Introduction to infectious disease modelling

Jamie Lloyd-Smith

Center for Infectious Disease Dynamics
Pennsylvania State University

with thanks to Ottar Bjornstad for sharing some slides…

Why do we model infectious diseases?

Following Heesterbeek & Roberts (1995)

1. Gain insight into mechanisms influencing disease spread, and link individual scale 'clinical' knowledge with population-scale patterns.
2. Focus thinking: model formulation forces clear statement of assumptions, hypotheses.
3. Derive new insights and hypotheses from mathematical analysis or simulation.
4. Establish relative importance of different processes and parameters, to focus research or management effort.
5. Thought experiments and "what if" questions, since real experiments are often logistically or ethically impossible.
6. Explore management options.

Note the absence of predicting future trends. Models are highly simplified representations of very complex systems, and parameter values are difficult to estimate.
→ quantitative predictions are virtually impossible.

Epidemic models: the role of data

Why work with data?
Basic aim is to describe real patterns, solve real problems.
Test assumptions!
Get more attention for your work
→ jobs, fame, fortune, etc
→ influence public health policy

Challenges of working with data
Hard to get good data sets.
The real world is messy! And sometimes hard to understand.
Statistical methods for non-linear models can be complicated.

What about pure theory?
Valuable for clarifying concepts, developing methods, integrating ideas.
(My opinion) The world (and Africa) needs a few brilliant theorists, and *many* strong applied modellers.


The SEIR framework for microparasite dynamics

Susceptible: naïve individuals, susceptible to disease

Exposed: infected by parasite but not yet infectious

**Infectious:** able to transmit parasite to others

Removed: immune (or dead) individuals that don't contribute to further transmission


The SEIR framework for microparasite dynamics

λ   "Force of infection"
$= \beta I$ under density-dependent transmission
$= \beta I/N$ under frequency-dependent transmission

ν   Rate of progression to infectious state
= 1/latent period

γ   Rate of recovery
= 1/infectious period

The SEIR framework for microparasite dynamics
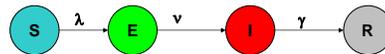
$$\frac{dS}{dt} = -\frac{\beta SI}{N}$$

$$\frac{dE}{dt} = \frac{\beta SI}{N} - \nu E$$

$$\frac{dI}{dt} = \nu E - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

Ordinary differential equations are **just one approach** to modelling SEIR systems.

SEI

SIRS

SIS
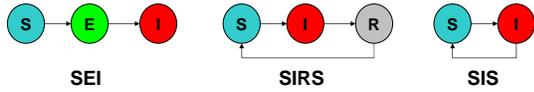
Adapt model framework to disease biology and to your problem!
No need to restrict to SEIR categories, if biology suggests otherwise.
e.g. leptospirosis has chronic shedding state → SICR
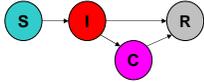
Depending on time-scale of disease process (and your questions),
add host demographic processes.

births

deaths
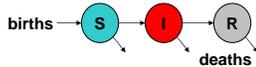
Disease with environmental reservoir (e.g. anthrax)

Death of pathogen in
environment

Vector-borne disease

Humans

birth

Vectors

death

TB treatment model

RN

ss– DOTS

ss– non-DOTS

ss+ DOTS

ss+ non-DOTS

Susc

Slow

Fast

ss–

ss+

ss–
det

ss+
det

Rec

Part.
rec

Residence times

E

ν
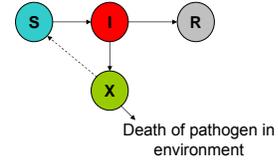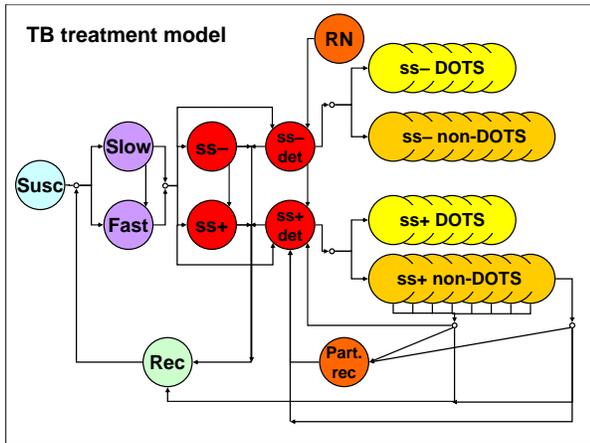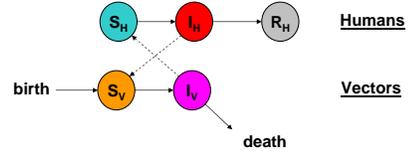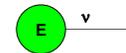
How long does an individual spend in the E compartment?
Ignoring further input from new infections:

$$\frac{dE}{dt} = -\nu E \quad \Rightarrow \quad E(t) = E(0)\,e^{-\nu t}$$

For a constant per capita rate of leaving compartment, the
<u>residence time</u> in the compartment is exponentially distributed.

ODE model

Data from
SARS

t

t

Residence times

Data from
SARS

t
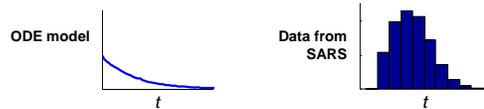
How to make the model fit the data better?
• "Box-car model" is one modelling trick

S $\quad \lambda \quad$ E_1 $\quad \nu/n \quad$ E_2 $\quad \nu/n \quad$ ... E_n $\quad \nu/n \quad$ I
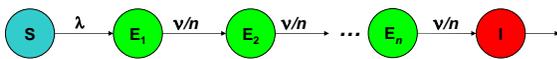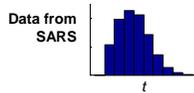
Divide compartment into n sub-compartments, each with constant
leaving rate of ν/n.

Residence time is now gamma-
distributed, with same mean and
flexible variance depending on the
number of sub-compartments.

$n=40$

$n=10$

$n=3$ $\quad n=1$

t
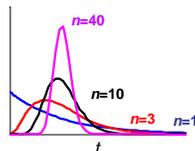
See Wearing et al (2005) PLoS Med 2: e174

Basic reproductive number, R_0

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

$R_0 \leq 1$

$R_0 > 1$

disease dies out

disease can invade

Outbreak dynamics
• probability of fade-out
• epidemic growth rate

Disease control
• threshold targets
• vaccination levels

2
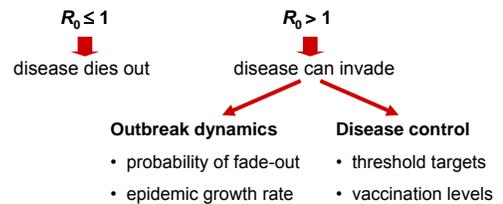
Calculating R_0 – Intuitive approach

$R_0 = $ Per capita rate $\times$ Duration of
of infecting others   infectiousness
… in a completely susceptible population.

Under frequency-dependent transmission:

Rate of infecting others $= \beta\, S/N$
$= \beta$ in wholly susceptible pop'n

Duration of infectiousness $= 1/$recovery rate
$= 1/\gamma$

$\rightarrow R_0 = \beta / \gamma$

Effective reproductive number

Expected number of cases caused by a typical infectious individual in a population that is not wholly susceptible.

$$R_{\text{effective}} = R_0 \times S/N$$

<u>Endemic disease:</u> At equilibrium $R_{\text{eff}} = 1$, so that $S^*/N = 1/R_0$

<u>Epidemic disease:</u> R_{eff} changes as epidemic progresses, as susceptible pool is depleted.
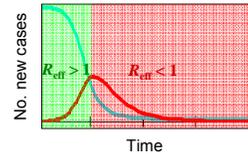
<u>Note:</u> Sometimes "effective reproductive number" is used to describe transmission in the presence of disease control measures. This is also called R_{control}.

$R_{\text{effective}}$ and herd immunity

$$R_{\text{effective}} = R_0 \times S/N$$

If a sufficiently high proportion of the population is immune, then $R_{\text{effective}}$ will be below 1 and the disease cannot circulate.

The remaining susceptibles are protected by **herd immunity**.

The critical proportion of the population that needs to be immune is determined by a simple calculation:

- For $R_{\text{eff}} < 1$, we need $S/N < 1/R_0$

- Therefore we need a proportion $1 - 1/R_0$ to be immune.

What does R_0 tell you?

- Epidemic threshold
  - NOTE: not every epidemic threshold parameter is R_0!
- Probability of successful invasion
- Initial rate of epidemic growth
- Prevalence at peak of epidemic
- Final size of epidemic (or the proportion of susceptibles remaining after a simple epidemic)
- Mean age of infection for endemic infection
- Critical vaccination threshold for eradication
- Threshold values for other control measures

The basic framework for macroparasite dynamics

For macroparasites the **intensity** of infection matters!
Basic model for a directly-transmitted macroparasite:

<u>State variables</u>

$N(t)$ = Size of host population

$M(t)$ = Mean number of sexually mature worms in host population

$L(t)$ = Number of infective larvae in the habitat

The basic framework for macroparasite dynamics

$$\frac{dM}{dt} = d_1 \beta\, L(t - \tau_1) - (\mu + \mu_1) M$$

$$\frac{dL}{dt} = s\, d_2 \lambda\, NM(t - \tau_2) - \mu_2 L - \beta N L$$

β infection rate
μ death rate of hosts
μ_1 death rate of adult worms within hosts
μ_2 death rate of larvae in environment
d_1 proportion of ingested larvae that survive to adulthood
d_2 proportion of eggs shed that survive to become infective larvae
τ_1 time delay for maturation to reproductive maturity
τ_2 time delay for maturation from egg to infective larva
s proportion of offspring that are female

<u>Further complexities:</u> parasite aggregation within hosts and density-dependent effects on parasite reproduction.

R_0 for macroparasites

For macroparasites, R_0 is the average number of female offspring (or just offspring in the case of hermaphroditic species) produced throughout the lifetime of a mature female parasite, which themselves achieve reproductive maturity in the absence of density-dependent constraints on the parasite establishment, survival or reproduction.
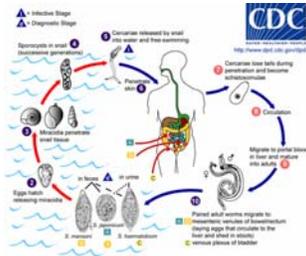
Effective R_0 for macroparasites

For macroparasites, R_{eff} is the average number of female offspring produced in a host population within which density dependent constraints limit parasite population growth.

For microparasites, R_{eff} is the reproductive number in the presence of competition for hosts at the population scale.

For macroparasites, R_{eff} is the reproductive number in the presence of competition at the within-host scale.

For both, under conditions of stable endemic infection, $R_{eff}=1$.

Major decisions in designing a model

Even after compartmental framework is chosen, still need to decide:

- Deterministic vs stochastic
- Discrete vs continuous time
- Discrete vs continuous state variables
- Random mixing vs structured population
- Homogeneous vs heterogeneous

  (and which heterogeneities to include?)

Deterministic vs stochastic models

<u>Deterministic models</u>

- Given model structure, parameter values, and initial conditions, there is no variation in output.

<u>Stochastic models</u> incorporate chance.

- Stochastic effects are important when numbers are small, e.g. during invasion of a new disease

- Demographic stochasticity: variation arising because individual outcomes are not certain

- Environmental stochasticity: variation arising from fluctuations in the environment (i.e. factors not explicitly included in the model)

Important classes of stochastic epidemic models

Monte Carlo simulation

- Any model can be made stochastic by using a pseudo-random number generator to "roll the dice" on whether events occur.

Branching process

- Model of invasion in a large susceptible population

- Allows flexibility in distribution of secondary infections, but does not account for depletion of susceptibles.

Important classes of stochastic epidemic models

Chain binomial

- Model of an epidemic in a finite population.

- For each generation of transmission, calculates new infected individuals as a binomial random draw from the remaining susceptibles.

Diffusion

- Model of an endemic disease in a large population.

- Number of infectious individuals does a random walk around its equilibrium value → quasi-stationary distribution

Continuous vs discrete time

<u>Continuous-time models (ODEs, PDEs)</u>

$$\frac{dN}{dt} = \lambda N$$

- Well suited for mathematical analysis
- Real events occur in continuous time
- Allow arbitrary flexibility in durations and residence times

<u>Discrete-time models</u>

$$N(t+1) = \lambda N(t)$$

- Data often recorded in discrete time intervals
- Can match natural timescale of system, e.g. generation time or length of a season
- Easy to code (simple loop) and intuitive
- Note: can yield unexpected behaviour which may or may not be biologically relevant (e.g. chaos).

Continuous vs discrete state variables

<u>Continuous state variables</u> arise naturally in differential equation models.

- Mathematically tractable, but biological interpretation is vague (sometimes called 'density' to avoid problem of fractional individuals).
- Ignoring discreteness of individuals can yield artefactual model results (e.g. the "atto-fox" problem).
- Quasi-extinction threshold: assume that population goes extinct if continuous variable drops below a small value

<u>Discrete state variables</u> arise naturally in many stochastic models, which treat individuals (and individual outcomes) explicitly.

Models for population structure
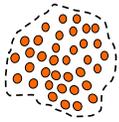
Random mixing **Multi-group** **Spatial mixing**
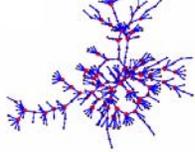
Network **Individual-based model**

Population heterogeneities

In real populations, almost everything is heterogeneous – no two individuals are completely alike.

Which heterogeneities are important for the question at hand? Do they affect epidemiological rates or mixing? Can parameters be estimated to describe their effect?

- often modelled using multi-group models, but networks, IBMs, PDEs also useful.

SIR output: the epidemic curve

S → I → R

$$\frac{dS}{dt} = -\frac{\beta S I}{N}$$

$$\frac{dI}{dt} = \frac{\beta S I}{N} - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

Susceptible
Removed
Infectious

Proportion of population

Time

SIR output: the epidemic curve

Proportion of population

$R_0=10$ $R_0=5$ $R_0=3$ $R_0=2$
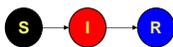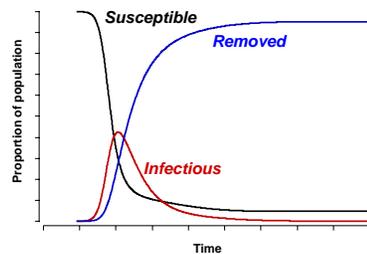
Time
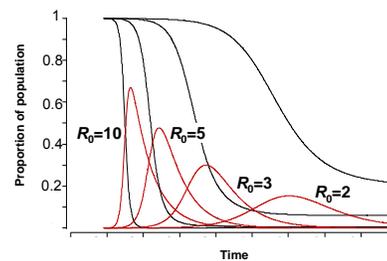
Basic model analyses (Anderson & May 1991):
Exponential growth rate, $r = (R_0 - 1)/D$
Peak prevalence, $I_{max} = 1 - (1 + \ln R_0)/R_0$
Final proportion susceptible, $f = \exp(-R_0[1-f]) \approx \exp(-R_0)$

SIR output: stochastic effects

SIR output: stochastic effects

6 stochastic epidemics with $R_0=3$.

Probability of disease extinction following introduction of 1 case.
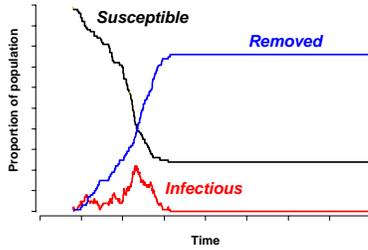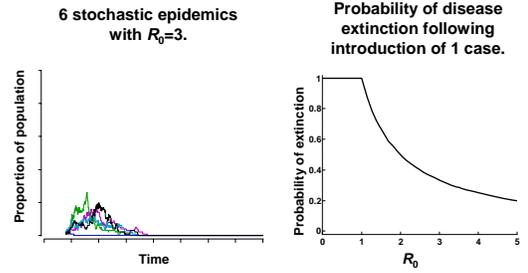
Stochasticity → risk of disease extinction when number of cases is small, even if $R_0>1$.

SIR with host demographics: epidemic cycles

births → S → I → R → deaths
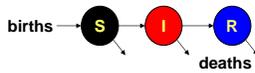
$$\frac{dS}{dt} = \mu N - \frac{\beta SI}{N} - \mu S$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - (\gamma + \mu)I$$

$$\frac{dR}{dt} = \gamma I - \mu R$$
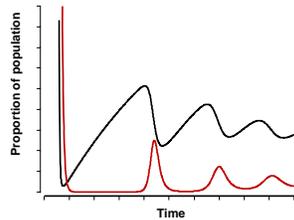
Cycle period $T \approx 2\pi (A\,D)^{1/2}$

where A = mean age of infection

D = disease generation interval

or can solve T in terms of SIR model parameters by linearization.
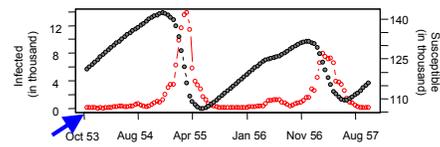
The S-I phase plot
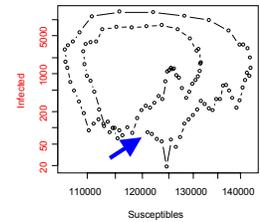
Summary of simple epidemic patterns

- Absence of recovery: logistic epidemic
- No susceptible recruitment (birth or loss of immunity): simple epidemics
- Susceptible recruitment through birth (or loss of immunity): recurrent epidemics
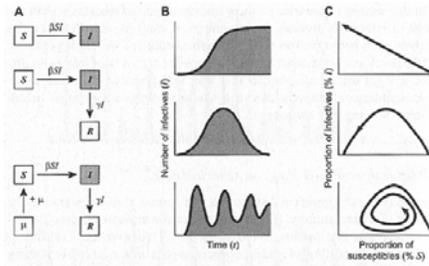
Herd immunity and epidemic cycling

Herd immunity prevents further outbreaks until S/N rises enough that $R_{eff} > 1$.
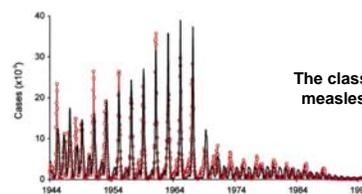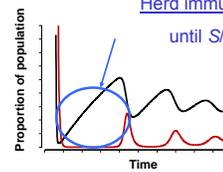
The classic example: measles in London

Herd immunity and epidemic cycling

Measles in London   Grenfell et al. (2001)

Cycle period depends on the effective birth rate.
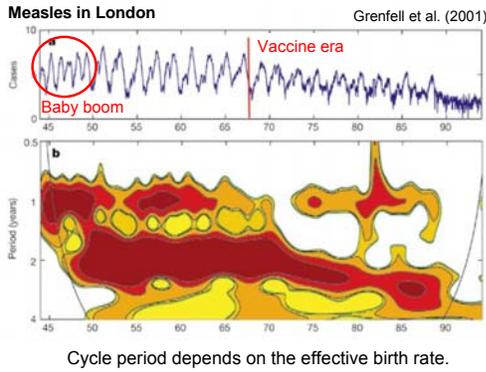
Persistence and fadeouts

Measles again…

Note that measles dies out between major outbreaks in Iceland, but not in the UK or Denmark.

What determines persistence of an acute infection?

NB: Questions like this are where "atto-foxes" can cause problems.
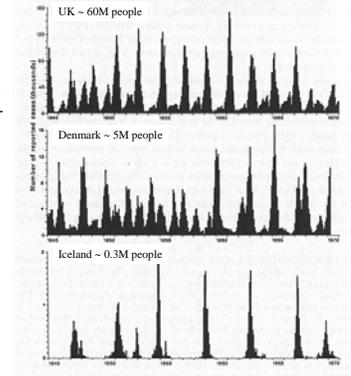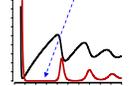
Intrinsic vs extrinsic forcing – what determines outbreak timing?

Untangling the relative roles of
intrinsic forcing (population dynamics and herd immunity)
and
extrinsic forcing (environmental factors and exogenous inputs)
is a central problem in population ecology.

This is particularly true for 'outbreak' phenomena such as infectious diseases or insect pests, where dramatic population events often prompt a search for environmental causes.

Leptospirosis in California sea lions
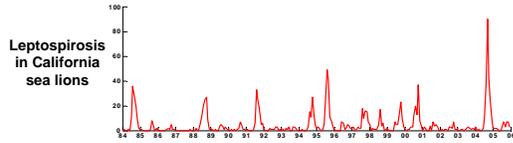
Intrinsic vs extrinsic forcing – what determines outbreak timing?

Example: leptospirosis in California sea lions

Intrinsic factors
Host population size and structure, recruitment rates and herd immunity

Extrinsic factors
Pathogen introduction: contact with reservoirs, invasive species, range shifts
Climate: ENSO events, warming temperatures
Malnutrition: from climate, fisheries or increasing N
Pollution: immunosuppressive chemicals, toxic algae blooms
Human interactions: Harvesting, protection, disturbance

Data needs I. What's needed to build a model?

Individual "clinical" data
- Latent period: time from infection to transmissibility
- Infectious period: duration (and intensity) of shedding infectious stages
- Immunity: how effective, and for how long?

Population data
- Population size and structure
- Birth and death rates, survival, immigration and emigration
- Rates of contact within and between population groups

Epidemiological data
- Transmissibility (R_0)
    - density dependence, seasonality

Data needs II. What's needed to validate a model?

Time series
- Incidence: number of new cases
- Prevalence: proportion of population with disease

Seroprevalence / sero-incidence: shows individuals' history of exposure.

Age/sex/spatial structure, if present.
    e.g. mean age of infection → can estimate R_0

Cross-sectional data
Seroprevalence survey (or prevalence of chronic disease)
    endemic disease at steady state → insight into mixing
    epidemic disease → outbreak size, attack rate, and risk groups

Contact tracing

SARS transmission chain, Singapore 2003
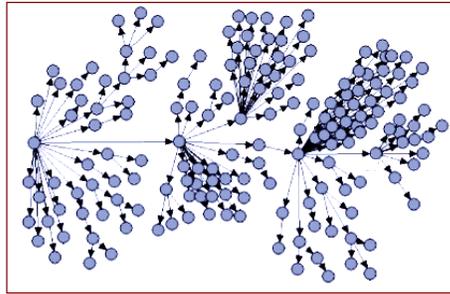
Morbidity & Mortality Weekly Report (2003)

Household studies

Observed time intervals between two cases of measles in families of two children. Data from Cirencester, England, 1946-1952 (Hope-Simpson 1952)

Presumed double primaries

Presumed within-family transmission
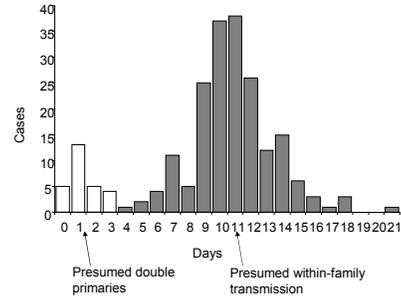
Measles:
Latent period 6-9 d, Infectious period 6-7 d, Average serial interval: 10.9 d

Long-term time series

Historical mortality records provide data:
London Bills of mortality for a week of 1665

Today: several infections are 'notifiable'

CDC Morbidity and Mortality
Weekly Report

http://www.who.int/research/en/

Outbreak time series

• Journal articles

http://www.who.int/wer/en/

http://www.cdc.gov/mmwr/
http://www.eurosurveillance.org

Grenfell & Anderson's (1989) study of whooping cough
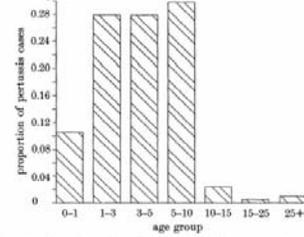
FIGURE 2. Proportional age distribution of whooping cough notifications in England and Wales for the period 1944 to 1946.

Age-incidence

e.g. Walsh (1983) of measles in urban vs rural settings in central Africa
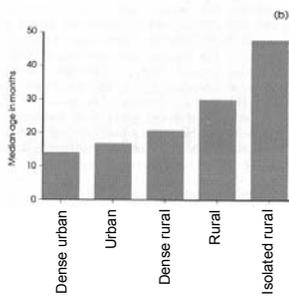
Age-seroprevalence curves

Rubella in Gambia

Rubella in UK

mumps

poliovirus

Hepatitis B virus

Age is in years
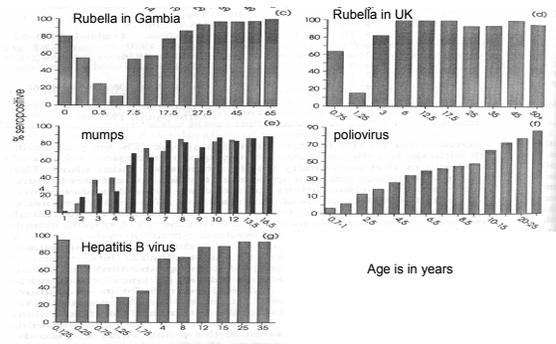
Seroprevalence: Proportion of population carrying antibodies indicating past exposure to pathogen.

Increased transmission leaves signatures in seroprevalence profiles

e.g. measles in small (grey) and large (black) families
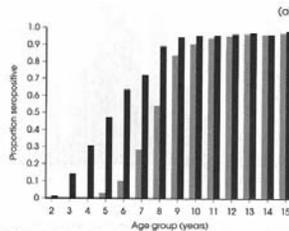
Fig. 3.10. (a) The proportion of an age group with antibodies specific to measles virus antigens in children from small and large families in the United States in 1957 prior to the introduction of mass vaccination (data from Black 1959). Family size clearly has an important influence on immunity to measles at different ages.

Two books full of data on important global health problems
 - PDF versions free to download.

http://www.dcp2.org/pubs/GBD   http://www.dcp2.org/pubs/DCP
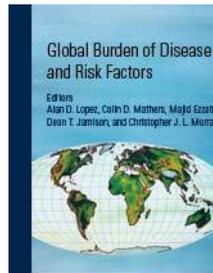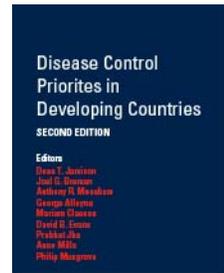
Other fields of disease modelling

<u>Within-host models</u>

• pathogen population dynamics and immune response
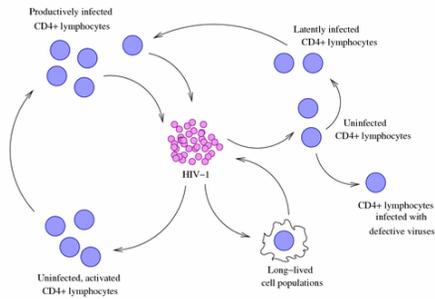
Other fields of disease modelling

<u>Pathogen evolution</u>

• adaptation to new host species, or evolution of drug resistance
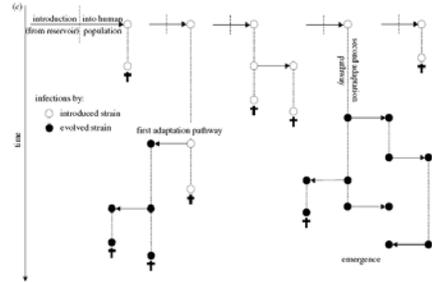
Other fields of disease modelling

<u>Phylodynamics</u>

• how epidemic dynamics interact with pathogen molecular evolution
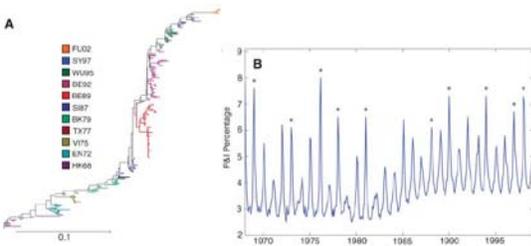
Community dynamics of disease

<u>Co-infections</u>

What happens when multiple parasites are present in the same host?

How do they interact? Resource competition? Immune-mediated indirect competition? Facilitation via immune suppression

<u>Multiple host species</u>

Many pathogens infect multiple species

  - when can we focus on one species?

  - how can we estimate importance of multi-species effects?

Zoonotic pathogens – many infections of humans have animal reservoirs, e.g. flu, bovine TB, yellow fever, Rift valley fever

Reservoir and spillover species

Host jumps and pathogen emergence