mixing**

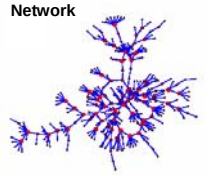


**Multi-group**

**Spatial mixing**



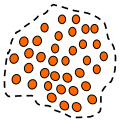
**Network**



**Individual-based model**

## Models for population structure

**Random mixing or mean-field**

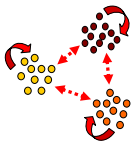


- Every individual in population has equal probability of contacting any other individual.
- Mathematically simple – "mass action" formulations borrowed from chemistry – but often biologically unrealistic.
- Sometimes basic $\beta SI$ form is modified to power law $\beta S^a I^b$ as a phenomenological representation of non-random mixing.

## Models for population structure

**Multi-group**

**or metapopulation**

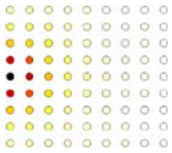


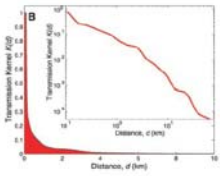
- Divides population into multiple discrete groupings, based on spatial or social differences.
- To model transmission, need contact matrix or Who Acquires Infection From Whom (WAIFW) matrix:

  $\beta_{ij}$ = transmission rate from infectious individual in group *i* to susceptible in group *j*
- Or use only within-group transmission (so $\beta_{ij}$ =0 when $i \neq j$), and explicitly model movement among groups.

## Models for population structure

**Spatial mixing**

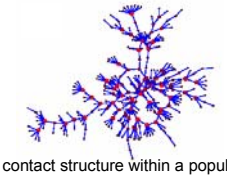




- Used when individuals are distributed (roughly) evenly in space.
- Can model many ways:
  - continuous space models (e.g. reaction-diffusion or contact kernel)
  - individuals as points on a lattice
  - patch models (metapopulation with spatial mixing)
- Used to study travelling waves, spatial control programs, influences of restricted mixing on disease invasion and persistence

## Models for population structure

**Social network**



- Precise representation of contact structure within a population
- "Nodes" are individuals and "edges" are contacts
- Important decisions: Binary vs weighted? Undirected vs directed? Static vs dynamic?
- Basic network statistics include degree distribution (number of edges per node) and clustering coefficient (How many of my friends are friends with each other?)
- Powerful tools of discrete mathematics can be applied.

## Models for population structure

**Individual-based model (IBM)**

**or microsimulation model**

- The most flexible framework.
- Every individual in the model carries its own attributes (age, sex, location, contact behaviour, etc etc)
- Can represent arbitrarily complex systems ( = realistic?) and ask detailed questions, but difficult to estimate parameters and to analyze model output; also difficult for others to replicate the model.
- STDSIM is a famous example, used to study transmission and control of sexually transmitted diseases including HIV in East Africa.

## Rabies in space

Rabies is an acute viral disease of mammals, that causes cerebral dysfunction, anxiety, confusion, agitation, progressing to delirium, abnormal behavior, hallucinations, and insomnia.

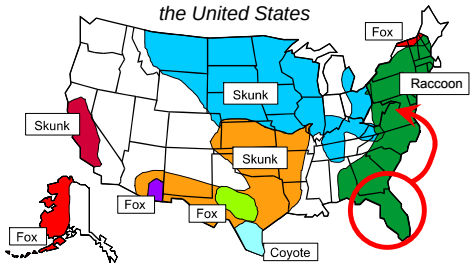
Transmitted by infected saliva, most commonly through biting.

Latent period = 3 – 12 weeks (in raccoons)
Infectious period = 1 week (ends in death)

Pre-exposure vaccination offers effective protection.

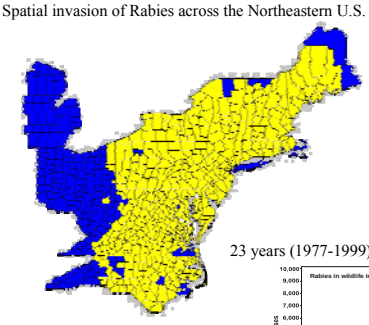
Post-exposure vaccination possible during latent period.

*Major Terrestrial Reservoirs of Rabies in the United States*

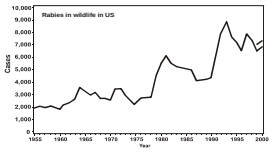



- Until mid 1970s, raccoon rabies was restricted to FL and GA.
- Then rabid raccoons were translocated to the WV-VA border, and a major epidemic began in the NE states.

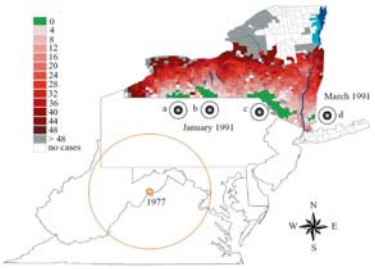
Spatial invasion of Rabies across the Northeastern U.S.

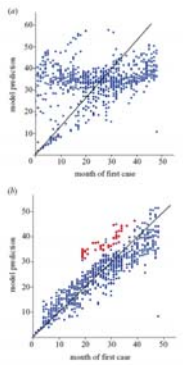


23 years (1977-1999)




## Models of the spatial spread of rabies



Simple patch model (+ small long-range transmission term) was able to fit data well.

Smith et al (2002) PNAS 99: 3668-3672

Russell et al (2004) Proc Roy Soc B 271: 21-25.

## Individual heterogeneity

**Macroparasitic diseases:**

worm burdens in individuals are overdispersed, and well-described by a negative binomial distribution.

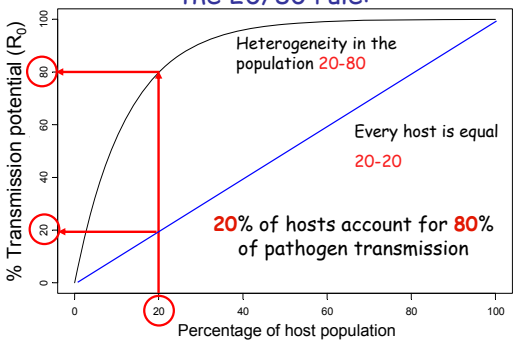
**STDs and vector-borne diseases:**

Woolhouse et al (PNAS, 1998) analyzed contact rate data and proposed a general **20/80 rule:**

20% of hosts are responsible for 80% of transmission

But how to approach **other directly-transmitted diseases**, for which contacts are hard to define?

## The vital few and insignificant many – the 20/80 rule:




Slide borrowed from Sarah Perkins

## A model for individual heterogeneity

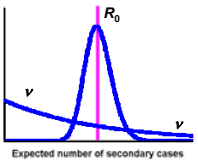
**Basic reproductive number, $R_0$**

Expected number of cases caused by a ***typical*** infectious individual in a susceptible population.

**Individual reproductive number, $\nu$**

Expected number of cases caused by a ***particular*** infectious individual in a susceptible population.

$\nu$ varies continuously among individuals, with population mean $R_0$.



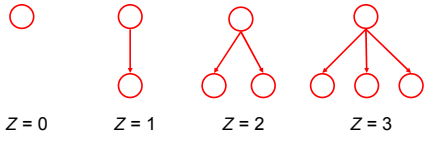

**Individual reproductive number, $\nu$**

***Expected number*** of cases caused by a particular infectious individual in a susceptible population.

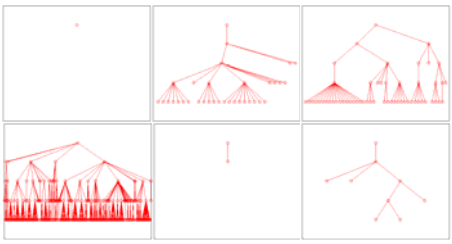
$Z$ = ***actual number*** of cases caused by a particular infectious individual.

Stochasticity in transmission → $Z \sim \text{Poisson}(\nu)$

The **offspring distribution** defines $\Pr(Z=j)$ for all $j$.



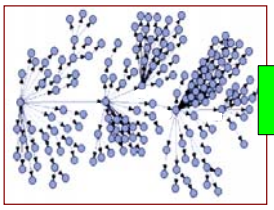
**Branching process**: a stochastic model for disease invasion into a large population.



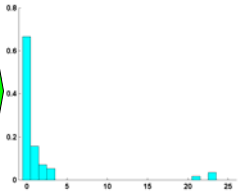
For any offspring distribution, it tells you:

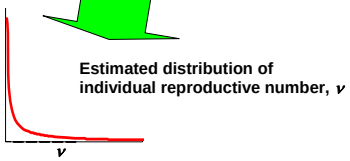
- Pr(extinction)
- Expected time of extinction and number of cases
- Growth rate of major outbreak

**Contact tracing for SARS**

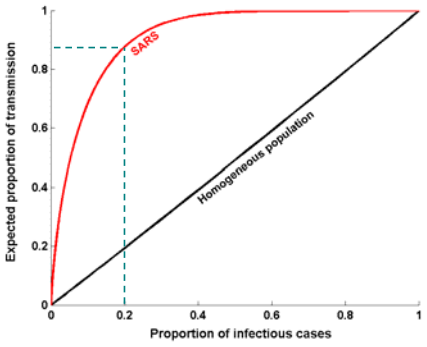


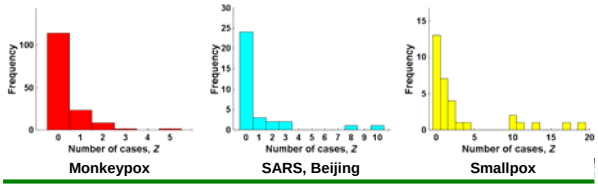
**Observed offspring distribution**



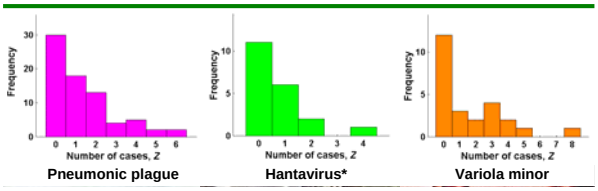



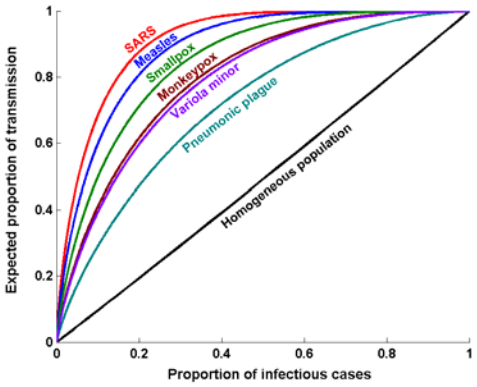

**Singapore SARS outbreak, 2003**



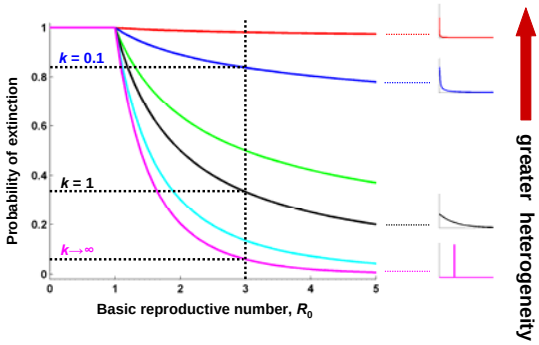




What about other emerging diseases?




**Dynamic effects: stochastic extinction of disease**




Read more about individual heterogeneity and superspreading in Lloyd-Smith et al (2005) Nature 438: 355-359.

## Transmission: mechanisms matter

Transmission dynamics are the core of epidemic models

→ Take time to think about the mechanisms underlying transmission, and to find the best tradeoff between model simplicity and biological realism.

e.g. Between-group transmission in metapopulations

Frequency-dependent transmission vs pair-formation models

## The **R**-matrix or next-generation matrix

Generalized $R_0$ for a multi-group population

$R_{ij}$ = E(# cases caused in group *j*|infected in group *i*)

Usual approach considers group membership as static.

$D_i$ = expected infectious period, spent entirely in group *i*

$\beta_{ij}$ = transmission rate from group *i* to group *j*

The expected number of cases in group *j* caused by an individual infected in group *i* is then:

$$R_{ij} = D_i\beta_{ij}$$

But what if the host moves and transmission is strictly local?

$D_{ij}$ = expected time spent in group *j* by individual infected in group *i*, while still infectious

$\beta_j$ = transmission rate within group *j*

Now $R_{ij} = D_{ij}\beta_j$

## Analytic approach to **R**

If movement rules are Markovian, so

$p_{ij}$ = Pr(move from group *i* to group *j*):

$m_j$ = Pr(recover or die while in group *j*)

The process can be described by an absorbing Markov chain, with overall transition matrix:

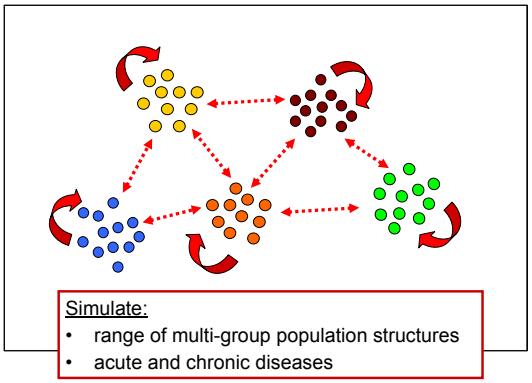
$$\begin{bmatrix} \mathbf{P}_{n\times n} & \mathbf{m} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}$$

The expected residence times $D_{ij}$ are then given by the fundamental matrix:

$$\mathbf{D} = (\mathbf{I} - \mathbf{P})^{-1}$$

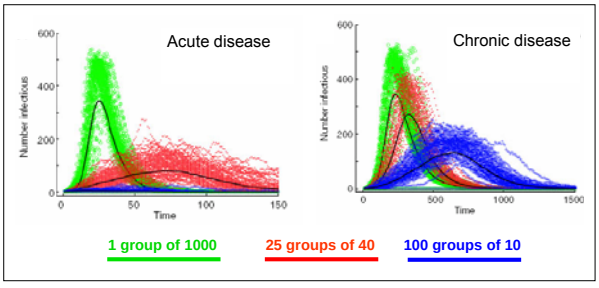
➡ R-matrix: $R_{ij} = D_{ij}\beta_j$

## Transmission in a metapopulation

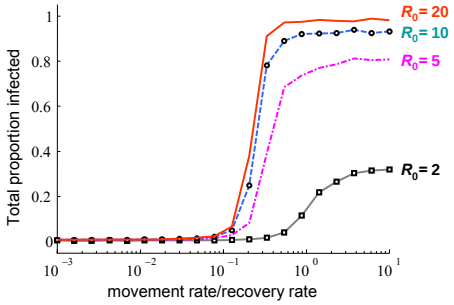


Simulate:

- range of multi-group population structures
- acute and chronic diseases

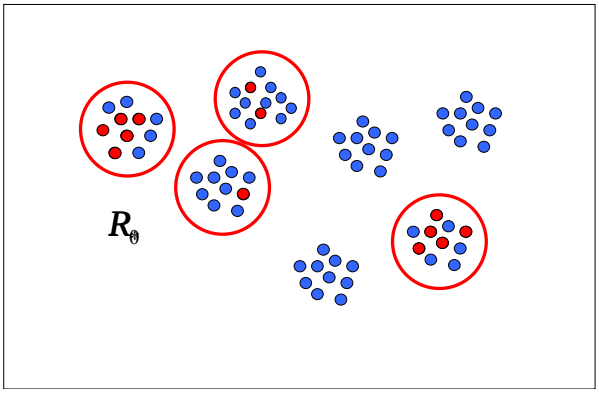



Acute and chronic diseases with **same $R_0$** behave very differently when invading a metapopulation.




**$R_0$ does not predict invasion for this system!**

### Units of analysis: $R_0$ versus $R_*$

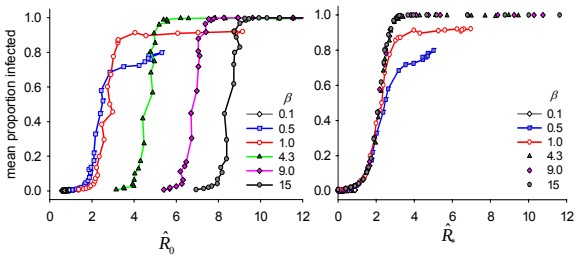



### Units of analysis: $R_0$ versus $R_*$

- $R_*$ = the expected number of groups infected by the first infected group (a group-level $R_0$).
  (Ball et al. 1997 Annals of Appl. Prob.)
- Analytical expressions for $R_0$ or $R_*$ are hard to find for systems with mechanistic movement, finite group sizes, and finite numbers of groups.
- Use "empirical" values: mean values from simulations where we track who infects whom.

$\hat{R}_0$ $\hat{R}_*$

### Predictors of disease invasion




$R_*$ is a much better predictor of disease invasion in a structured population than $R_0$

### Approaching $R_*$

$\mu/\gamma$ = expected number of movements between groups by an individual during its infectious period

$p_I$ = expected proportion of initial group infected following the initial outbreak. If $R_0$ is large, then $p_I$ ~1.

$p_I n\mu/\gamma$ = expected number of infectious individuals that will disperse from the initial group

$R_0 \approx \beta/\gamma$

So: for a pandemic, we require $R_0$ >1 and $p_I\, n\mu/\gamma$ >1.
crudely, $R_*$ will increase with $p_I\, n\, \mu/\gamma$ and with $\beta/\gamma$.

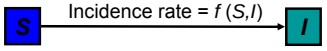
A proper mathematical treatment of this problem is needed!

### Summary on mechanisms in multi-group models

- Need to consider timescales of relevant processes: mixing, recovery, transmission, (susc. replenishment)
- In some limits, simpler models do OK.
- In general, and especially when different processes occur on **similar timescales**, **mechanistic models** are needed to capture dynamics.
- Appropriate "units" aid prediction.

Read more about disease invasion in structured populations in
Cross et al (2005) Ecol Lett 8: 587-595
Cross et al (2007) JRS Interface 4:315-324..

### A mechanistic model for STD transmission



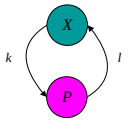

STDs are often modelled using frequency-dependent incidence:

$$f(S,I) = c_{FD}\, p_{FD} \left(\frac{S}{N}\right) I$$

$c_{FD}$ = rate of acquiring new partners
$p_{FD}$ = prob. of transmission in S-I partnership
$S/N$ = prob. that partner is susceptible
$I$ = density of infectives

Read more about pair-formation models for STDs in
Lloyd-Smith et al (2004) Proc Roy Soc B 271: 625-634

## Pair dynamics



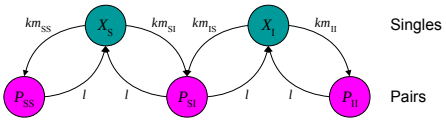

Singles

Pairs

$X$ = single individual
$P$ = pair

$k$ = pairing rate (per capita)
$l$ = pair dissolution rate

## Pair dynamics



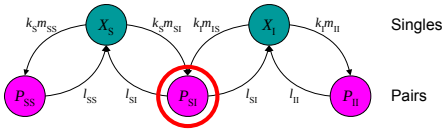

Singles

Pairs

$X_y$ = single individual of type $y$ (where $y$ = S or I)
$P_{yz}$ = pair of types $y$ and $z$ (where $y,z$ = S or I)

$k$ = pairing rate (per capita)
$l$ = pair dissolution rate
$m_{yz}$ = "mixing matrix"

## Pair dynamics




Singles

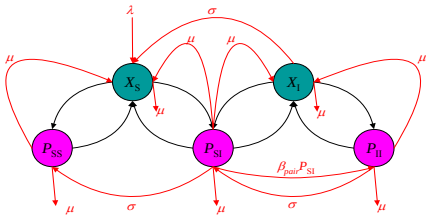
Pairs

$X_y$ = single individual of type $y$ (where $y$ = S or I)
$P_{yz}$ = pair of types $y$ and $z$ (where $y,z$ = S or I)

$k_y$ = pairing rate (per capita)
$l_{yz}$ = pair dissolution rate
$m_{yz}$ = "mixing matrix"

Transmission occurs only in S-I pairs ($P_{SI}$), at rate $\beta_{pair}$

## Pair-formation epidemic

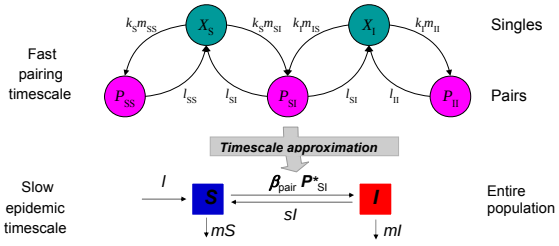



Consider populations where pairing dynamics are much faster than disease dynamics.

Timescale approximation: pairing dynamics are at quasi-steady-state relative to disease dynamics

(*c.f.* Heesterbeek & Metz (1993) J. Math. Biol. 31: 529-539.)

## Timescale approximation for pairing




$S = X_S + 2P_{SS} + P_{SI}$ $\quad I = X_I + 2P_{II} + P_{SI}$

**Challenge:** find $P^*_{SI}$ in terms of $S$, $I$, and pairing parameters. Then incidence rate = $\beta_{pair} P^*_{SI}$

Fast pairing dynamics

$$\begin{cases} \frac{dX_S}{dt} = -k_S X_S + 2l_{SS}P_{SS} + l_{SI}P_{SI} \\ \frac{dX_I}{dt} = -k_I X_I + 2l_{II}P_{II} + l_{SI}P_{SI} \\ \frac{dP_{SS}}{dt} = \frac{1}{2}k_S m_{SS} X_S - l_{SS}P_{SS} \\ \frac{dP_{SI}}{dt} = \frac{1}{2}k_S m_{SI} X_S + \frac{1}{2}k_I m_{IS} X_I - l_{SI}P_{SI} \\ \frac{dP_{II}}{dt} = \frac{1}{2}k_I m_{II} X_I - l_{II}P_{II} \end{cases}$$

Slow disease dynamics

$$\begin{cases} \frac{dS}{dt} = \lambda - \beta_{pair} P^*_{SI} + \sigma I - \mu S \\ \frac{dI}{dt} = \beta_{pair} P^*_{SI} - (\sigma + \mu) I \end{cases}$$

## Finding $P^*_{SI}$ from fast equations

Substitute:

$$\left.\begin{aligned} S &= X_S + 2P_{SS} + P_{SI} \\ I &= X_I + 2P_{II} + P_{SI} \end{aligned}\right\}$$ Total all susceptibles and infectives, single and paired

$$\left.\begin{aligned} m_{SS} = m_{IS} &= \frac{k_S X_S}{k_S X_S + k_I X_I} \\ m_{SI} = m_{II} &= \frac{k_I X_I}{k_S X_S + k_I X_I} \end{aligned}\right\}$$ Assume random mixing in pair formation

Set $dP_{yz}/dt$'s = 0, solve for $P^*_{SI}$

## Simplest case: uniform behaviour

Disease status has no effect on pairing behaviour.

$k$ = pairing rate for all individuals

$l$ = break-up rate for all partnerships

Incidence rate = $\beta_{pair} P^*_{SI} = \beta_{pair}\left(\frac{k}{k+l}\right)\frac{SI}{N}$

Recall the FD incidence: $c_{FD} p_{FD}\left(\frac{S}{N}\right) I$

## Pair-based transmission and frequency dependence

$c_{FD}$ = rate of acquiring partners

$$= \frac{1}{1/l + 1/k} = \frac{kl}{k+l}$$

$p_{FD}$ = probability of transmission in S-I partnership

$$= 1 - \exp(-\beta_{pair} \times 1/l) \approx \beta_{pair}/l \quad (\text{since } \beta_{pair} << l)$$

$$c_{FD} p_{FD} \frac{SI}{N} \approx \beta_{pair}\left(\frac{k}{k+l}\right)\frac{SI}{N}$$

## Pair-based transmission and frequency dependence

Frequency dependence can represent pair-based transmission but timescale approximation is required.

Conversely:

We know STD dynamics are driven by pair-based transmission.

➡ FD models implicitly make timescale approximation.

Use mechanistic derivation of FD to assess this assumption.

## Application to STD models

Transient, highly-transmissible STDs

- High chance of infection per exposure
- Most individuals recover within a month
- e.g. gonorrhoea, chlamydia

Many bacterial STDs

Chronic, less-transmissible STDs

- Low chance of infection per exposure
- No recovery!
- e.g. HIV, HSV-2

Many viral STDs

## When does FD adequately represent pair-based transmission?

Compare simulations:

frequency-dependent incidence

*vs.*

full simulation of pair dynamics and disease

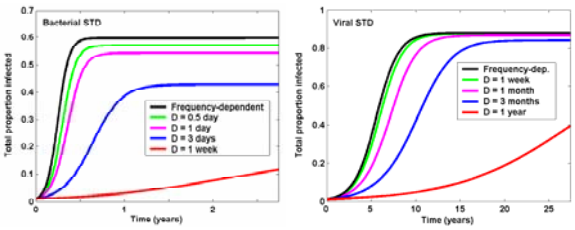
for different timescales of:

- disease – bacterial and viral STDs
- pairing dynamics – define average pair lifetime, D

$$D = \frac{1}{l} = \frac{1}{k}$$

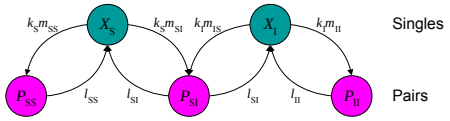
Transient, highly-transmissible STD

Chronic, less-transmissible STD




Frequency dependence is a good depiction of pair-based transmission only when mixing occurs fast compared to disease timescales.

## Modelling disease-induced behaviour changes




Four cases:

1. No effect on behaviour
2. Disease alters pair-formation rate, $k_S \neq k_I$
3. Disease alters break-up rate, $l_{SS} \neq l_{SI} \neq l_{II}$
4. Disease alters both $k_y$ and $l_{yz}$ ($y$, $z$ = S or I)

## Modelling disease-induced behaviour changes

For all four cases, the incidence rate takes a generalized frequency-dependent form:

$$\beta_{\text{pair}} \phi_{\kappa}(s,i) \frac{SI}{N}$$

where $\phi_{\kappa}(s,i)$ is a function of $s=S/N$, $i=I/N$ and the pairing parameters.

| Case | Rates, k | $\phi_k(s,i)$ |
|---|---|---|
| 1 | $k_S=k_I=k$<br>$l_{SS}=l_{SI}=l_{II}=l$ | $\frac{k}{k+l}$ |
| 2 | $k_S \neq k_I$<br>$l_{SS}=l_{SI}=l_{II}=l$ | $\frac{\pi_S \pi_I}{\pi_S s + \pi_I i}$ |
| 3 | $k_S=k_I=k$<br>$l_{SS} \neq l_{SI} \neq l_{II}$ | $\frac{\pi}{\frac{1}{2}+\frac{1}{2}\sqrt{1-4a\pi^2 si}}$ |
| 4 | $k_S \neq k_I$<br>$l_{SS} \neq l_{SI} \neq l_{II}$ | $\frac{\pi_S \pi_I}{\frac{1}{2}\left(\pi_S s + \pi_I i + \sqrt{(\pi_S s + \pi_I i)^2 - 4a(\pi_S \pi_I)^2 si}\right)}$ |

where $s=S/N$, $i=I/N$, $\pi_y=k_y/(k_{y+}l_{SI})$ and $a = \frac{l_{SI}}{k_I}\left(1-\frac{l_{SI}}{l_{SS}}\right)+\frac{l_{SI}}{k_S}\left(1-\frac{l_{SI}}{l_{II}}\right)+\left(1-\frac{l_{SI}^2}{l_{SS}l_{II}}\right)$

If $k_S=k_I$, then $\pi_S = \pi_I \equiv \pi$.

## Calculation of $R_0$

$R_0 = \lim_{S \to N}$ (transmission rate per I individual × duration of infectiousness)

$$= \lim_{S \to N}\left(\beta_{\text{pair}} \phi_{\kappa}(s,i) \frac{S}{N} \times \frac{1}{\sigma + \mu}\right)$$

$$= \frac{\beta_{\text{pair}}}{\sigma + \mu} \lim_{S \to N}\left(\phi_{\kappa}(s,i)\right)$$

$$= \frac{\beta_{\text{pair}}}{\sigma + \mu}\left(\frac{k_I}{k_I + l_{SI}}\right) \quad \text{in all four cases.}$$

- No dependence on $k_S$
- No dependence on $l_{SS}$ or $l_{II}$
- Identical to standard FD result, $R_0 = \frac{c_{FD} p_{FD}}{\sigma + \mu}$

  if $c_{FD}$ = contact rate of *infected* individuals

## Calculation of stability threshold

Consider stability threshold of the no-infection equilibrium, when population is wholly susceptible:

$R_0 > 1 \leftrightarrow$ no-infection equilibrium is unstable to perturbations in $I$

$$R_0 > 1 \leftrightarrow \left[\frac{\partial f_I}{\partial I}\right]_{S \to N} > 0, \quad \text{where } \frac{dI}{dt} = f_I(S,I)$$

Yields the same result as $R_0$ calculation in all four cases – though note that just because a quantity is an epidemic threshold parameter does not mean that it equals $R_0$!!

e.g. $(R_0)^k$ for any $k>0$ also has an epidemic threshold at 1.