

# Modelling Infectious Diseases and Health Economic Evaluation of Vaccines

Day 1: Mathematical Models of Infectious Diseases

Day 2: Estimating Infectious Disease Parameters from Serological and Social Contact Data

Niel Hens



Interuniversity Institute for Biostatistics  
and statistical Bioinformatics



Health Economics & Modelling Infectious Diseases  
Vaccine & Infectious Disease Institute  
University of Antwerp

Antwerp, September 6-11, 2021

# Course outline

Day I: Introduction to mathematical models for infectious diseases  
(11h-17h)

Day II: Inferring model parameters from data  
(9h-17h)

Day III: Meta-population and individual-based Models  
(9h-17h)

Day IV: Introduction to health economic evaluation and dealing with uncertainty  
(9h-17h)

Day V: Economic evaluation of vaccination programmes, specific issues  
(9h-15h)

# Introduction

- There are countless of **infectious agents** that can infect human, animal and plant hosts.
- They can be **transmitted directly** between hosts via respiratory air droplets or bodily fluids.
- They can also be **transmitted indirectly** through an intermediary source, for instance via mosquitoes, ticks, rodents, environmental particles or contaminated blood products.
- Infectious agents **evolve and transform** while new agents **emerge** regularly, implying their supply can be considered infinite.

# Introduction

- Microscopically small infectious agents with relatively short life spans, which replicate within their hosts are called **microparasites** such as viruses, bacteria and fungi.
- Larger infectious agents with relatively longer life spans called **macroparasites** such as parasitic worms.
- Many infectious agents live inside or on the surface of their hosts' bodies without causing illness or even discomfort. In fact, hosts even **depend for their survival on infectious agents** (e.g., bacteria in the human gut).
- Infectious agents causing disease in their hosts are often referred to as **pathogens**.
- **Infectious diseases** are caused by pathogens, which are transmissible between hosts, either directly or indirectly.

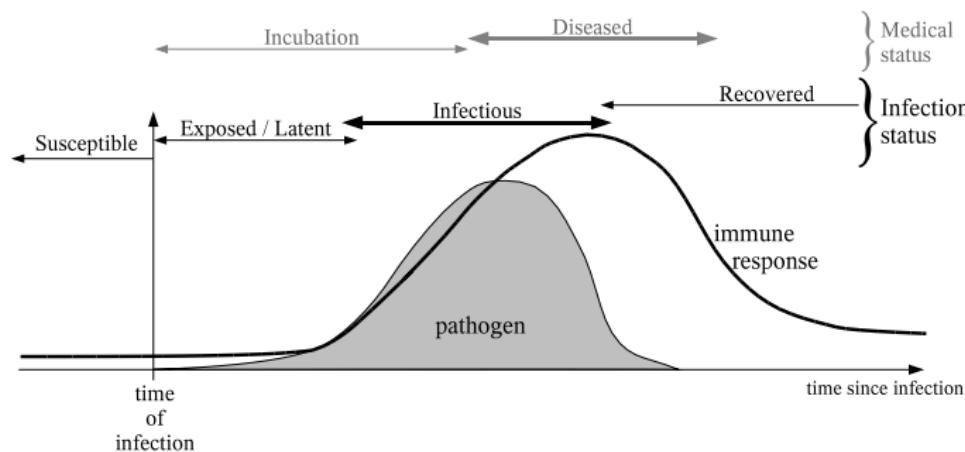
# Introduction

An infectious disease is often characterized by its

- transmission route
- transmission potential
- natural history
  - incubation time
  - latent or pre-infectious period
  - infectious period
- acquired immunity
- symptomatic/asymptomatic

# Natural history

## The Infectious Disease Process:



Keeling and Rohani (2008)

# Transmission routes

- **Close contacts** transmission (Varicella, Influenza, Mumps, Tuberculosis, Measles, SARS-CoV-2, ...)
- **Sexual** transmission (Hepatitis B, HIV, HPV, ...)
- **Parenteral** transmission (Hepatitis B and C, HIV, ...)
- **Orofecal** transmission (Hepatitis A, Cholera, Rotavirus, ...)
- **Vertical** transmission (Hepatitis B and C, HIV, Syphilis, ...)
- **Vectorborne** transmission (Malaria, Dengue, Leishmaniasis, ...)
- ...

## Historic pandemics

- Plague of Justinian, from 541 to 750, killed between 50% and 60% of Europe's population
- The Black Death of 1347 to 1352 killed 25 million in Europe over 5 years (estimated to be between 25 and 50% of the populations of Europe, Asia, and Africa - the world population at the time was 500 million)
- The introduction of smallpox, measles, and typhus to the areas of Central and South America by European explorers during the 15th and 16th centuries caused pandemics among the native inhabitants
- Between 1518 and 1568 disease pandemics are said to have caused the population of Mexico to fall from 20 million to 3 million

## Historic pandemics

- The first European influenza epidemic occurred between 1556 and 1560, with an estimated mortality rate of 20%
- Smallpox killed an estimated 60 million Europeans during the 18th century
- In the 19th century, tuberculosis killed an estimated one-quarter of the adult population of Europe; by 1918 one in six deaths in France were still caused by TB.
- The Influenza Pandemic of 1918 (or the Spanish Flu) killed 25-50 million people
- UNAIDS estimates that in 2019, 38 million people live with HIV and 690k people died due to AIDS worldwide
- Numbers for SARS-CoV-2 (status 05-09-2021): 4.6 million deaths officially → estimated true number of deaths: 15.2 million deaths (The Economist)

## Part I

# Mathematical Models of Infectious Diseases

# Outline

## 2 Introduction

## 3 Compartmental Models

- Discrete time models
- A mathematical intermezzo
- Kermack and McKendrick's Model
  - The SIR model with vital dynamics
  - The SIR model with vital dynamics and vaccination
- Building your own model
- Transmission within multiple subpopulations
- Who Acquires Infection From Whom?

## 4 Epilogue

# Mathematical modelling of infectious diseases

- Purposes:
  - **prediction**: requires the inclusion of known complexities and population-level heterogeneity
  - **understanding**: investigating the factors that drive dynamics
- Building a model presents a trade-off:
  - **accuracy**: reproduce what is observed and predict future dynamics
  - **transparency**: ability to understand how model components influence the dynamics and interact
  - **flexibility**: ease of adapting the model to new situations

# Mathematical modelling of infectious diseases

- Limitations:

- models present a **simplification of reality**
- chance events of infectious disease transmission **hinder perfect prediction**

- A good model:

- suited to its purpose: **simple as possible, but no simpler**
- **balance accuracy, transparency, flexibility**
- **parametrisable from available data**

# Mathematical modelling of infectious diseases

- Daniel **Bernoulli** was the first to present a [mathematical model for smallpox](#) in 1760, published in 1766
- Since then many people have developed models to describe infectious disease dynamics, see e.g. [Bailey \(1975\)](#); [Anderson and May \(1992\)](#); [Grenfell and Dobson \(1995\)](#); [Daley and Gani \(1999\)](#); [Hethcote \(2000\)](#)
- Several textbooks are available; amongst which [Vynnycky and White \(2010\)](#) provides an excellent introduction to mathematical modelling

Q & A on Vynnycky and White (2010)

# Contact between individuals

- Predicting the number of infections at time  $t + 1$  based on the circumstance at time  $t$
- The force of infection  $\lambda$ 
  - the per capita rate at which a susceptible individual contracts infection
  - it is assumed proportional to the number of infectious persons at time  $t$  and depending on how the contact structure is assumed to change with population size  $N$  it is given by:

$$\lambda_t = \beta I_t$$

$$\lambda_t = \beta I_t / N_t$$

# Contact between individuals

- The number of new infections at time  $t + 1$  is given by  $\lambda_t S_t$  and thus:

$$\begin{aligned}I_{t+1} &= \beta S_t I_t \\I_{t+1} &= \beta S_t I_t / N_t\end{aligned}$$

This is referred to as the mass action principle

- density-dependent transmission:  $I_{t+1} = \beta S_t I_t$ :
  - as the population size increases, so does the contact rate
  - mostly applicable to plant and animal diseases (homogeneity)
- frequency-dependent transmission:  $I_{t+1} = \beta S_t I_t / N_t$ :
  - the contact rate is assumed constant regardless of a change in population size
  - mostly applicable to human and vectorborne diseases (heterogeneity)

## Contact between individuals

- Note that when in a **constant population**, both density- and frequency-dependent are equivalent

$$I_{t+1} = \beta S_t I_t / N_t = \beta S_t I_t / N = \tilde{\beta} S_t I_t$$

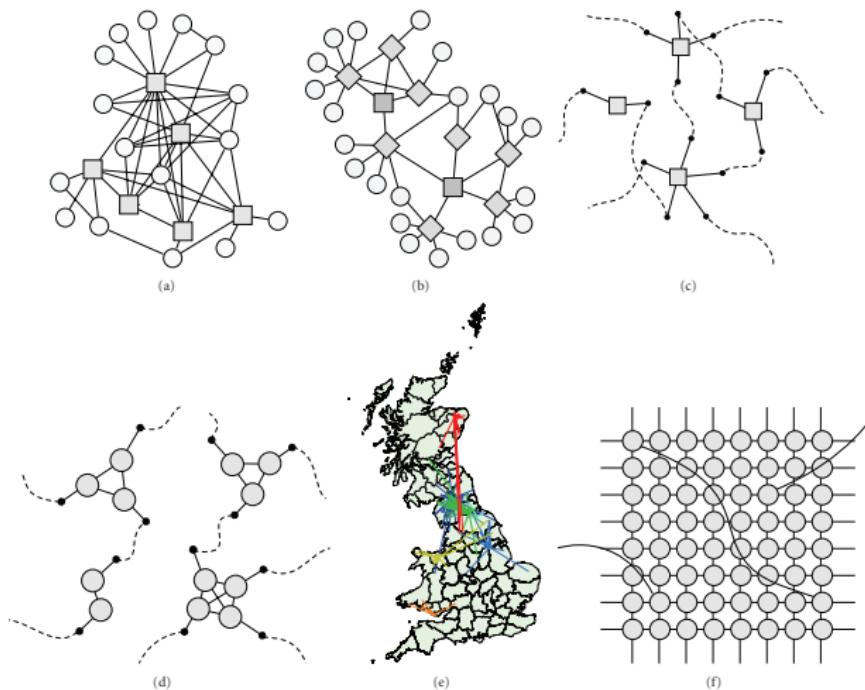
- In what follows I will use  $I_{t+1} = \beta S_t I_t$  without loss of generality unless specified differently.

# Contact between individuals

## Heterogeneity

- airborne infections: *age* - example: children at school have more contacts with children of the same age
- sexually transmitted infections: *age and sexual behavior*
- **temporal heterogeneity**
  - seasonality,
  - week vs weekend,
  - holiday vs non-holiday, ...
- ...

# Contact between individuals, plants, herds, ...



# Modeling frameworks

- compartmental models
  - the population is subdivided into broad subgroups (compartments)
  - individuals are tracked collectively
  - roughly either deterministic or stochastic (probabilistic)
    - deterministic models describe what happens ‘on average’ in a population
    - stochastic models allow the number of individuals who move between compartments to vary through chance
  - transmission dynamic or static models
- metapopulation models
- network models
- microsimulation or agent-based models

# Rules of engagement

| Entity                    | Symbol 1            | Symbol 2              |
|---------------------------|---------------------|-----------------------|
| maternal immune           | $M$ (number)        | $m_{pi}$ (proportion) |
| susceptibles              | $S$ (number)        | $s$ (proportion)      |
| exposed/latent*           | $E$ (number)        | $e$ (proportion)      |
| infected                  | $I$ (number)        | $i$ (proportion)      |
| recovered                 | $R$ (number)        | $r$ (proportion)      |
| total population size     | $N$ (number)        | -                     |
| age-stratified pop. size  | $N(a)$ (number)     | -                     |
| life expectancy           | $L$ (dec. number)   | -                     |
| survivor function         | -                   | $m$ (proportion)      |
| recovery rate             | $\nu$ (rate)        | -                     |
| force of infection        | $\lambda$ (rate)    | -                     |
| transmission parameter    | $\beta$ (rate)      | -                     |
| basic reproduction number | $R_0$ (dec. number) | -                     |

\*depending on the context  $e$  refers to  $\exp(1) = 2.71828\dots$

Table: Glossary of the most important symbols in this section.

# The basic SIR compartmental model

## ■ State variables

$S$ : susceptible individuals

$I$ : infectious individuals (able to transmit the infectious agent)

$R$ : removed/recovered ind. (not transmitting the infectious agent)

$N = S + I + R$ : total and constant population size

## ■ Parameters

$\beta$ : transmission rate

$\nu$ : recovery rate

## ■ Force of infection

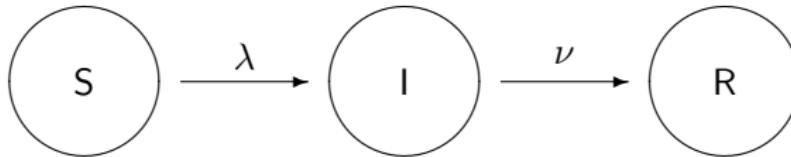
follows the law of mass action:  $\lambda_t = \beta I_t$

■ → lifelong immunity after infection

■ ...

# Model formulation

- SIR model:



- Extensions
  - maternal immunity/passive immunity: MSIR
  - exposed or latent infection (pre-infectious): SEIR
  - waning of immunity: SIRS
  - susceptible upon recovery: SIS
  - infected for life: SI

## The discrete time deterministic SIR model

- SIR compartmental model
- Equations:

$$\begin{aligned}S_{t+1} &= S_t - \lambda_t S_t \\I_{t+1} &= I_t + \lambda_t S_t - \nu I_t \\R_{t+1} &= R_t + \nu I_t\end{aligned}$$

with  $N_{t+1} = S_{t+1} + I_{t+1} + R_{t+1} = S_t + I_t + R_t = N_t$ .

- $\lambda_t = \beta I_t$  and  $\nu$  are risks
- risks are related to rates as follows:

$$\text{risk} = 1 - e^{-\text{rate}}$$

- if the 'rate' is small:

$$\text{risk} \approx \text{rate}$$

## The discrete time stochastic SIR model

- The number of newly infected cases arises from a stochastic process with mean  $\beta I_t S_t$
- The number of newly recovered individuals arises from a stochastic process with mean  $\nu_t I_t$
- What distribution would you assume for both these quantities?

## The discrete time stochastic SIR model

- The number of **newly infected** cases arises from a **stochastic process with mean  $\beta I_t S_t$**
- The number of **newly recovered** individuals arises from a **stochastic process with mean  $\nu_t I_t$**
- What distribution would you assume for both these quantities?
  - a Poisson distribution
  - a binomial distribution
- What are the (dis)advantages for both distributions?

## The discrete time deterministic SIR model

- When do you expect the number of new infections to decrease?
- Focus on the second equation:

$$I_{t+1} = I_t + \beta I_t S_t - \nu I_t$$

clearly  $I_{t+1} = I_t$  if  $\beta S_t = \nu$

- The epidemic will
  - decrease if  $S_t < \nu/\beta$
  - increase if  $S_t > \nu/\beta$
- $\nu/(\beta N)$  is called the relative removal rate
- $N\beta/\nu = N\beta D$  is called the basic reproduction number  $R_0$

Here  $D$  denotes the average infectious period

- Consequently  $\beta = R_0/(ND)$

- └ Compartmental Models
  - └ Discrete time models

## R software

- R: [www.r-project.org](http://www.r-project.org)
- Rstudio: [www.rstudio.com](http://www.rstudio.com)
- Other software: Matlab, Berkely Madonna ...

## R-code for the discrete stochastic SIR model

```
# Initial parameters
N = 1E6 # Total population
I = 1 # Number of Infectious at time 0
S = N-1 # Number of Susceptibles at time 0
R = 0 # Number of Recovered at time 0

# Transmission parameters
R0 = 2.48 # Basic Reproduction number
nu = 1/6 # Recovery rate (in days)
b = nu*R0/N # Infection rate (in days)

# Initial states
S0 = S # Number of Susceptibles at time t=0
I0 = I # Number of Infectious at time t=0
R0 = R # Number of Recovered at time t=0
```

## R-code discrete stochastic SIR model: option 1: Poisson

```
# Placeholders
Svec = Sold; Ivec = Iold; Rvec = Rold
stop = FALSE
# Loop - continue until stop = TRUE
while (!stop){
    Ih = rpois(1,b*Iold*Sold); print(Ih)
    Rh = rpois(1,nu*Iold)
    if ((Iold+Ih-Rh)<=0) break
    Sold = Sold-Ih
    if (Sold<=0){Rold=N; break}
    Iold = Iold+Ih-Rh
    Rold = Rold+Rh
    Svec = c(Svec,Sold)
    Ivec = c(Ivec,Iold)
    Rvec = c(Rvec,Rold)
    if (Iold==0){stop=T}
}
```

## R-code discrete stochastic SIR model: option 2: Binomial

```
# Placeholders
Svec = Sold; Ivec = Iold; Rvec = Rold
stop = FALSE
# Loop - continue until stop=TRUE
while (!stop){
    Ih = rbinom(1,Sold,(1-exp(-b*Iold)))
    print(Ih)
    Rh = rbinom(1,Iold,(1-exp(-nu)))
    Sold = Sold-Ih
    Iold = Iold+Ih-Rh
    Rold = Rold+Rh
    Svec = c(Svec,Sold)
    Ivec = c(Ivec,Iold)
    Rvec = c(Rvec,Rold)
    if (Iold==0){stop=T}
}
```

## Class exercise

- Run the discrete SIR model 100,000 times
- Keep track of the final size
- Study the distribution of final sizes. What do you observe?
- How does this distribution change with increasing  $R_0$ ?
- What happens if  $R_0$  decreases below 1?

## The basic reproduction number

- Recall  $R_0 = N\beta D$ , the basic reproduction number
- What does  $R_0$  represent?
- Consider the total number of new infections in the population between time  $t$  and  $t + 1$ :

$$\beta I_t S_t$$

- At the start of an epidemic, say  $t = 0$ :  $I_0 = 1$  and  $S_0 = N$  and thus the total number of new infections between  $t = 0$  and  $t = 1$  equals

$$\beta N$$

- By the end of the infectious period of duration  $D$  time units, the infectious person would have infected  $\beta ND$  individuals
- Therefore  $R_0$  is the number of secondary cases caused by a single infective introduced into a wholly susceptible population of size  $N$  during the infective's infectious period.

# The basic reproduction number

- $R_0$  constitutes a threshold:
  - if  $R_0 > 1$  then the epidemic can grow
  - if  $R_0 \leq 1$  then the epidemic will die out
- Using  $R_0$ , one defines the number of effective contacts by each person per unit time by

$$c_e = R_0/D.$$

- Therefore  $\beta = c_e/N$  is the “per capita number of effective contacts made by a given individual per unit time”, or equivalently “the per capita rate at which two specific individuals come into effective contact per unit time”
- For a given pathogen, it is difficult to define an effective contact
- $\beta$  and  $c_e$  are therefore often determined by specifying  $R_0$  and  $D$

## The basic reproduction number: examples

| Infectious Disease | $R_0$ |
|--------------------|-------|
| measles            | 10-20 |
| chickenpox         | 5-10  |
| mumps              | 5-10  |
| rubella            | 4-7   |
| smallpox           | 3-5   |

Farrington (*Modelling Epidemics*, 2003)

More on this later on

## Discrete time models: an alternative notation

- Other notation:

$$S_{t+\delta t} = S_t - \lambda_t S_t \delta t$$

$$I_{t+\delta t} = I_t + \lambda_t S_t \delta t - \nu I_t \delta t$$

$$R_{t+\delta t} = R_t + \nu I_t \delta t$$

with  $\delta t$  symbolizing the size of a small time step.

- The time step in a discrete time model could be very important.  
Why?

- Next:

→ consider continuous time models

→ focus on deterministic continuous time models

## From discrete to continuous time models

- Reordering the equation:

$$\frac{S_{t+\delta t} - S_t}{\delta t} = -\lambda_t S_t$$

$$\frac{I_{t+\delta t} - I_t}{\delta t} = \lambda_t S_t - \nu I_t$$

$$\frac{R_{t+\delta t} - R_t}{\delta t} = \nu I_t$$

## From discrete to continuous time models

- Limit for  $\delta t \rightarrow 0$ :

$$\lim_{\delta t \rightarrow 0} \frac{S_{t+\delta t} - S_t}{\delta t} = -\lambda_t S_t$$

$$\lim_{\delta t \rightarrow 0} \frac{I_{t+\delta t} - I_t}{\delta t} = \lambda_t S_t - \nu I_t$$

$$\lim_{\delta t \rightarrow 0} \frac{R_{t+\delta t} - R_t}{\delta t} = \nu I_t$$

## From discrete to continuous time models

- Using derivatives:

$$\begin{aligned}\frac{dS_t}{dt} &= -\lambda_t S_t \\ \frac{dI_t}{dt} &= \lambda_t S_t - \nu I_t \\ \frac{dR_t}{dt} &= \nu I_t\end{aligned}$$

# Population dynamics: Malthus (1798)

Evolution is determined by the balance between births and deaths:

$$\frac{dx}{dt} = bx - mx = (b - m)x \equiv rx.$$

- For  $r = 0$ , steady population:

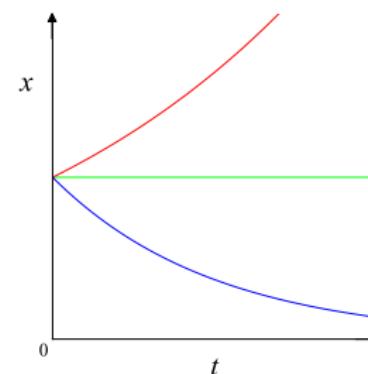
$$x(t) = x(0).$$

- For  $r > 0$ , exponential growth:

$$x(t) = x(0) e^{|r|t}.$$

- For  $r < 0$ , exponential decay:

$$x(t) = x(0) e^{-|r|t}.$$



# Population dynamics: Verhulst (1836)

Some limiting factor must prevent the population to grow indefinitely:

$$\frac{dx}{dt} = rx \left(1 - \frac{x}{K}\right).$$

- All solutions scale to one:

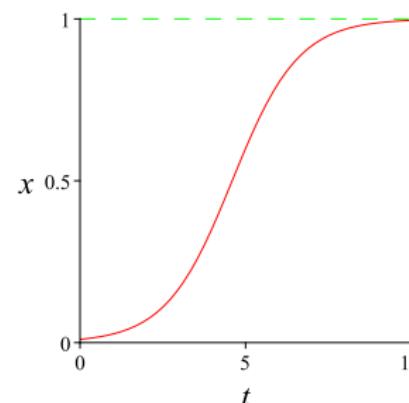
$$t \rightarrow r^{-1}t, \quad x \rightarrow Kx.$$

- Analytical solution:

$$x(t) = \frac{x(0) e^t}{1 + x(0)(e^t - 1)}.$$

- Horizontal asymptote:

$$\lim_{t \rightarrow +\infty} x(t) = 1.$$



# Linear stability analysis

Steady states are points where evolution stops:

$$\frac{dx}{dt} = F(x) = x(1-x) = 0 \quad \rightarrow \quad x = 0, 1.$$

- Small deviation from 0 increases:

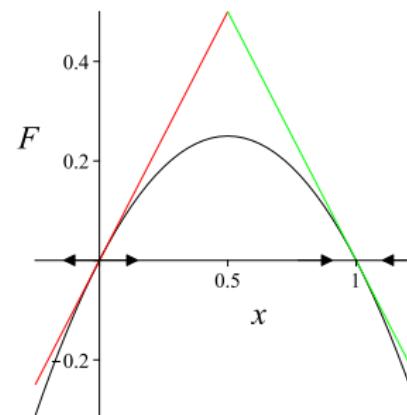
$$x(t) = \delta x(t) \rightarrow \delta x(t) \propto e^t,$$

the steady state is **unstable**.

- Small deviation from 1 decreases:

$$x(t) = 1 - \delta x(t) \rightarrow \delta x(t) \propto e^{-t},$$

the steady state is **stable**.



# A Contribution to the Mathematical Theory of Epidemics

*A Contribution to the Mathematical Theory of Epidemics.*

By W. O. KERMACK and A. G. MCKENDRICK.

(Communicated by Sir Gilbert Walker, F.R.S.—Received May 13, 1927.)

more than a contribution, a basis...

"Reference may here be made to the work of Ross and Hudson (1915-17) in which the same problem is attacked. The problem is here carried to a further stage, and it is considered from a point of view which is in one sense more general.[...]"

Kermack and McKendrick (1927)

# Kermack-McKendrick's model

The SIR model without vital dynamics  
(and without a general infectious period distribution)

## ■ State variables

$S$ : Susceptible individuals

$I$ : Infectious individuals (able to transmit the infectious agent)

$R$ : Removed/Recovered ind. (not transmitting the infectious agent)

$N = S + I + R$ : Total (and constant) population

## ■ Parameters

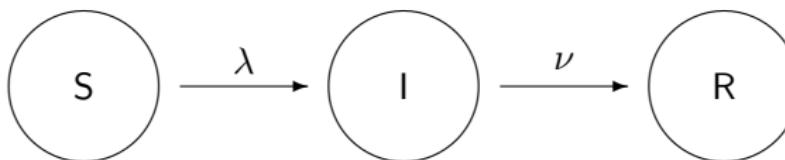
$\beta$ : transmission rate

$\nu$ : recovery rate

## ■ Force of infection (frequency dependent mass action)

follows the law of mass action:  $\lambda(t) = \beta \frac{I(t)}{N}$

## The SIR model



$$\begin{aligned}\frac{dS(t)}{dt} &= -\beta S \frac{I}{N}, \\ \frac{dI(t)}{dt} &= \beta S \frac{I}{N} - \nu I, \\ \frac{dR(t)}{dt} &= \nu I,\end{aligned}$$

with  $N = S + I + R$ .

Note that the time dependence is not always made explicit in the notation; here in the right-hand side (RHS) of the equation.

## Analytic approximation of the incidence

- Let  $s, i, r$  denote the proportion of population in the susceptible, infectious and removed classes, respectively:

$$s = \frac{S}{N}, \quad i = \frac{I}{N}, \quad r = \frac{R}{N},$$

$$\tau = \nu t, \quad \sigma = \frac{\beta}{\nu}.$$

- The equations for the SIR model can be written as

$$\begin{aligned}\frac{ds(\tau)}{d\tau} &= -\sigma si, \\ \frac{di(\tau)}{d\tau} &= \sigma si - i, \\ \frac{dr(\tau)}{d\tau} &= i,\end{aligned}$$

with  $s + i + r = 1$ .

## Analogy with ecology

- For very slow recovery,  $\sigma \gg 1$ , the system reduces to

$$\begin{aligned}\frac{ds}{d\tau} &= -\sigma si, \\ \frac{di}{d\tau} &= \sigma si,\end{aligned}$$

with  $s + i = 1$ .

- It is equivalent to two logistic equations

$$\begin{aligned}\frac{ds}{d\tau} &= -\sigma s(1 - s), \\ \frac{di}{d\tau} &= \sigma(1 - i)i.\end{aligned}$$

## Analytic approximation of the incidence

- Substituting  $i = 1 - s - r$  in 1<sup>st</sup> and 3<sup>rd</sup> equation of system our SIR model:

$$\begin{aligned}\frac{dr}{d\tau} &= 1 - s - r \\ \frac{ds}{dr} &= -\sigma s\end{aligned}$$

with initial values  $s(0) = s_0$  and  $r(0) = 0$ , one obtains

$$\frac{dr}{d\tau} = 1 - s_0 e^{-\sigma r} - r.$$

- Assuming  $\sigma r \ll 1$  and developing the exponential term

$$\frac{dr}{d\tau} \approx 1 - s_0 + (s_0 \sigma - 1)(r) + \frac{1}{2} s_0 \sigma^2 r^2 := F(r).$$

# Analytic approximation of the incidence

## ■ Analytic solution

$$r(t) = \frac{1}{\sigma^2 s_0} \left( \sigma s_0 - 1 + \sqrt{-q} \tanh \left( \frac{\sqrt{-q}}{2} t - \varphi \right) \right)$$

where

$$\sqrt{-q} = \sqrt{(\sigma s_0 - 1)^2 + 2s_0 i_0 \sigma^2}$$

and

$$\varphi = \tanh^{-1} \left( \frac{\sigma s_0 - 1}{\sqrt{-q}} \right)$$

## Analytic approximation of the incidence

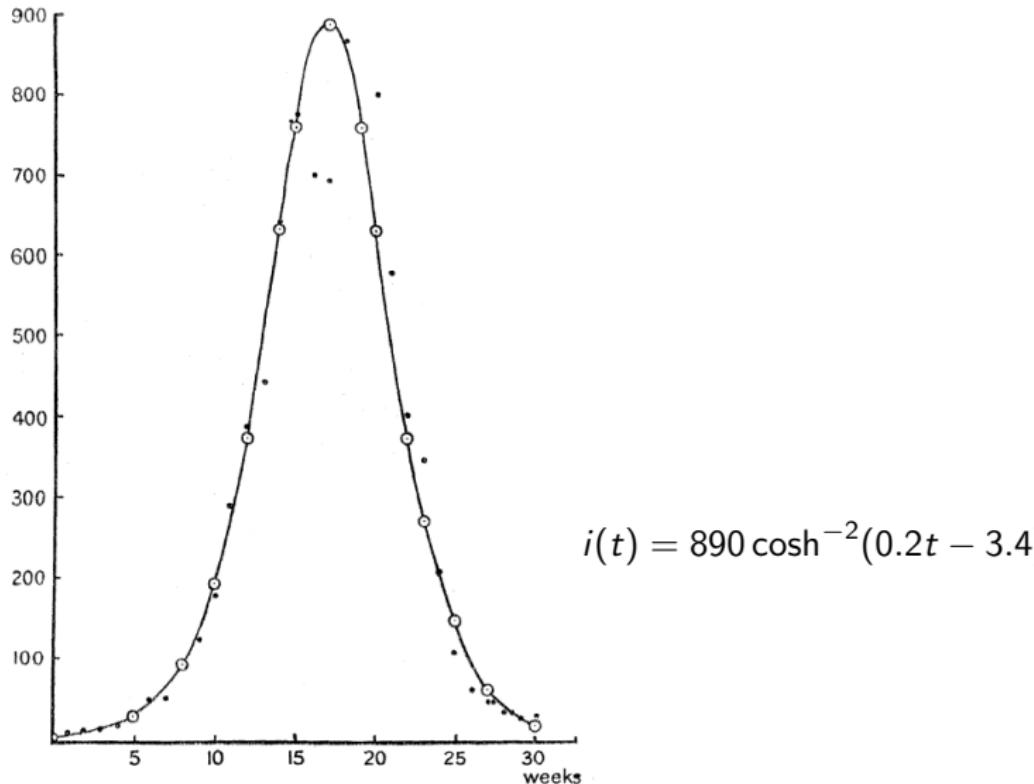
- Finally, since

$$i(t) = \frac{dr}{dt}$$

$$i(t) = \frac{1}{2s_0\sigma^2} \sqrt{-q} \cosh^{-2} \left( \frac{\sqrt{-q}}{2} t - \varphi \right)$$

- Application: Plague-related deaths in Bombay (1905-1906)

## Kermack and McKendrick: theory of epidemics



## Kermack and McKendrick: the threshold phenomenon

- Let's focus now on the second equation:

$$\begin{aligned}\frac{di(t)}{dt} &= \beta si - \nu i, \\ &= (\beta s - \nu)i.\end{aligned}$$

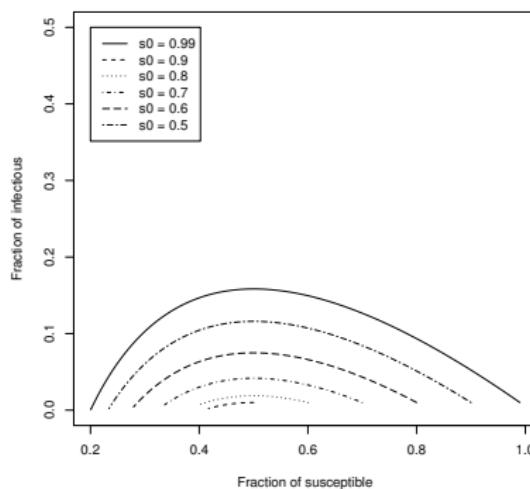
- If the initial fraction of the susceptible population is less than  $\nu/\beta$  then we have  $\frac{di(t)}{dt} < 0$  and the epidemic dies out:

- $\nu/\beta$  is called the **relative removal rate**
- The inverse of the relative removal rate is called **the basic reproduction number**

A pathogen will invade if and only if  $R_0 > 1$ .

## Kermack and McKendrick: phase-plot

$$\beta = 0.2856, \nu = 1/7, R_0 = 2$$



- **Application:** Vaccination can be used to reduce the proportion of susceptible population below  $1/R_0$  for disease eradication.  
This matter will be discussed later.

# Kermack and McKendrick: the threshold phenomenon

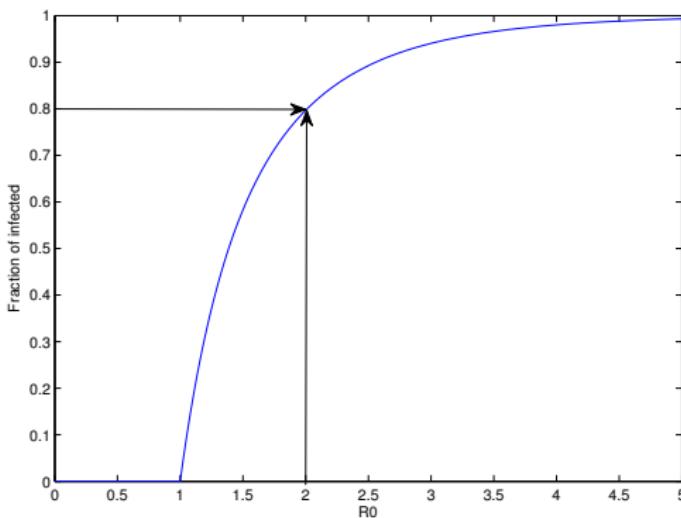
- Having defined  $R_0$ ,
- assuming an entirely susceptible population  $s_0 \approx 1$ , the threshold phenomenon can be re-expressed as:

*A pathogen will invade if and only if  $R_0 > 1$ .*

- Note that in a stochastic model this is not true: '[can invade](#)'

# Kermack and McKendrick: final size and $R_0$

Total fraction of infected population according to  $R_0$



## Some first conclusions

- Assumption:

the time course of the epidemic is sufficiently fast not to be influenced by population demography.

- Conclusion:

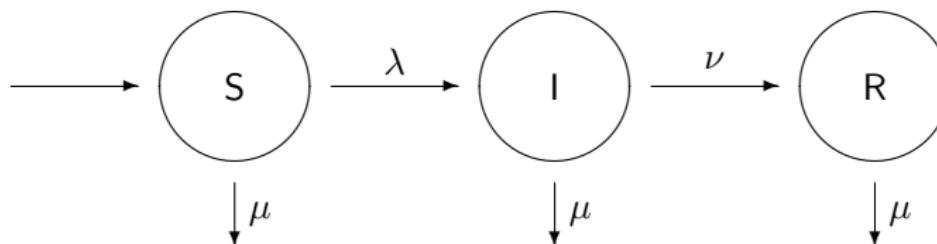
extinction of the pathogen

- Question:

How to explore the long-term persistence  
and endemic dynamics of an infectious disease?

Inclusion of demographic processes

## SIR model with demography: model formulation



$$\begin{aligned}\frac{dS(t)}{dt} &= B - \beta S \frac{I}{N} - \mu S, \\ \frac{dI(t)}{dt} &= \beta S \frac{I}{N} - \nu I - \mu I, \\ \frac{dR(t)}{dt} &= \nu I - \mu R.\end{aligned}$$

## SIR model with demography: model formulation

- Working with constant population:

$$S + I + R = N \Rightarrow B = \mu N$$

- Let  $s, i, r$  denote the proportion of population in the susceptible, infectious and removed classes, respectively.

$$\begin{aligned}\frac{ds(t)}{dt} &= \mu - \beta si - \mu s, \\ \frac{di(t)}{dt} &= \beta si - \nu i - \mu i, \\ \frac{dr(t)}{dt} &= \nu i - \mu r.\end{aligned}$$

# Equilibrium values

- Setting the right-hand side of the system to 0:

$$\begin{aligned}\mu - \beta si - \mu s &= 0, \\ \beta si - \nu i - \mu i &= 0, \\ \nu i - \mu r &= 0.\end{aligned}$$

- Existence of two equilibria:

- The trivial equilibrium (disease free equilibrium):

$$E_0 = (1, 0, 0)$$

- The endemic equilibrium:

$$\begin{aligned}E_1 &= (s_1, i_1, r_1) \\ &= \left( \frac{\mu + \nu}{\beta}, \frac{\mu}{\beta} \left( \frac{\beta}{\mu + \nu} - 1 \right), 1 - s_1 - i_1 \right).\end{aligned}$$

$$E_1 \text{ is positive} \Leftrightarrow \frac{\beta}{\mu + \nu} > 1.$$

# Threshold value

- Focus on the second equation of the system:

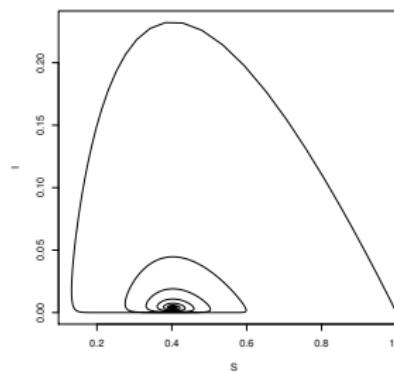
$$\frac{di(t)}{dt} = \beta si - \nu i - \mu i$$

$$\beta = 5, \mu = 1/75, \nu = 2 \Rightarrow R_0 = 2.49$$

- We have

$$\frac{di(t)}{dt} > 0 \Leftrightarrow \beta s - (\nu + \mu) > 0$$

$$\frac{di(t)}{dt} > 0 \Leftrightarrow s > \frac{\nu + \mu}{\beta}$$



# Threshold value: the basic reproduction number

- Intuitively:

- $\beta$ : the transmission rate

*the mean number of new cases produced by an infectious individual per time unit.*

- An infected individual remains infectious for  $1/(\mu + \nu)$  time unit
  - During that time period, a typical infectious individual produces  $\frac{\beta}{\mu + \nu}$  infections in a totally susceptible population. Thus,

$$R_0 = \frac{\beta}{\mu + \nu}.$$

## Threshold value: the basic reproduction number

- If  $R_0 > 1$ , an epidemic occurs and the solution of the system approaches the endemic equilibrium:

$$E_1 = \left( \frac{1}{R_0}, \frac{\mu}{\mu + \nu} \left( 1 - \frac{1}{R_0} \right), 1 - s_1 - i_1 \right).$$

## R code for solving dynamical systems

- R package desolve (Soetaert et al, 2010)  
Solving Initial value differential equations in R
- Calling the library

```
library(deSolve)
```

- Function *ode* in R, general call:

```
ode(y=vector containing state variables,  
     times=vector containing the time units for integration,  
     func=function containing the model equations,  
     parms=vector containing the model parameters)
```

# The SIR model in R

- Definition of the ODE system:
  - Derivatives of state variables are expressed as a list

```
SIR<-function(t,state,parameters)
{
  with(as.list(c(state, parameters)),
  {
    N=S+I+R
    dS = mu*N-beta*I*S/N - mu*S
    dI = beta*I*S/N - nu*I - mu*I
    dR = nu*I -mu*R
    list(c(dS, dI, dR))
  })
}
```

- └ Compartmental Models

- └ Kermack and McKendrick's Model

# The SIR model in R

- Parameter definition/initialisation

```
N=50000  
R0=2; mu0=1/75; nu0=1/(1/24)  
beta0=R0*(mu0+nu0)  
parameters = c(mu=mu0,beta=beta0,nu=nu0)
```

- Initial values

```
state = c(S=N-1,I=1,R=0)
```

- Time range:

from  $T_0 = 0$  to  $T_{\max} = 1000$  by step of 0.01

```
times<-seq(0,1000,by=0.01)
```

- Solution of Initial value differential equations

```
out<-as.data.frame(  
ode(y=state,times=times,func=SIR,parms=parameters)  
)
```

└ Compartmental Models

└ Kermack and McKendrick's Model

# Summary

```
library(deSolve)
N=50000;R0=2; mu0=1/75; nu0=1/(1/24)
beta0=R0*(mu0+nu0)
parameters = c(mu=mu0,beta=beta0,nu=nu0)
state = c(S=N-1,I=1,R=0)
times = seq(0,1000,by=0.01)
SIR<-function(t,state,parameters)
{
  with(as.list(c(state, parameters)),
  {
    N=S+I+R
    dS <- mu*N-beta*I*S/N - mu*S
    dI <- beta*I*S/N - nu*I - mu*I
    dR <- nu*I -mu*R
    list(c(dS, dI, dR))
  })
}
require(deSolve)
out<-as.data.frame(ode(y=state,times=times,func=SIR,parms=parameters))
```

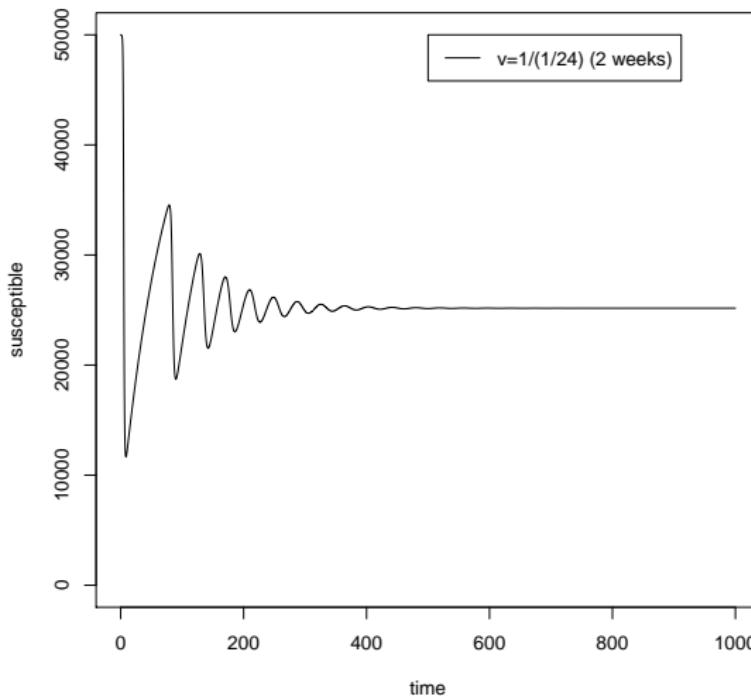
## Number of susceptible individuals

```
tail(out)
```

|        | time    | S        | I        | R        |
|--------|---------|----------|----------|----------|
| 99996  | 999.95  | 25166.63 | 164.4616 | 24668.91 |
| 99997  | 999.96  | 25166.63 | 164.4615 | 24668.91 |
| 99998  | 999.97  | 25166.63 | 164.4615 | 24668.91 |
| 99999  | 999.98  | 25166.63 | 164.4615 | 24668.91 |
| 100000 | 999.99  | 25166.63 | 164.4615 | 24668.91 |
| 100001 | 1000.00 | 25166.63 | 164.4615 | 24668.91 |

- Thus, for  $R_0 = 2$ , the proportion of susceptible individuals at equilibrium is around 50%.

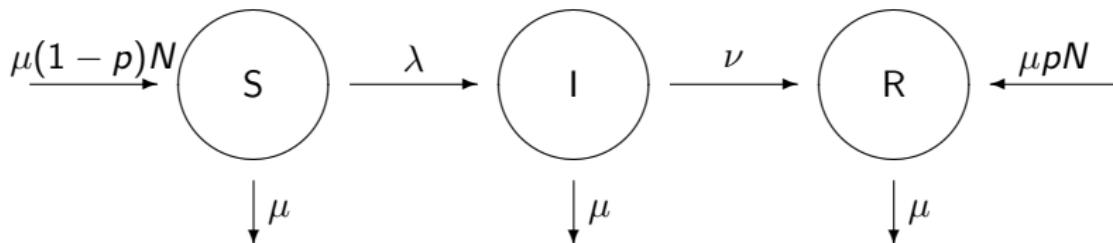
## Number of susceptible individuals



# The basic reproduction number and vaccination programs

- Assumptions
  - SIR model with demography
  - a proportion  $p$  of each cohort of new-born is vaccinated immediately at birth
- Model representation

## Model formulation



$$\begin{aligned}\frac{dS(t)}{dt} &= \mu(1-p)N - \beta S \frac{I}{N} - \mu S, \\ \frac{dI(t)}{dt} &= \beta S \frac{I}{N} - \nu I - \mu I, \\ \frac{dR(t)}{dt} &= \mu p N + \nu I - \mu R.\end{aligned}$$

## Model formulation

- Working with constant population:

$$S + I + R = N \Rightarrow B = \mu N.$$

- Let  $s, i, r$  denote the proportion of population in the susceptible, infectious and removed classes, respectively.

$$\begin{aligned}\frac{ds(t)}{dt} &= \mu(1 - p) - \beta si - \mu s, \\ \frac{di(t)}{dt} &= \beta si - \nu i - \mu i, \\ \frac{dr(t)}{dt} &= \mu p + \nu i - \mu r.\end{aligned}$$

# Equilibrium Values

- Setting the right-hand side of the system to 0:

$$\begin{aligned}\mu(1-p) - \beta si - \mu s &= 0, \\ \beta si - \nu i - \mu i &= 0, \\ \mu p + \nu i - \mu r &= 0.\end{aligned}$$

- Existence of two equilibria
  - The trivial equilibrium (disease free equilibrium):

$$E_0 = (1-p, 0, p)$$

- The endemic equilibrium:

$$\begin{aligned}E_1 &= (s_1, i_1, r_1) \\ &= \left( \frac{\mu + \nu}{\beta}, \frac{\mu}{\beta} ((1-p)R_0 - 1), 1 - s_1 - i_1 \right).\end{aligned}$$

$$E_1 \text{ is positive} \Leftrightarrow (1-p)R_0 > 1$$

## Threshold value

- Vaccination efficacy:

The efficacy of vaccination programs is governed by the relationship between vaccination coverage and the basic reproduction number.

- The proportion of vaccinated new-born population should attain the critical value

$$p_c = 1 - \frac{1}{R_0}$$

for disease eradication. This quantity is also known as the herd immunity threshold

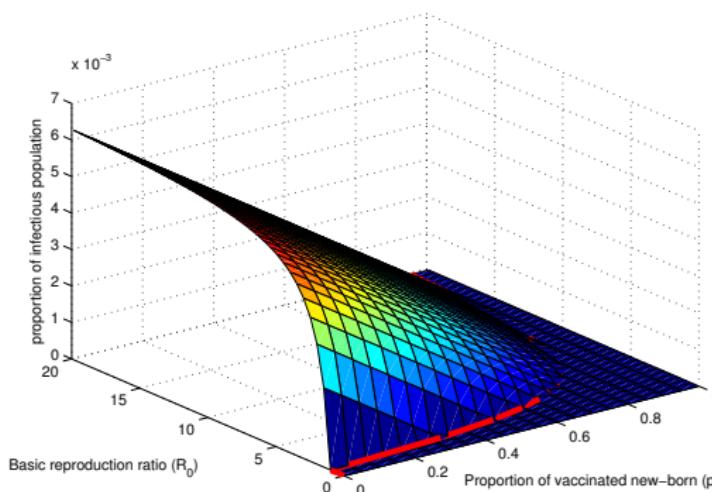
- Herd immunity refers to the indirect protection experienced by unvaccinated individuals resulting from immune individuals in the population

- This leads us to the definition of the effective reproduction number

$$R_e = p_s R_0,$$

where  $p_s$  is the proportion susceptible to infection

## Threshold value



Proportion of infected population regarding different values of  $R_0$  and the proportion of vaccinated population.

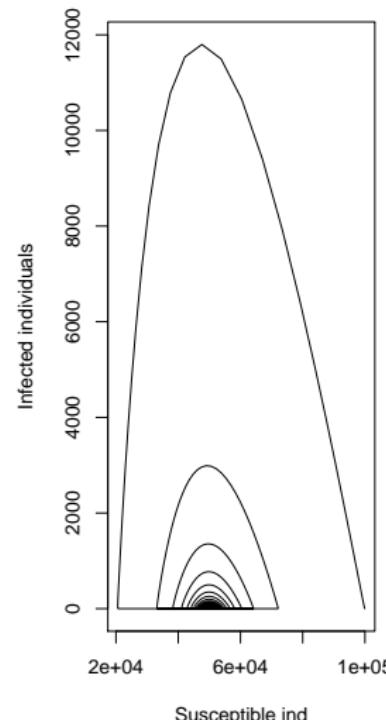
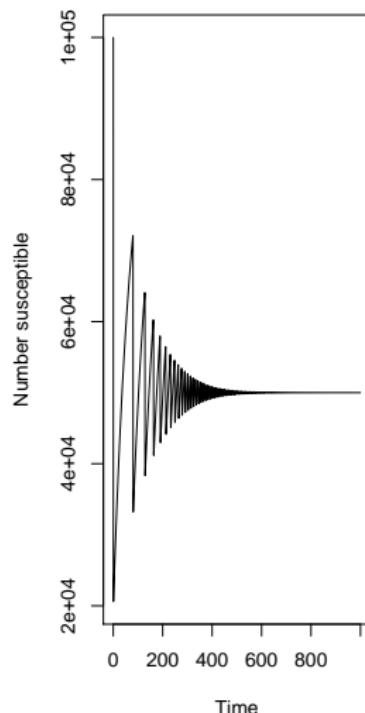
## SIR model with vaccination in R

```
N=50000
R0=2; mu0=1/75; nu0=1/(1/24); p0=0.4
beta0=R0*(mu0+nu0)
parameters = c(mu=mu0,beta=beta0,nu=nu0,p=p0)
state = c(S=N-1,I=1,R=0)
times = seq(0,1000,by=0.01)
SIRv=function(t,state,parameters)
{
  with(as.list(c(state, parameters)),
  {
    N=S+I+R
    dS = mu*(1-p)*N-beta*I*S/N - mu*S
    dI = beta*I*S/N - nu*I - mu*I
    dR = nu*I -mu*R+mu*p*N
    list(c(dS, dI, dR))
  })
}
out = as.data.frame(ode(y=state,times=times,func=SIRv,parms=parameters))
```

└ Compartmental Models

└ Kermack and McKendrick's Model

# Results



## Building your own model

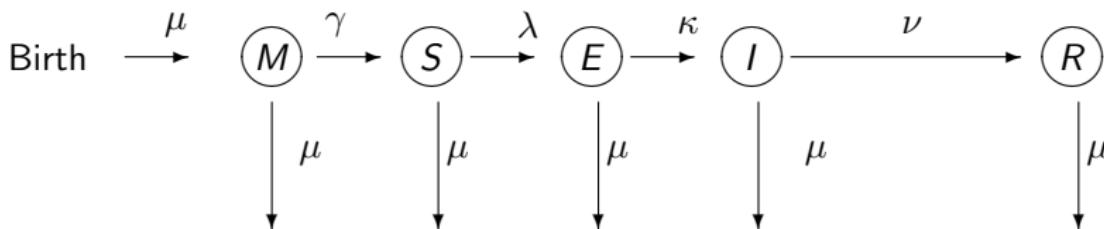
Consider a constant population and an infectious agent with the following characteristics:

- New-born individuals are passively immune (maternal antibodies) for time period  $1/\gamma$  after which they are susceptible to infection
- Force of infection: mass action principle
- Newly infected individuals enter into a latent phase (duration  $1/\kappa$ ): these individuals are infected but not infectious; they were Exposed to the infectious agent.
- Life-long immunity

What does the flow diagram look like?

## Building your own model: an MSEIR example

| State:      | Maternal Susceptible | Latent  | Infected | Immune      |
|-------------|----------------------|---------|----------|-------------|
| Time Scale: | (Months)             | (Years) | (Days)   | (Life long) |



**Figure:** Illustration of the MSEIR model. The individuals enter to the population protected by maternal antibodies after loosing those they move into the susceptible class, after infection to the exposed class (infected but not infectious), after a latent period to the infectious class and after recovering to the immune class.

ODE system?

## MSEIR example: ODE System

$$\begin{aligned}\frac{dM(t)}{dt} &= \mu N - (\gamma + \mu)M \\ \frac{dS(t)}{dt} &= \gamma M - (\lambda + \mu)S \\ \frac{dE(t)}{dt} &= \lambda S - (\mu + \kappa)E \\ \frac{dI(t)}{dt} &= \kappa E - (\mu + \nu)I \\ \frac{dR(t)}{dt} &= \nu I - \mu R.\end{aligned}$$

where  $\lambda = \beta \frac{I}{N}$ .

In this case the basic reproductive number is given by:

$$R_0 = \frac{\kappa\beta}{(\kappa + \mu)(\nu + \mu)}.$$

## MSEIR example

- Basic reproduction number:  $R_0=2$
- Duration of maternal immunity: 3 months
- Duration of the latent period: 2 days
- Duration of the infectious period: 1 week

## MSEIR example

- Basic reproduction number:  $R_0=2$
- Duration of maternal immunity: 3 months
- Duration of the latent period: 2 days
- Duration of the infectious period: 1 week
- Parameters:
  - $\beta = \frac{R_0(\kappa+\mu)(\nu+\mu)}{\kappa}$
  - $\gamma = 1/(3/12); \mu = 1/75; \nu = 52; \kappa = 1/(2/365)$
- Work with numbers:
  - Implementation in R?

## MSEIR R-code

```
N=1E5
mu0=1/75; sigma0=1/(3/12);kappa0=1/(2/365);nu0=1/(7/365);
beta0=R0*(kappa0+mu0)*(nu0+mu0)/kappa0
parameters = c(mu=mu0, sigma=sigma0,kappa=kappa0,nu=nu0,beta=beta0)
state = c(M=1, S=N-2,E=0, I=1,R=0)
times=seq(0,1000,by=0.01)

MSEIR=function(t,state,parameters)
{
  with(as.list(c(state, parameters)),
  {
    N=M+S+E+I+R
    dM= mu*N-(mu+sigma)*M
    dS = sigma*M-beta*I*S/N - mu*S
    dE = beta*I*S/N-(kappa+mu)*E
    dI = kappa*E - nu*I - mu*I
    dR = nu*I -mu*R
    list(c(dM, dS, dE, dI, dR))
  })
}
```

- └ Compartmental Models

- └ Building your own model

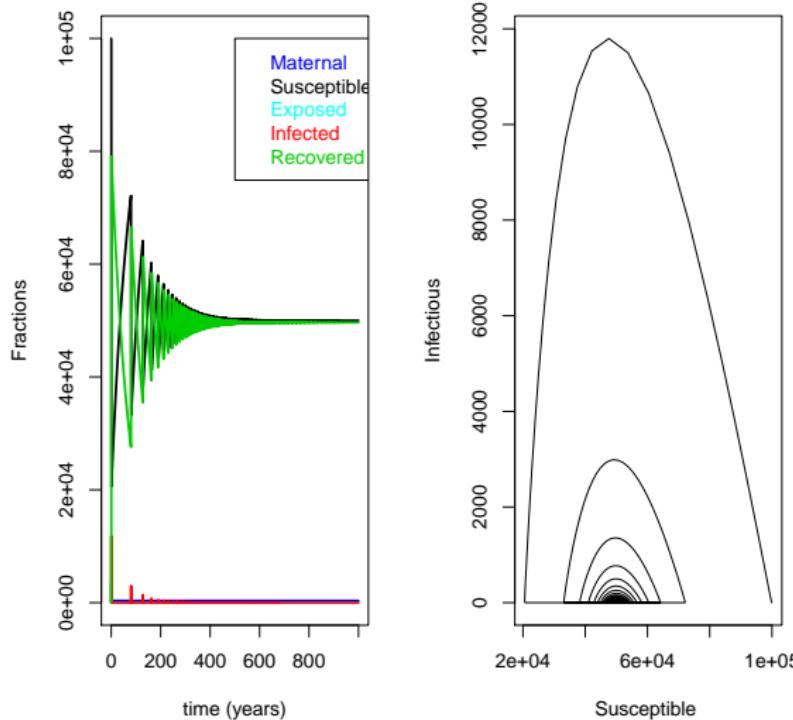
# Summary

```
out=as.data.frame(ode(y=state,times=times,func=MSEIR,parms=parameters))

par(mfrow=c(1,2))
##### Time series plot
plot(range(times),c(0,N),xlab="time (years)",
      ylab="Fractions",type="n",ylim=c(0,N))
lines(out$time,out$M,lwd=2,col=4)
lines(out$time,out$S,lwd=2,col=1)
lines(out$time,out$E,lwd=2,col=5)
lines(out$time,out$I,lwd=2,col=2)
lines(out$time,out$R,lwd=2,col=3)
legend(mean(times),N,c("Maternal","Susceptible","Exposed",
                      "Infected","Recovered"),text.col=c(4,1,5,2:3))

##### Phase plan
plot(out$S,out$I ,type="l",main=" ", xlab="Susceptible ", ylab="Infecti
```

# The MSEIR model results



## Class exercise

- Implement [vaccinating children](#) in the MSEIR model
- Code the model using [proportions](#) rather than numbers
- Code an extension of the model in which [recovered individuals become susceptible again](#): MSEIRS
- What would happen if one switches to a [resolution of days](#) for the time scale?

## Transmission within multiple subpopulations

The law of mass action as stated before relies on the hypothesis that infected and susceptible individuals mix homogeneously.

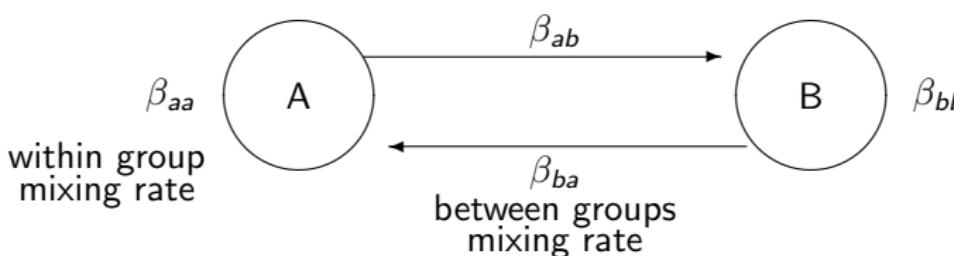
$\beta$  is thus considered as age and time independent.

- Objective Taking into account the social structures in the population.
- Assumptions
  - The population is constructed from multiple subpopulations.
  - the transmission process is governed by the mixing patterns in the population

# Who Acquires Infection From Whom?

The transmission process is governed by the mixing pattern commonly represented by the WAIFW 'Who Acquires Infection From Whom' matrix.

## ■ A first example



**Figure:** Mixing patterns between two subpopulations:  $\beta_{aa}$  is the mixing rate within group A,  $\beta_{bb}$  is the mixing rate within group B.  $\beta_{ab}$  and  $\beta_{ba}$  are the mixing rates between the two groups.

## The WAIFW matrix:

$$C = \begin{pmatrix} \beta_{aa} & \beta_{ab} \\ \beta_{ba} & \beta_{bb} \end{pmatrix}.$$

## Application 1: The SIS transmission model for Gonorrhea

- Gonorrhea is a disease transmitted through the population by heterosexual contacts.
- Immunity to reinfection does not exist.
- The two subpopulations, males and females, interact according to the following mixing matrix

$$C = \begin{pmatrix} 0 & \beta_{fm} \\ \beta_{mf} & 0 \end{pmatrix}.$$

# Application 1: The SIS transmission model for Gonorrhea

- The SIS transmission model for Gonorrhea (Capasso, 2008):

$$\begin{aligned}\frac{dS_f}{dt} &= -\beta_{fm} S_f(t) I_m(t) + \nu_f I_f(t) \\ \frac{dI_f}{dt} &= \beta_{fm} S_f(t) I_m(t) - \nu_f I_f(t) \\ \frac{dS_m}{dt} &= -\beta_{mf} S_m(t) I_f(t) + \nu_m I_m(t) \\ \frac{dI_m}{dt} &= \beta_{mf} S_m(t) I_f(t) - \nu_m I_m(t).\end{aligned}$$

- Assuming  $N_f = S_f + I_f$  and  $N_m = S_m + I_m$

$$\begin{aligned}\frac{dI_f}{dt} &= \beta_{fm}(N_f - I_f(t)) I_m(t) - \nu_f I_f(t), \\ \frac{dI_m}{dt} &= \beta_{mf}(N_m - I_m(t)) I_f(t) - \nu_m I_m(t).\end{aligned}$$

# SIS model for Gonorrhea: endemic equilibrium

- System:

$$\begin{aligned}\beta_{fm}(N_f - I_f(t))I_m(t) - \nu_f I_f(t) &= 0, \\ \beta_{mf}(N_m - I_m(t))I_f(t) - \nu_m I_m(t) &= 0.\end{aligned}$$

- After some calculations, and noting the **relative removal rates as**  
 $\rho_f = \nu_f / \beta_f, \rho_m = \nu_m / \beta_m$

$$I_f(\infty) = \frac{N_f N_m - \rho_f \rho_m}{\rho_f + N + N_m} \text{ and } I_m(\infty) = \frac{N_f N_m - \rho_f \rho_m}{\rho_m + N + N_f}.$$

- This equilibrium is positive  $\Leftrightarrow N_f N_m - \rho_f \rho_m > 0$  or

$$\left( \frac{N_f \beta_f}{\nu_f} \times \frac{N_m \beta_m}{\nu_m} \right) = R_{0,f} \times R_{0,m} > 1.$$

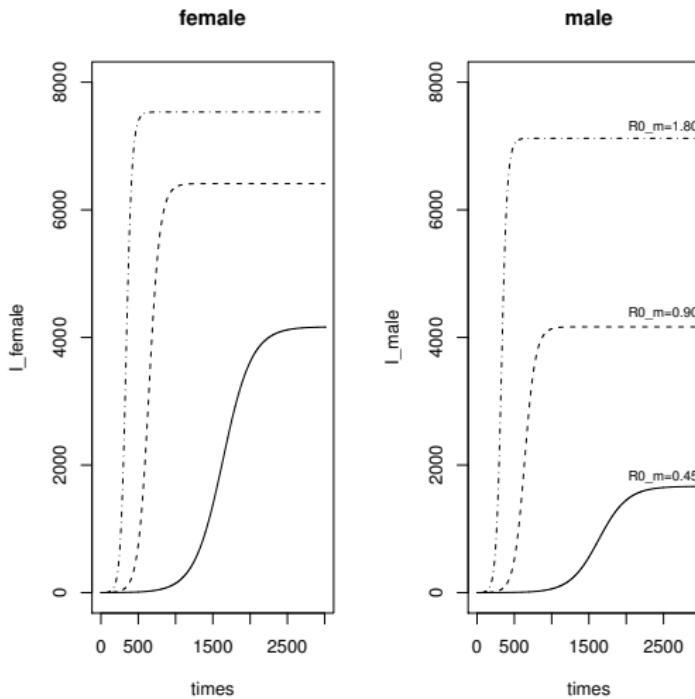
# SIS Model for Gonorrhea in R

```
parameters <- c(beta1=0.000003,beta2=0.000006,v1=0.007,v2=0.05,N1=10000,N2=15000)
state <- c(IF=1,IM=0)
Gonorrhea<-function(t,state,parameters){
with(as.list(c(state, parameters)),{
dIF <- beta1*(N1-IF)*IM-v1*IF
dIM <- beta2*(N2-IM)*IF-v2*IM
list(c(dIF,dIM))})
}
times<-seq(0,3000,by=0.1)

beta2.i<-0.000006*c(0.25,0.5,1)
beta2.i

y1mat<-y2mat<-matrix(0,length(times),length(beta2.i))
for(i in 1:length(beta2.i)){
parameters <- c(beta1=0.000003,beta2=beta2.i[i],v1=0.007,v2=0.05,N1=10000,N2=15000)
require(deSolve)
out <- as.data.frame(ode(y=state,times=times,func=Gonorrhea,parms=parameters))
y1mat[,i]<-out$IF
y2mat[,i]<-out$IM}
```

## SIS Model for Gonorrhea: results



## Application 2: SIR - 2 interacting subpopulations

- Model structure: Conversely to gonorrhea model, the out-diagonal elements of the WAIFW matrix are positive

$$C = \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix}.$$

representing contacts within and across subpopulations.

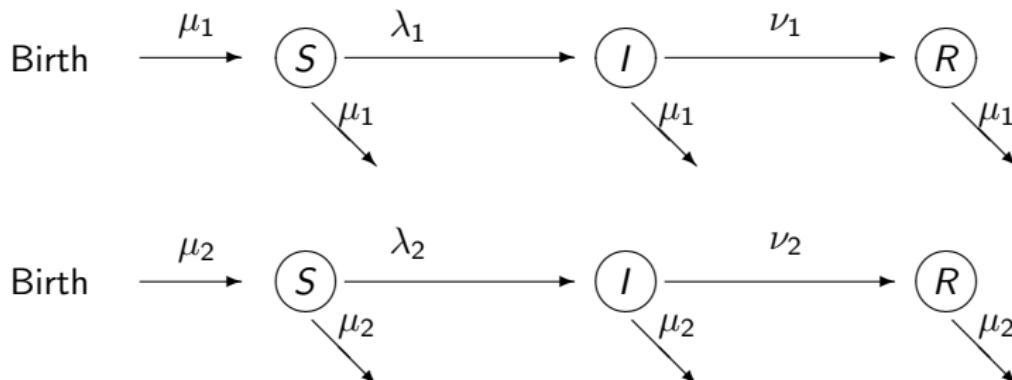


Figure: Illustration of a SIR model with two subpopulations.

## Application 2: SIR - 2 interacting subpopulations

- Force of infection

$$\left(\sum_{j=1}^2 \beta_{ij} I_j\right) S_i = \lambda_i S_i.$$

- ODE system

$$\frac{dS_i(t)}{dt} = - \left(\sum_{j=1}^2 \beta_{ij} I_j\right) S_i + N_i \mu_i - \mu_i S_i$$

$$\frac{dI_i(t)}{dt} = \left(\sum_{j=1}^2 \beta_{ij} I_j\right) S_i - (\mu_i + \nu_i) I_i$$

$$\frac{dR_i(t)}{dt} = \nu_i I_i - \mu_i R_i.$$

- We will now consider the symmetric case setting  $\beta_{12} = \beta_{21} = \alpha$ .

└ Compartmental Models

└ Who Acquires Infection From Whom?

## Application 2: implementation in R

```
alpha<-0
parameters <- c(beta11=0.05,beta12=alpha,beta21=alpha,beta22=0.05,
                  nu1=1/30,nu2=1/30,mu=0.001)
state <- c(S1=0.8,I1=0.2,R1=0,S2=0.8,I2=0.2,R2=0)
times<-seq(0,10000,by=0.01)

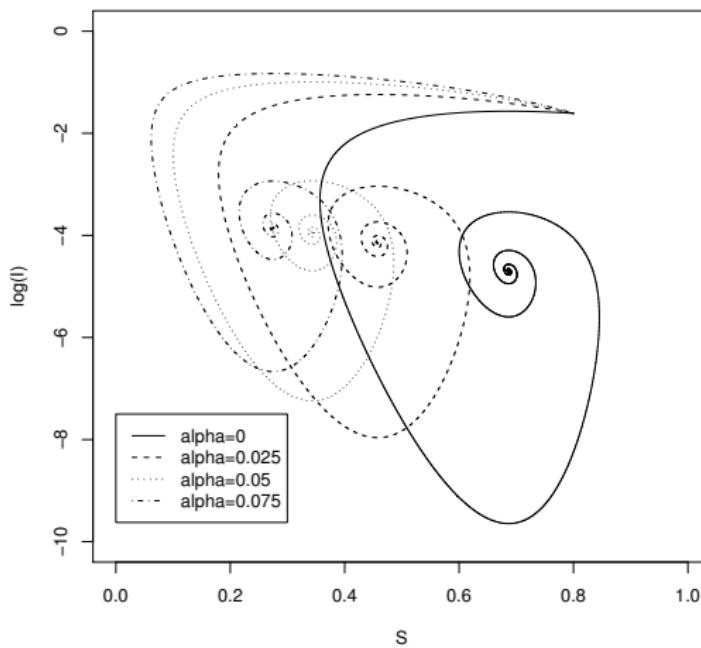
SIRtwo<-function(t,state,parameters)
{
  with(as.list(c(state, parameters)),
  {
    dS1 <- -(beta11*I1+beta12*I2)*S1+mu-mu*S1
    dI1 <- (beta11*I1+beta12*I2)*S1-nu1*I1-mu*I1
    dR1 <- nu1*I1 - mu*R1
    dS2 <- -(beta21*I1+beta22*I2)*S2+mu-mu*S2
    dI2 <- (beta21*I1+beta22*I2)*S2-nu2*I2-mu*I2
    dR2 <- nu2*I2 - mu*R2
    list(c(dS1,dI1,dR1,dS2,dI2,dR2))
  })
}
```

## Application 2: implementation in R

```
times<-seq(0,10000,by=0.01)
require(deSolve)
out <- as.data.frame(ode(y=state,times=times,func=SIRtwo,
                           parms=parameters))

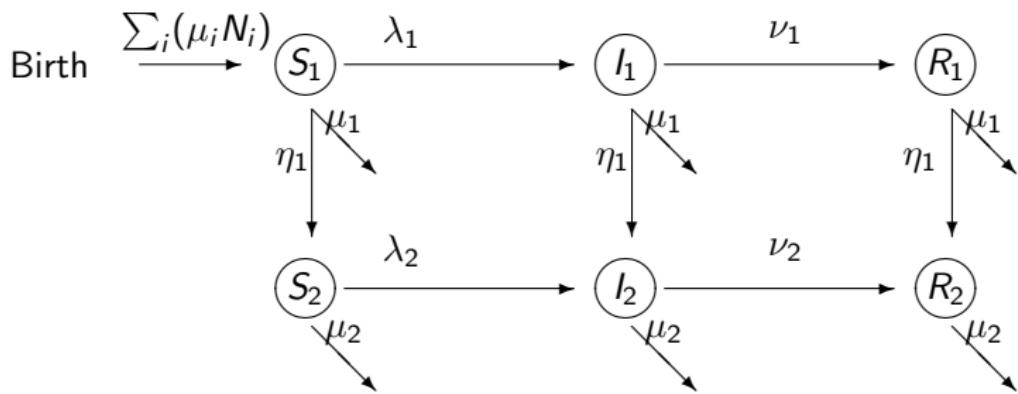
par(mfrow=c(1,1))
plot(out$S1,log(out$I1),type="l",main=" ",
      xlab="S",ylab="log(I)",ylim=c(-10,0),xlim=c(0,1))
legend(0,-7.5,c("alpha=0"),lty=c(1:4))
```

## Application 2: results



# Transmission over age and time in the SIR model

- The model with interacting subpopulations can be seen as an age-time dependent SIR model or age-structured SIR model
- The population is divided into a finite number of age groups interacting with each other



**Figure:** Illustration of age structured SIR model with two age groups.

# Transmission over age and time in the SIR model

- A general approach: partial differential equations
- Each differential equation represent the change (over time and age) in the compartment.

$$\left\{ \begin{array}{l} \frac{\partial S(a,t)}{\partial a} + \frac{\partial S(a,t)}{\partial t} = -(\lambda(a,t) + \mu(a))S(a,t), \\ \frac{\partial I(a,t)}{\partial a} + \frac{\partial I(a,t)}{\partial t} = \lambda(a,t)S(a,t) - (\nu + \alpha + \mu(a))I(a,t), \\ \frac{\partial R(a,t)}{\partial a} + \frac{\partial R(a,t)}{\partial t} = \nu I(a,t) - \mu(a)R(a,t), \end{array} \right.$$

where  $N(a, t) = S(a, t) + I(a, t) + R(a, t)$  and  $S(0, T) = B(t)$ , the number of births all susceptible to infection and  $\lambda(a, t)$  is given by an integral equation with, in general, no explicit solution

- Quite complicated to solve

## The Cohort Age Structured model

- For a population with  $K$  age groups, the system of ordinary differential equations for the first age group ( $i = 1$ ) in an age-structured SIR model is given by

$$\frac{dS_1(t)}{dt} = - \left( \sum_{j=1}^K \beta_{1j} I_j \right) S_1 + \sum_i (\mu_i N_i) - \mu_1 S_1 - \eta_1 S_1,$$

$$\frac{dI_1(t)}{dt} = \left( \sum_{j=1}^K \beta_{1j} I_j \right) S_1 - \nu_1 I_1 - \eta_1 I_1,$$

$$\frac{dR_1(t)}{dt} = \nu_1 I_1 - \mu_1 R_1 - \eta_1 R_1.$$

## The Cohort Age Structured model

- Denoting  $\eta_i$  is the rate at which individuals of age-class  $i$  pass from  $S_i, I_i, R_i$  to  $S_{i+1}, I_{i+1}, R_{i+1}$ , we have for  $i = 2$  to  $K$

$$\begin{aligned}\frac{dS_i(t)}{dt} &= - \left( \sum_{j=1}^K \beta_{ij} I_j \right) S_i + \eta_{i-1} S_{i-1} - \mu_i S_i - \eta_i S_i, \\ \frac{dI_i(t)}{dt} &= \left( \sum_{j=1}^K \beta_{ij} I_j \right) S_i + \eta_{i-1} I_{i-1} - (\nu_i + \mu_i) I_i - \eta_i I_i, \\ \frac{dR_i(t)}{dt} &= \nu_i I_i - \mu_i R_i + \eta_{i-1} R_{i-1} - \eta_i R_i.\end{aligned}$$

- This model is a cohort age-structured model (CAS-model) with the disadvantage that people can instantaneously grow older.

## The Realistic Age Structured model

- A better alternative to the CAS-model is the RAS-model: Realistic Age-structured Model
- In the RAS-model the SIR model is applied to each age-class of one year and after one year people instantaneously move to the next age-class
- The model resembles the educational system where children move to another class after one year
- The RAS-model consists of the following two-step iteration:  
Assuming one-year age-groups, let  $\{S_i(t), I_i(t), R_i(t)\}$  denote the number of susceptible, infected and recovered individuals of age  $i = 0, \dots, K - 1$  at time  $t$  (in years).

# The Realistic Age Structured model

**Step 1:** Given initial values

$\{S_i(t), I_i(t), R_i(t)\} = \{S_i(t_0), I_i(t_0), R_i(t_0)\}, i = 0, \dots, K - 1$  we solve the following set of ODEs:

$$\begin{cases} \frac{dS_i(t)}{dt} &= -(\beta I_i(t) + \mu_i) S_i(t), \\ \frac{dI_i(t)}{dt} &= \beta I_i(t) S_i(t) - (\nu + \mu_i) I_i(t), \\ \frac{dR_i(t)}{dt} &= \nu I_i(t) - \mu_i R_i(t), \end{cases}$$

to obtain  $\{S_i(t + 1), I_i(t + 1), R_i(t + 1)\}, i = 0, \dots, K - 1$  after one year.

# The Realistic Age Structured model

**Step 2:** Individuals are then shifted by one year:

- $\{S_i(t+1), I_i(t+1), R_i(t+1)\} \rightarrow \{S_{i+1}(t+1), I_{i+1}(t+1), R_{i+1}(t+1)\}, i = 0, \dots, K-2$
- all newborns  $B$  are assumed susceptible to infection:  
 $\{S_0(t+1), I_0(t+1), R_0(t+1)\} = \{B, 0, 0\}.$

This process is iterated throughout the time period of interest.

# Two examples (CAS and RAS)

Kovac et al. *BMC Bioinformatics* (2018) 19:101  
<https://doi.org/10.1186/s12859-018-2108-3>

BMC Bioinformatics

METHODOLOGY ARTICLE

Open Access



## Heterogeneous computing for epidemiological model fitting and simulation

Thomas Kovac<sup>1,2\*</sup> , Tom Haber<sup>2†</sup>, Frank Van Reeth<sup>2</sup> and Niel Hens<sup>1,3</sup>



Contents lists available at ScienceDirect

Epidemics

journal homepage: [www.elsevier.com/locate/epidemica](http://www.elsevier.com/locate/epidemica)



Estimating dynamic transmission model parameters for seasonal influenza by fitting to age and season-specific influenza-like illness incidence



Nele Goeyvaerts<sup>a,b,\*</sup>, Lander Willem<sup>a,b,c,†</sup>, Kim Van Kerckhove<sup>a,b</sup>,  
Yannick Vandendriessche<sup>a</sup>, Germaine Hanquet<sup>d</sup>, Philippe Beutels<sup>b</sup>, Niel Hens<sup>a,b</sup>

<sup>a</sup> Interuniversity Institute for Biostatistics and statistical Bioinformatics, Hasselt University, Agoralaan Gebouw D, B3590 Diepenbeek, Belgium

<sup>b</sup> Centre for Health Economics Research and Modelling Infectious Diseases, Vaccine & Infectious Disease Institute, University of Antwerp, Universiteitsplein 1, B2610 Wilrijk, Belgium

<sup>c</sup> Department of Mathematics and Computer Science, University of Antwerp, Middelheimlaan 1, B2020 Antwerp, Belgium

<sup>d</sup> KCE – Belgian Health Care Knowledge Centre, Boulevard du Jardin Botanique 55, B1000 Brussels, Belgium

Goeyvaerts et al. (2015); Kovac et al. (2018)

# Summary

- We have introduced the basic building blocks of mathematical models of infectious diseases
- Models are a simplification of reality and model assumptions are crucial in the development of mathematical models
- However, several questions still remain:
  - How do we inform models using data?
  - How can models be compared?
  - How can models be fitted to data? ...

## Part II

Estimating Infectious Disease  
Parameters: the force of infection

# Overview

## 5 Introduction

- The SIR model
- The force of infection
- Is homogeneous mixing realistic?
- Estimating  $R_0$  from incidence data
- Endemic equilibrium & the static model

## 6 Serological Data

- Data examples
- Current status data

## 7 Estimating the force of infection

- Modelling current status data
- The early work
- The outbreak in the 90's
- The issue of monotonicity

# Rules of engagement

| Entity                    | Symbol 1            | Symbol 2              |
|---------------------------|---------------------|-----------------------|
| maternal immune           | $M$ (number)        | $m_{pi}$ (proportion) |
| susceptibles              | $S$ (number)        | $s$ (proportion)      |
| infected                  | $I$ (number)        | $i$ (proportion)      |
| recovered                 | $R$ (number)        | $r$ (proportion)      |
| total population size     | $N$ (number)        | -                     |
| age-stratified pop. size  | $N(a)$ (number)     | -                     |
| life expectancy           | $L$ (dec. number)   | -                     |
| survivor function         | -                   | $m$ (proportion)      |
| recovery rate             | $\nu$ (rate)        | -                     |
| force of infection        | $\lambda$ (rate)    | -                     |
| transmission parameter    | $\beta$ (rate)      | -                     |
| basic reproduction number | $R_0$ (dec. number) | -                     |

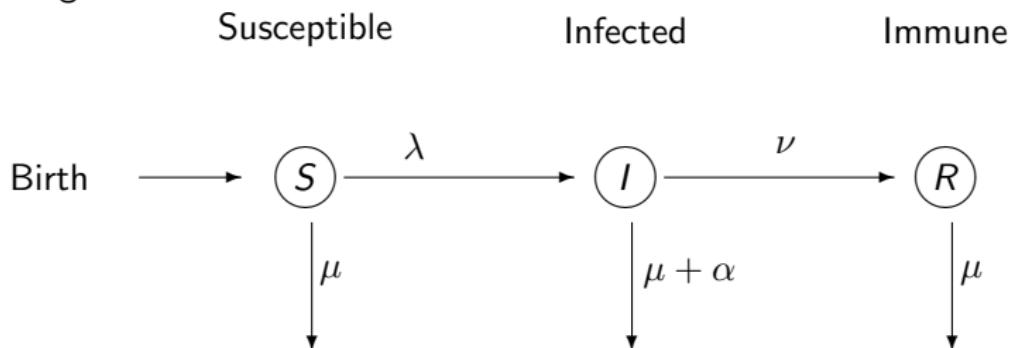
Table: Glossary of the most important symbols.

## The SIR model

- It is assumed that for 'simple' infectious diseases the individuals in the population can be classified, according to their infection status, into three states:
  - Susceptible to infection: individuals who have not been exposed yet - the population at risk
  - Infected and infectious to others: individuals who have been infected
  - Immune to reinfection: individuals that recovered from infection and are immune
- Simple but effective model
  - influenza (one season), measles, rubella, ...

# The SIR model

- Let us recall the basic SIR model
- Flow diagram:



- parameters:
  - $\mu$ : the death rate
  - $\lambda$ : the force of infection
  - $\alpha$ : the disease-related mortality rate
  - $\nu$ : the recovery rate

## The SIR model

- Compartments and time scales.
- Each differential equation represent the change (over time and age) in the compartment.

$$\begin{cases} \frac{\partial S(a,t)}{\partial a} + \frac{\partial S(a,t)}{\partial t} = -(\lambda(a,t) + \mu(a))S(a,t), \\ \frac{\partial I(a,t)}{\partial a} + \frac{\partial I(a,t)}{\partial t} = \lambda(a,t)S(a,t) - (\nu + \alpha + \mu(a))I(a,t), \\ \frac{\partial R(a,t)}{\partial a} + \frac{\partial R(a,t)}{\partial t} = \nu I(a,t) - \mu(a)R(a,t), \end{cases}$$

where  $N(a, t) = S(a, t) + I(a, t) + R(a, t)$  and  $S(0, T) = B(t)$ , the number of births all susceptible to infection.

## The SIR assumptions

SIR-assumptions:

- All newborns enter directly into the susceptible class
  - the model ignores the maternal antibodies period
- The infection and the infectious period occur simultaneously
  - the SIR model ignores the latent period (the period in which the individual is infected but not yet infectious to others).

## Duration in the SIR model

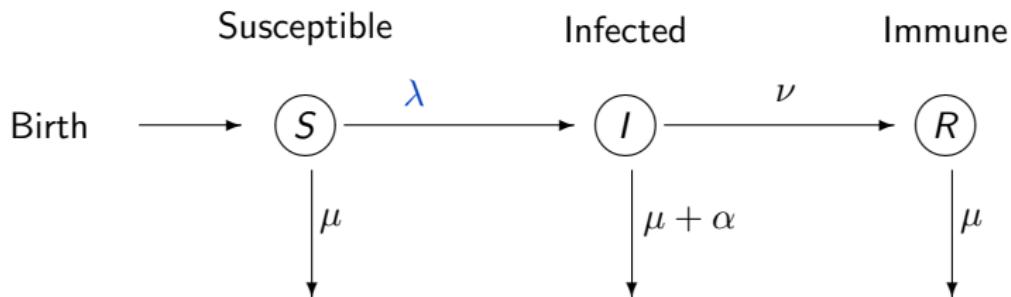
- The duration within each class varies from class to class.
  - e.g. measles (in developed countries):
    - mean age of exposure is approximately 4-5 years,
    - infectious period is around 7 days,
    - immunity is life long.
- These lengths of time hold for most of childhood infectious diseases: the susceptible period lasts years, the infectious period days and immunity is assumed to be life long.

└ Introduction

└ The force of infection

## The force of infection & mass action

- The rate at which susceptibles become infected.
- The rate at which individuals leave the susceptible class.



└ Introduction

└ The force of infection

# The force of infection & mass action

## ■ Mapping of individuals

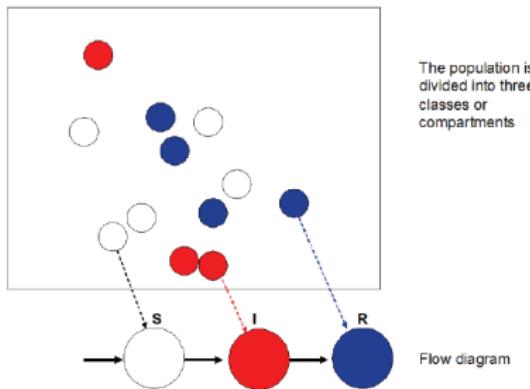
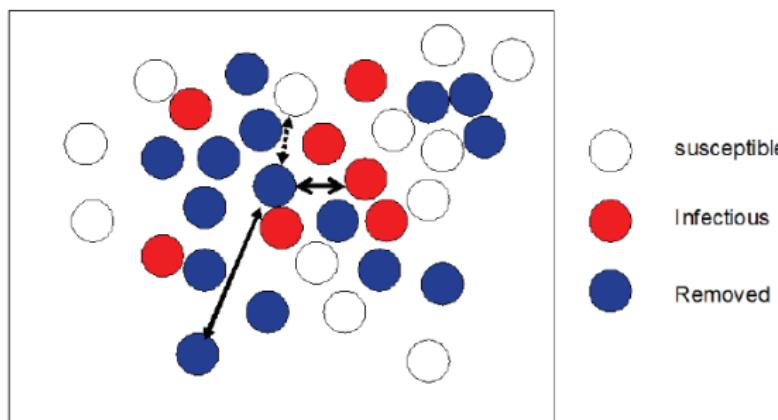


Figure: The different compartments of the SIR model.

## The force of infection & mass action

- New infections are the result of susceptibles meeting infectious individuals:



**Susceptible** **Infectious** **Removed**

Figure: The mass action principle.

## The force of infection & mass action

- A central characteristic of the population dynamics of infection diseases is the transmission of infection from the infected (& infectious) class to the susceptible class:  $\beta$  = transmission rate.
- The **pseudo** or density-dependent **mass-action** principle states:  
new cases of infection (the incidence) =  $\beta I(t)S(t)$ .
- The force of infection given by (Hamer, 1906 - Measles):

$$\lambda(t) = \beta I(t),$$

does not depend on the population size:  $\beta = c_e/N = \text{constant}$ , where  $c_e$  is the effective contact rate:  $c_e \propto N$ .

- Often used for animal & plant infections

## The force of infection & mass action

- The **true** or frequency-dependent **mass-action principle** states:  
new cases of infection (the incidence) =  $\beta(t)I(t)S(t)/N(t)$ .  
where  $\beta(t) = c_e(t)$  - with  $c_e(t)/N(t) = \text{constant}$  - depends on the population size.
- The force of infection is given by

$$\lambda(t) = \beta I(t)/N(t)$$

- Most often used assumption for human infections.
- Both frequency- and density-dependent mass action are the same in a constant population.

## The force of infection & mass action

- In the case of homogeneous mixing:

$$R_0 = \beta ND,$$

where  $N$  is the total population size and  $D$  is the mean infectious period.

- $\beta$  is often recalculated based on  $R_0$  when transferring populations. This corresponds to the assumption of frequency-dependent mass-action.

## The force of infection & mass action

- The basic reproduction number  $R_0$  expresses the average number of secondary cases produced by a typical infectious individual during his/her entire infectious period when introduced into a completely susceptible population.

| Infectious Disease | $R_0$ |
|--------------------|-------|
| measles            | 10-20 |
| chickenpox         | 5-10  |
| mumps              | 5-10  |
| rubella            | 4-7   |
| smallpox           | 3-5   |

Farrington (*Modelling Epidemics*, 2003)

└ Introduction

  └ Is homogeneous mixing realistic?

## Is homogeneous mixing realistic?

- The underlying assumption of the mass-action principle, as stated here, is that the **infectious and susceptible individuals mix in a homogeneous way**, i.e., that  $\beta$  is age and time independent.
- In endemic equilibrium  $I(t) \equiv I = \text{constant}$  and thus the **force of infection is constant**.
- In general this is not true, however:
  - it is a reasonable assumption to start from for outbreaks
  - for non-vaccinated childhood infections (e.g. parvovirus B19, VZV, etc.) the assumption of endemic equilibrium is tenable

└ Introduction

└ Estimating  $R_0$  from incidence data

## Estimating $R_0$ from incidence data

- $R_0$  determines the epidemic potential of the pathogen
- $R$  ( $R_e$ ) determines the effective epidemic spread potential
- The formulas used here assume homogeneous mixing and are therefore approximate
- However they are useful!
- The alternative is to develop a model for the pathogen at hand and estimate its parameters

# The growth rate

- Assume we are in the early stage of an epidemic:

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

where  $\Lambda$  is called the growth rate of the epidemic

- If we take the logarithm of both sides:

$$\log(I(t)) = \log(I(0)) + \Lambda t.$$

- So how can we estimate  $\Lambda$ ?

## $R_0$ and growth rates

- In case the pre-infectious period is short in comparison with the infectious period and the infectious period is assumed to follow an exponential distribution:

$$R_0 = 1 + \Lambda D,$$

where  $D$  is the average infectious period.

- In case the pre-infectious period and the infectious period both follow an exponential distribution:

$$R_0 = (1 + \Lambda D)(1 + \Lambda D'),$$

where  $D'$  and  $D$  are the average pre-infectious and infectious period, respectively.

└ Introduction

└ Estimating  $R_0$  from incidence data

## $R_0$ and growth rates

- In case the pre-infectious and infectious periods are unknown but assumed to follow an exponential distribution but the serial interval is known:

$$R_0 = 1 + \Lambda T_s$$

- Given  $R_0 = 1 + \Lambda D$ , and assumptions, one can show:

$$R_0 = 1 + \frac{\log(2)}{T_d} D,$$

where  $T_d$  is the doubling time

- A more general expression:

$$R_0 = \frac{\Lambda D \left( \frac{\Lambda D'}{m} + 1 \right)^m}{\left( 1 - \left( \frac{\Lambda D}{n} + 1 \right)^{-n} \right)}$$

where  $m$  and  $n$  reflect the tightness of the gamma distribution for the pre-infectious and the infectious period, respectively.

└ Introduction

└ Estimating  $R_0$  from incidence data

## $R_0$ and final size

- Assume we have the final size of our epidemic. If  $R_0 < 3$  we can estimate it using the following methods:
  - Denote  $z_f$  the proportion of the population which has been infected by the end of the epidemic and assume that all individuals are susceptible at the start of the epidemic:

$$R_0 = -\frac{\log(1 - z_f)}{z_f}$$

- Assume that  $s_0$  and  $s_f$  are the proportions of the population susceptible at the start and at the end of the epidemic, respectively:

$$R_0 = \frac{\log(s_f) - \log(s_0)}{s_f - s_0}$$

- Denote  $N$  the population size and  $C$  the number of cases:

$$R_0 = \frac{N - 1}{C} \log \left( \frac{s_0 + 0.5}{s_f - 0.5} \right)$$

## Class exercise

- If  $R_0 = 2$  and our pre-infectious and infectious period are exponential and both, on average, 2 days long. What is the growth rate of the epidemic?
- What happens in the previous situation when the pre-infectious period is very short? What is the doubling time in this situation?
- Draw a graph of the relationship of  $R_0$  and the final size

└ Introduction

└ Estimating  $R_0$  from incidence data

# Some illustrations



RESEARCH ARTICLE

## Spatiotemporal Evolution of Ebola Virus Disease at Sub-National Level during the 2014 West Africa Epidemic: Model Scrutiny and Data Meagreness

Eva Santermans<sup>1\*</sup>, Emmanuel Robesyn<sup>2</sup>, Tapiwa Ganyani<sup>1</sup>, Bertrand Sudre<sup>2</sup>, Christel Faes<sup>1</sup>, Chantal Quinten<sup>2</sup>, Wim Van Bortel<sup>2</sup>, Tom Haber<sup>3</sup>, Thomas Kovac<sup>1,3</sup>, Frank Van Reeth<sup>3</sup>, Marco Testa<sup>2,4</sup>, Niel Hens<sup>1,5</sup>, Diamantis Plachouras<sup>2</sup>

**1** Interuniversity Institute for Biostatistics and Statistical Bioinformatics, Hasselt University, Diepenbeek, Belgium, **2** European Centre for Disease Prevention and Control, Stockholm, Sweden, **3** Expertise centre for Digital Media, iMinds, iJL, Diepenbeek, Belgium, **4** Department of Public Health, University of Turin, Turin, Italy, **5** Centre for Health Economics Research and Modelling Infectious Diseases, Vaccine & Infectious Disease Institute, University of Antwerp, Antwerp, Belgium



On the estimation of the basic reproduction number using non-linear mass action principle based epidemic models

Tapiwa Ganyani<sup>1</sup>, Christel Faes<sup>1</sup>, Gerardo Chowell<sup>2,3</sup>, Niel Hens<sup>1,4</sup>

Santermans et al. (2016)

└ Introduction

└ Estimating  $R_0$  from incidence data

## Homogeneous mixing

- Niels Becker: 'It is not so important which specific model you use to estimate  $R_0$ . It is however important to use the appropriate method when studying intervention strategies'

└ Introduction

└ Estimating  $R_0$  from incidence data

# Overview

## 5 Introduction

- The SIR model
- The force of infection
- Is homogeneous mixing realistic?
- Estimating  $R_0$  from incidence data
- Endemic equilibrium & the static model

## 6 Serological Data

- Data examples
- Current status data

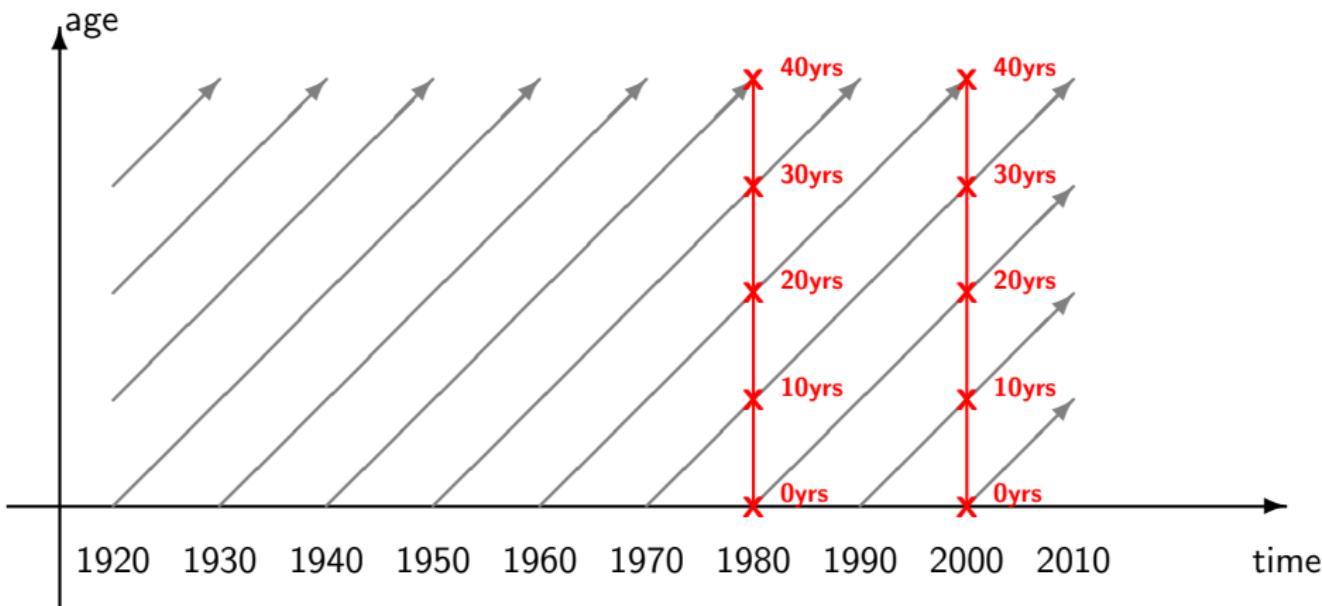
## 7 Estimating the force of infection

- Modelling current status data
- The early work
- The outbreak in the 90's
- The issue of monotonicity

# Endemic equilibrium

- Endemic Equilibrium = Time Homogeneity:
- Only one time scale: age.
- Following one cohort of individuals from birth to death.
- Main assumption: there is no change over time.

## The Lexis diagram



└ Introduction

└ Endemic equilibrium &amp; the static model

## The static model

- The static model is the steady state of the basic model, i.e., the parameters of the model do not depend on time but only on the host age.
- We assume that the force of infection,  $\lambda$ , is constant or age dependent but independent of time and that  $\alpha = 0$ .
- Differential equations:

$$\frac{dS}{da} = -(\lambda(a) + \mu(a))S(a), \quad \text{susceptible},$$

$$\frac{dI}{da} = \lambda(a)S(a) - (\nu + \mu(a))I(a), \quad \text{infected},$$

$$\frac{dR}{da} = \nu I(a) - \mu(a)R(a), \quad \text{Immune}.$$

- $\lambda(a)$ : age-dependent force of infection and  $S(0) = N(0) = B$  the number of births
- The age-dependent force of infection does not imply homogeneous mixing.

## The number of hosts at age $a$

- The number of hosts at age  $a$  is given by

$$N(a) = N(0)P(\text{survive until age } a)$$

- where with mortality rate  $\mu(a)$ :

$$P(\text{survive until age } a) = e^{-\int_0^a \mu(u)du}$$

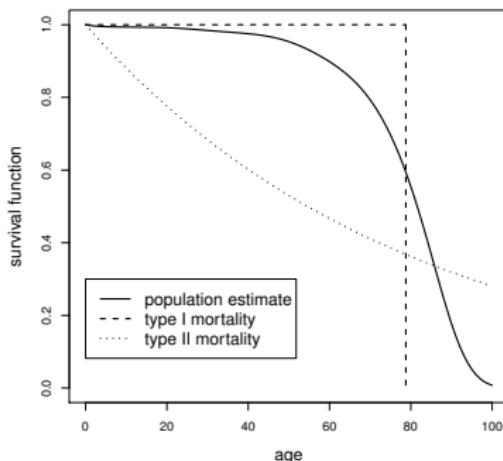
- For type I mortality we obtain ( $L$  = life expectancy)

$$N(a) = \begin{cases} N(0) & a \leq L, \\ 0 & a > L. \end{cases}$$

- For type II mortality the number of hosts at age  $a$  is given by

$$N(a) = N(0)e^{-\mu a}.$$

## The number of hosts at age $a$



**Figure:** The three models entail the same life expectancy: 78.8 years.

└ Introduction

└ Endemic equilibrium & the static model

## Number of susceptibles at age $a$

- The ordinary differential equation:

$$\frac{dS}{da} = -(\lambda(a) + \mu(a))S(a),$$

can be solved.

- The solution is given by

$$S(a) = S(0)e^{-\int_0^a (\lambda(u) + \mu(u))du},$$

where  $S(0) = N(0)$ .

- We assume all newborns are susceptible.

└ Introduction

└ Endemic equilibrium &amp; the static model

## Number of susceptibles at age $a$

- Assuming type I mortality we obtain

$$S(a) = \begin{cases} N(0)e^{-\int_0^a \lambda(u)du} & a \leq L, \\ 0 & a > L. \end{cases}$$

- Indeed, the change in the susceptible class ( $a \leq L$ ):

$$\frac{dS}{da} = -\lambda(a)S(a).$$

- Assuming type II mortality we obtain

$$S(a) = N(0)e^{-\int_0^a \lambda(u)du - \mu a}.$$

- The derivative with respect to the age is

$$\frac{dS}{da} = -(\lambda(a) + \mu)S(a),$$

## Proportion (fraction) of susceptibles at age $a$

- Instead of the total number of susceptibles we can use the proportion of susceptible hosts at age  $a$ :

$$s(a) = \frac{S(a)}{N(a)} = \frac{S(0)e^{-\int_0^a (\lambda(u) + \mu(u))du}}{N(0)e^{-\int_0^a \mu(u)du}} = e^{-\int_0^a \lambda(u)du}.$$

Note that we eliminate the natural rate of death,  $\mu$ , when we use the proportion susceptible.

- The change in the susceptible class:

$$\frac{ds(a)}{da} = -\lambda(a)s(a).$$

└ Introduction

└ Endemic equilibrium &amp; the static model

## Infected class

- The corresponding differential equation is given by

$$\frac{dI}{da} = \lambda(a)S(a) - (\nu + \mu(a))I(a).$$

which can be solved to obtain  $I(a)$ :

$$I(a) = N(0)e^{-\int_0^a (\nu + \mu(u))du} \int_0^a \lambda(u)e^{-\int_0^u (\lambda(v) - \nu)dv} du$$

- The proportion infected is given by

$$i(a) = e^{-\nu a} \int_0^a \lambda(u)e^{-\int_0^u (\lambda(v) - \nu)dv} du$$

- Simplifying assumption:  $\lambda(a) = \lambda$

$$i(a) = \frac{\lambda}{\lambda - \nu} [e^{-\nu a} - e^{-\lambda a}].$$

## Immune class

- The differential equation for the immune class is given by

$$\frac{dR}{da} = \nu I(a) - \mu(a)R(a).$$

- The total number of host in the immune class can be calculated by

$$R(a) = N(a) - S(a) - I(a).$$

- The proportion of immune

$$r(a) = 1 - s(a) - i(a).$$

└ Introduction

└ Endemic equilibrium & the static model

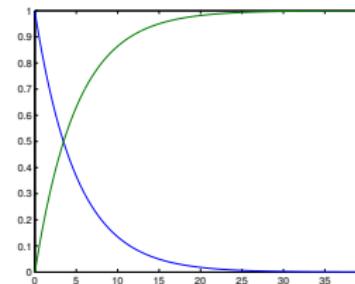
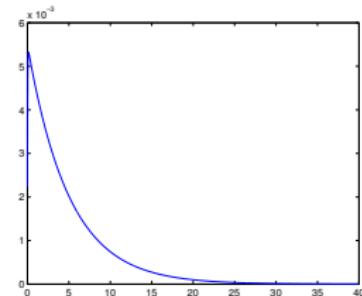
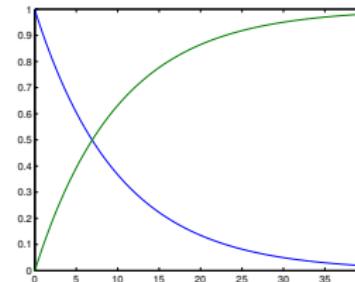
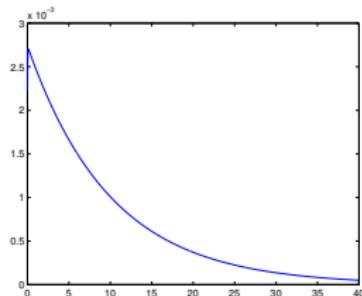
## The static model: an example

- Assume type I mortality with life expectancy  $L = 75$  years and recovery rate of  $\nu^{-1} = 10$  days.
- Proportion of individuals in each compartment (infected, susceptible and immune):  $\lambda = 0.1$  and  $\lambda = 0.2$
- What do you expect in terms of  $(s, i, r)$  and age?

└ Introduction

└ Endemic equilibrium &amp; the static model

## The static model: an example



**Figure:** Proportion infected (left column), susceptible and recovered (right column, blue and green, resp.). Top row:  $\lambda = 0.1$ . Bottom row:  $\lambda = 0.2$ .

# Overview

## 5 Introduction

- The SIR model
- The force of infection
- Is homogeneous mixing realistic?
- Estimating  $R_0$  from incidence data
- Endemic equilibrium & the static model

## 6 Serological Data

- Data examples
- Current status data

## 7 Estimating the force of infection

- Modelling current status data
- The early work
- The outbreak in the 90's
- The issue of monotonicity

# Data

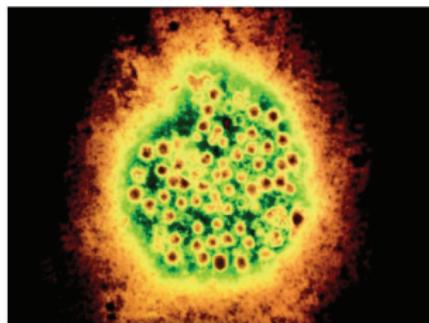
- Serological data:
  - measuring antigen-specific antibody levels in the blood.
  - cross-sectional
- Data:
  - Hepatitis A (Bulgaria)
  - Parvovirus B19 (Belgium)
  - Rubella (UK)
  - Varicella Zoster Virus (Belgium)

# Data

Table: Summary of the serological data sets.

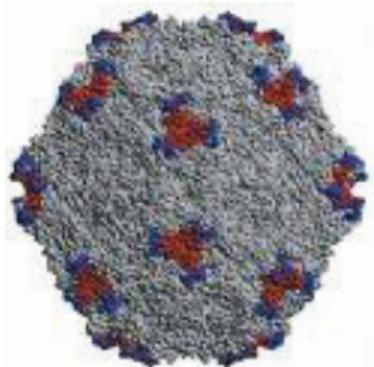
| Infection      | Main Transmission Route | Time Frame | Country  | Age range |
|----------------|-------------------------|------------|----------|-----------|
| Hepatitis A    | orofecal                | 1964       | Bulgaria | 1-86      |
| Parvovirus B19 | airborne                | 2001-2003  | Belgium  | 0-82      |
| Rubella        | airborne                | 1986-1987  | UK       | 1-44      |
| VZV            | airborne                | 2001-2003  | Belgium  | 0-82      |

## Hepatitis A: Bulgaria



- Acute inflammatory disease of the liver.
- Transmission: orofecal
- Infectiousness: two weeks before symptom onset
- Keiding (1991)

## Parvovirus B19: Belgium



Baby with the typical "slapped-cheek" rash, which is characteristic of fifth disease.

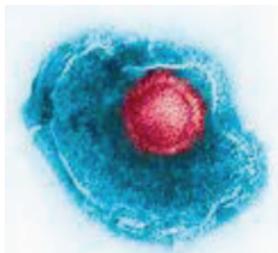
- B19 infection causes the so-called 'fifth disease', a mild rash illness ('slapped-cheek' rash)
- Transmission: respiratory droplets
- Infectious ( $\pm 6$  days) during the incubation period ( $\pm 14$  days)
- Disease burden: for pregnant women there is a potential for the fetus to have severe anemia, possibly leading to miscarriage.

## Rubella: UK



- German measles
- Transmission: direct or aerosol contact
- Incubation period of 2 to 3 weeks
- Farrington (1990)

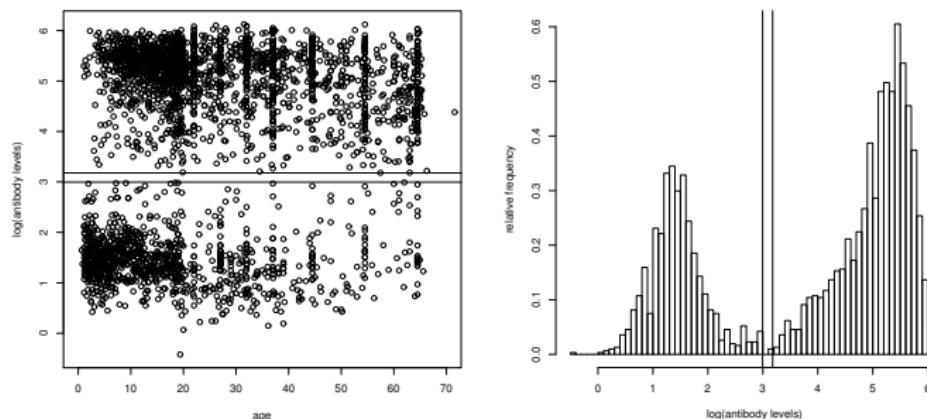
## Varicella Zoster Virus: Belgium



- Primary VZV infection results in chickenpox
- Transmission: direct or aerosol contact
- When infected, infectious for about 7 days
- Incubation period of two weeks
- Reactivation later in life (10 – 20%): herpes zoster or shingles
- Disease burden zoster: 25% is in constant pain

## Serological data: antibody levels

- We consider an age-specific cross-sectional prevalence sample of size  $N$  and let  $a_i$  be the age at sampling for the  $i$ th subject.
- Let  $Z_i$  the antibody activity level (in U/ml) for the  $i$ th subject.



**Figure:** Belgian Parvovirus B19 data:  $\log(\text{antibody titers})$  versus age (left panel); a histogram of the  $\log(\text{antibody titers})$  ignoring age (right panel).

└ Serological Data

  └ Current status data

## From serology to current status data

- Define the dichotomized version of  $Z_i$  as the binary variable  $Y_i$

$$Y_i = \begin{cases} 1 & \text{if } Z_i > \tau_u & \text{seropositive,} \\ 0 & \text{if } Z_i < \tau_\ell & \text{seronegative,} \end{cases}$$

- Note that  $Y_i$  is missing if  $\tau_\ell < Z_i < \tau_u$ : ignoring equivocals
- $\tau_\ell$  and  $\tau_u$  are the lower and upper threshold values.

Example: 20 resp. 24 U/ml for parvovirus B19

- Result:  $(a_i, Y_i), i = 1 \dots N$  (redefine  $N$  if you left out equivocals).

└ Serological Data

└ Current status data

## From serology to current status data

- From continuous to binary data, to seroprevalence = age-specific proportion of seropositives

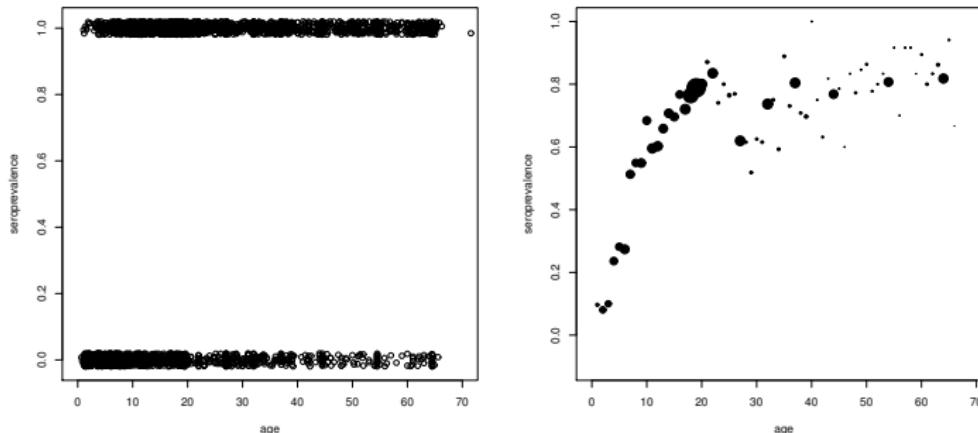


Figure: Belgian data on parvovirus B19.

└ Serological Data

  └ Current status data

## Hepatitis A: Bulgaria

Bulgarian data on hepatitis A:

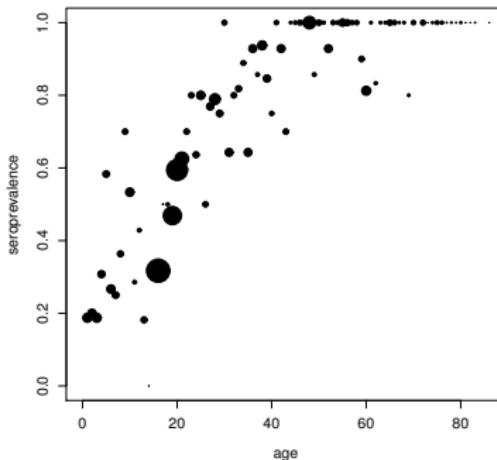


Figure: Age-specific proportion positive samples of hepatitis A based on cross-sectional survey in Bulgaria anno 1964.

- └ Serological Data

- └ Current status data

## Rubella: UK

UK data on rubella

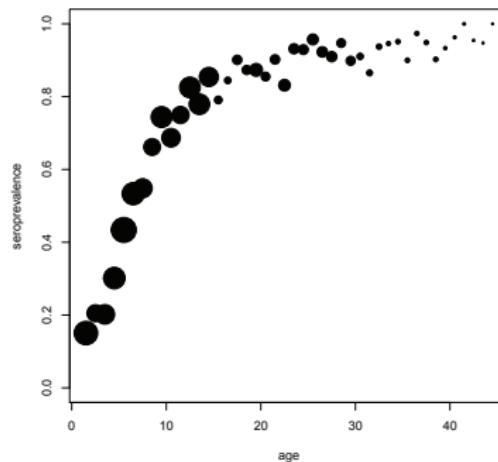


Figure: Age-specific proportion positive samples of rubella based on cross-sectional survey in the UK.

Serological Data

Current status data

## Varicella Zoster Virus: Belgium

Belgian data on varicella zoster virus

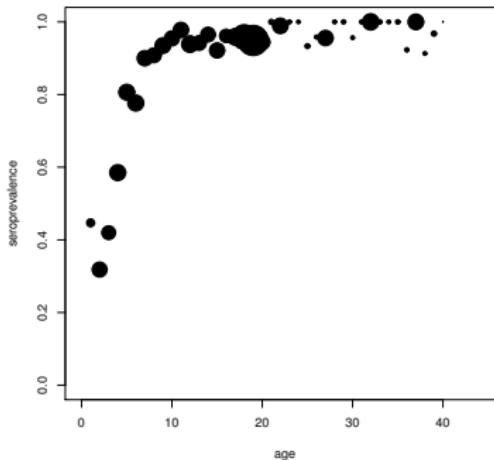


Figure: Age-specific proportion positive samples of VZV based on cross-sectional survey in Belgium (up to age 40 years).

# Overview

## 5 Introduction

- The SIR model
- The force of infection
- Is homogeneous mixing realistic?
- Estimating  $R_0$  from incidence data
- Endemic equilibrium & the static model

## 6 Serological Data

- Data examples
- Current status data

## 7 Estimating the force of infection

- Modelling current status data
- The early work
- The outbreak in the 90's
- The issue of monotonicity

## Modelling current status data

- Consider a prevalence sample of size  $N$  and let  $a_i$  be the age of the  $i$ th subject.
- Instead of observing the age of infection we observe a binary variable:

$$Y_i = \begin{cases} 1 & \text{subject } i \text{ experienced infection before age } a_i \text{ (left-censored)} \\ 0 & \text{elsewhere (right-censored)} \end{cases}$$

- The prevalence is assumed equal to the seroprevalence:

$$\pi(a_i) = P(Y_i = 1 | a_i).$$

- We ignore misclassifications and diagnostic uncertainty.
- The log likelihood (parameter vector  $\theta$ ):

$$L(\theta) = \sum_{i=1}^N \{ Y_i \log(\pi(a_i)) + (1 - Y_i) \log(1 - \pi(a_i)) \}.$$

## Modeling current status data: GLMs

- In the terminology of generalized linear models the age-dependent probability  $\pi(a)$  is modeled as

$$\pi(a) = g^{-1}(\eta(a))$$

- $\eta(a)$ : linear predictor.
- $g$ : link function.

## Modeling current status data: link functions

- For binary responses,  $g$  is often taken to be

- A logit link function:

$$\log(\pi(a)/(1 - \pi(a))).$$

- A complementary log-log link:

$$\log(-\log(1 - \pi(a))).$$

- A log link:

$$-\log(1 - \pi(a)).$$

## Force of infection

- The force of infection is given by

$$\lambda(a) = \frac{\pi'(a)}{1 - \pi(a)}, \text{ where } \pi'(a) = \frac{d\pi(a)}{da}.$$

- Link to survival analysis:

- $\pi(a)$  is the distribution function
- $\pi'(a)$  is the density function
- $1 - \pi(a) = s(a)$  is the survival function
- $\lambda(a)$  is the infection hazard

- In the general case for binary response the force of infection has the form of

$$\lambda(a) = \eta'(a)\delta[\eta(a)]$$

where  $\delta[\eta(a)]$  is determined by the link function.

- └ Estimating the force of infection
  - └ Modelling current status data

## Force of infection

- For a model with logit link function we have

$$\lambda(a) = \frac{\pi'(a)}{1 - \pi(a)} = \eta'(a) \frac{e^{\eta(a)}}{1 + e^{\eta(a)}} = \eta'(a)\pi(a)$$

- For models with a complementary loglog link we have

$$\lambda(a) = \frac{\pi'(a)}{1 - \pi(a)} = \eta'(a)e^{\eta(a)}$$

- For models with log link function we have

$$\lambda(a) = \frac{\pi'(a)}{1 - \pi(a)} = \frac{\eta'(a)e^{-\eta(a)}}{e^{-\eta(a)}} = \eta'(a)$$

## Examples

- Some parametric models:
  - Exponential model
  - Weibull model
  - Log-logistics model
- Applied to Rubella in the UK
- Software R

## The exponential model applied

- The time spent in the susceptible class is an exponential distribution:

$$s(a) = e^{-\lambda a}$$

- = model with constant force of infection.
- We use the `glm()` function with log link function.

```
>fit.exp<-glm(NEG/NTOT ~ -1+AGE,  
                family = binomial(link=log),  
                data = Rub1)
```

## The exponential model applied

- Constant force of infection.

```
> summary(fit.exp)
```

Call:

```
glm(formula = NEG/NTOT ~ -1 + AGE, family = binomial(link = log),  
     data = Rub1)
```

Coefficients:

|     | Estimate | Std. Error | z value | Pr(> z )     |
|-----|----------|------------|---------|--------------|
| AGE | -0.09668 | 0.02117    | -4.566  | 4.97e-06 *** |

---

Signif. codes: 0 \*\*\* 0.001 \*\* 0.01 \* 0.05 . 0.1 1  
(Dispersion parameter for binomial family taken to be 1)

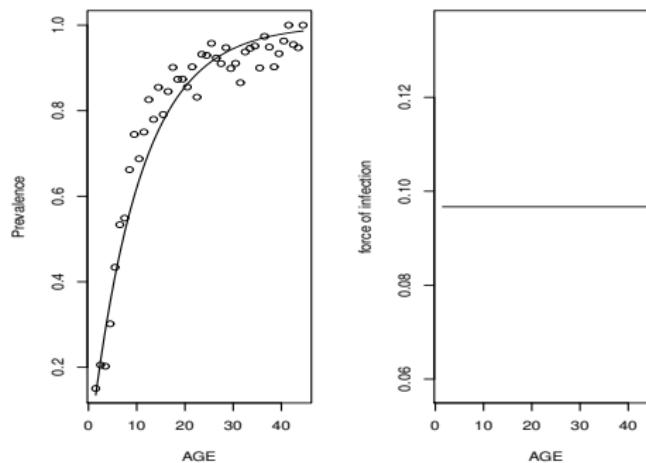
Null deviance: Inf on 44 degrees of freedom  
Residual deviance: 1.2493 on 43 degrees of freedom  
AIC: 18.914

└ Estimating the force of infection

└ Modelling current status data

# The exponential model applied

- Constant force of infection.



## The Weibull model applied

- Assume that the time spent in the susceptible class is Weibull:

$$s(a) = e^{-\zeta a^\alpha}.$$

- The force of infection is given by

$$\lambda(a) = \frac{\alpha \zeta a^{\alpha-1} e^{-\zeta a^\alpha}}{e^{-\zeta a^\alpha}} = \zeta \alpha a^{\alpha-1}.$$

## The Weibull model applied

- The prevalence

$$\pi(a) = 1 - e^{-\zeta a^\alpha}.$$

- GLM with complementary log-log link function

$$g(\pi(a)) = \log(-\log(1 - \pi(a))) = \log(\zeta) + \alpha \log(a).$$

- Age dependent force of infection

$$\lambda(a) = \zeta \alpha a^{\alpha-1},$$

which is monotone (increasing or decreasing) force of infection, depending on the sign of  $\alpha$ .

- Note that for  $\alpha = 1$  the Weibull model is an exponential model.

## The Weibull model applied

- A model with monotone force of infection.
- We fit the model using  $\log(\text{age})$  as a predictor.
- We use the `glm()` function with cloglog link function.

```
> fit.weibul<-glm(POS/NTOT ~ log(AGE),  
+ family = binomial(link=cloglog),  
+ data = Rub1)
```

## The Weibull model applied

### ■ Age-dependent force of infection

```
> summary(fit.weibul)
```

Call:

```
glm(formula = POS/NTOT ~ log(AGE), family=binomial(link=cloglog),  
     data = Rub1)
```

Deviance Residuals:

| Min        | 1Q         | Median    | 3Q        | Max       |
|------------|------------|-----------|-----------|-----------|
| -0.3317246 | -0.1083039 | 0.0008287 | 0.1086089 | 0.2385534 |

Coefficients:

|             | Estimate | Std. Error | z value | Pr(> z )   |
|-------------|----------|------------|---------|------------|
| (Intercept) | -1.8924  | 0.9236     | -2.049  | 0.04046 *  |
| log(AGE)    | 0.8498   | 0.3082     | 2.758   | 0.00582 ** |
| ---         |          |            |         |            |

Signif. codes: 0 \*\*\* 0.001 \*\* 0.01 \* 0.05 . 0.1 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 12.0671 on 43 degrees of freedom

Residual deviance: 1.0133 on 42 degrees of freedom

AIC: 21.276

## The Weibull model applied

- A model with monotone force of infection:

$$\lambda(a) = \zeta \alpha a^{\alpha-1},$$

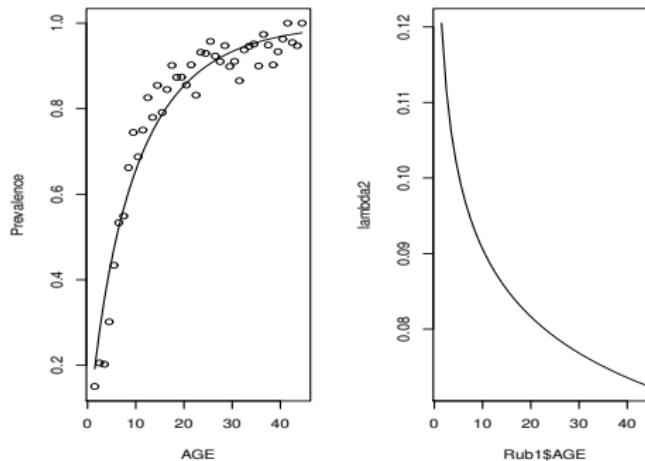
```
> fit.weibul$coeff
(Intercept)      log(AGE)
-1.8923888     0.8497694
> zeta<-exp(fit.weibul$coeff[1])
> alpha<-fit.weibul$coeff[2]
> lambda2<-zeta*alpha*Rub1$AGE^(alpha-1)
```

└ Estimating the force of infection

└ Modelling current status data

# The Weibull model applied

- Monotone force of infection.



## The log-logistics model applied

- The time spent in the susceptible class follows a log-logistic distribution.

$$s(a) = \frac{1}{1 + \zeta a^\alpha}.$$

- The hazard (or the force of infection):

$$\lambda(a) = \frac{\alpha a^{\alpha-1} \zeta}{1 + \zeta a^\alpha}.$$

- └ Estimating the force of infection
  - └ Modelling current status data

## The log-logistics model applied

- The prevalence

$$\pi(a) = \frac{\zeta a^\alpha}{1 + \zeta a^\alpha}.$$

$$1 - \pi(a) = 1 - \frac{\zeta a^\alpha}{1 + \zeta a^\alpha} = \frac{1}{1 + \zeta a^\alpha}.$$

- GLM with logit link function

$$\log\left(\frac{\pi(a)}{1 - \pi(a)}\right) = \log(\zeta a^\alpha) = \log(\zeta) + \alpha \log(a).$$

- Age-dependent force of infection

$$\lambda(a) = \frac{\alpha a^{\alpha-1} \zeta}{1 + \zeta a^\alpha}.$$

which is a single peak force of infection.

## The log-logistics model applied

- We fit the model using  $\log(\text{age})$  as a predictor.
- We use the `glm()` function with logit link function.
- ```
> fit.loglogistic<-glm(POS/NTOT ~ log(AGE),  
family = binomial(link=logit),  
data = Rub1)
```

# The log-logistics model applied

## ■ The R output.

```
> summary(fit.loglogistic)
Call:
glm(formula = POS/NTOT ~ log(AGE), family=binomial(link=logit),
     data = Rub1)

Deviance Residuals:
    Min      1Q  Median      3Q      Max
-0.25303 -0.07079  0.01551  0.08371  0.29270

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.9380    1.5088  -1.947  0.05151 .
log(AGE)     1.6284    0.5725   2.844  0.00445 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1  1
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 12.06714  on 43  degrees of freedom
Residual deviance: 0.67183  on 42  degrees of freedom
AIC: 19.459
```

## The log-logistics model applied

- A model with a single peak force of infection:

$$\lambda(a) = \frac{\alpha a^{\alpha-1} \zeta}{1 + \zeta a^\alpha}.$$

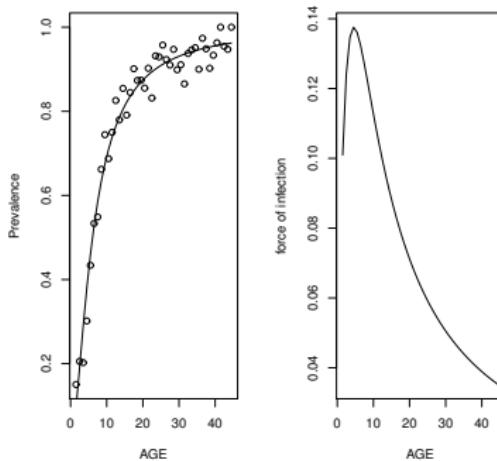
```
> fit.loglogistic$coeff
(Intercept)    log(AGE)
-2.938030     1.628396
> zeta<-exp(fit.loglogistic$coeff[1])
> alpha<-fit.loglogistic$coeff[2]
> aa<-alpha*(Rub1$AGE^(alpha-1))*zeta
> bb<-1+zeta*(Rub1$AGE^(alpha))
> lambda3<-aa/bb
```

└ Estimating the force of infection

  └ Modelling current status data

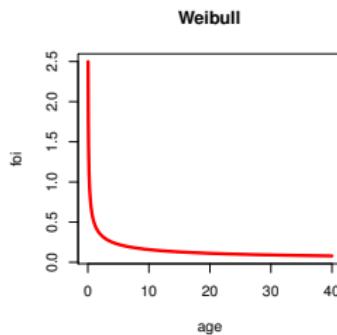
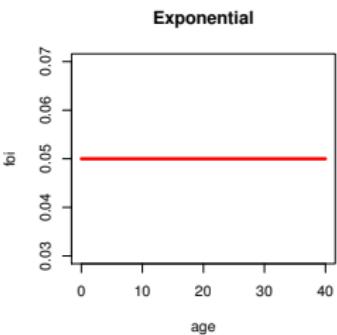
## The log-logistics model applied

- Single peak force of infection.



- └ Estimating the force of infection
  - └ Modelling current status data

## Summary: parametric models for the force of infection



- └ Estimating the force of infection
  - └ Modelling current status data

# Estimating the force of infection

## *Binomial likelihood*

### ■ Direct Parametrization

$$\ell(\theta) = \sum_{i=1}^N Y_i \log(1 - e^{-\int_0^{a_i} \lambda(u) du}) + (1 - Y_i) \log(e^{-\int_0^{a_i} \lambda(u) du}).$$

$$\hookrightarrow \hat{\lambda}(a)$$

### ■ Indirect Parametrization

$$\ell(\theta) = \sum_{i=1}^N Y_i \log(\pi(a_i)) + (1 - Y_i) \log(1 - \pi(a_i)).$$

$$\hookrightarrow \hat{\lambda}(a) = \hat{\pi}'(a)/(1 - \hat{\pi}(a))$$

## Estimating the force of infection in R

- Use generalized linear models

```
glm(...)
```

- Define your own (log)likelihood and maximize it:

```
loglik=function(theta){  
  ...  
  return(totalloglik)  
}
```

```
mle(loglik,...)
```

└ Estimating the force of infection

  └ The early work

## Early work

- How did it all start ?
- Time traveling from 1934 to 1989:
  - Muench (1934)
  - Griffiths (1974)
  - Grenfell and Anderson (1985)
  - Becker (1989)

# The thing to do . . . The catalytic model of Muench

25]

*Derivation of Rates by the Catalytic Curve*

25

DERIVATION OF RATES FROM SUMMATION DATA  
BY THE CATALYTIC CURVE<sup>1</sup>

BY HUGO MUENCH, International Health Division, The Rockefeller Foundation

One of the problems confronting the epidemiologist is the evaluation of the effective contact rate of a given disease among a given population group. What proportion of individuals are exposed to infection during, say, a year? What fraction will not develop the disease because they have had previous attacks, or because, perhaps, they have a natural immunity?

The answers to these questions are generally obtained from the study of morbidity rates. Now an official morbidity table is a notoriously tricky thing; it is a compilation, usually, of the reports of some of the practicing physicians of a community, on those cases which they have seen and have been able to diagnose. Results based on such data can never be more than approximations.

For this reason, more critical studies are founded on surveys which keep under observation a sample of the population over a considerable length of time. The accumulation of sufficient material in this way is apt to be tedious and expensive.

There is still another approach to the problem in the case of diseases which leave traces of their attacks in the shape of permanent immunity. The fact that an effective exposure has taken place can be ascertained, sometimes by a history, at times by means of some test. In this case, the figures represent a *summation*. They give the sum of all children ten years old who have had the disease at any time in the past; the sum of all effective exposures, at any age, during the years up to that age.

Knowing the sum of infections at different ages, it is possible to go back and find the exposure rate which would produce such a sum. Mathematically, the gradually increasing proportion of positive results with increasing ages is an integral, the derivative of which represents the rate of increase. The thing to do, then, is to find out what curve describes the growth of the summation data and to find its derivative, which will be the rate at which the curve is rising at different ages. This was done in a notable manner by Collins,<sup>2</sup> who fitted various

"The thing to do, then, is to find out what curve describes the growth of the summation data and to find its derivative, which will be the rate at which the curve is rising at different ages."

(Hugo Muench, 1934)

<sup>1</sup> Revision of a paper delivered at the Annual Meeting of the American Statistical Association in Philadelphia, December, 1933.

<sup>2</sup> Selwyn D. Collins, "Age Incidence of the Common Communicable Diseases of Children," U. S. Public Health Reports, Vol. 44, No. 14, April 8, 1929.

## Just do it

- A model for the prevalence:

$$\pi(a) = 1 - e^{-\lambda a}.$$

- Derivative=rate of change:

$$\pi'(a) = \lambda e^{-\lambda a}$$

- The rate of change per susceptible:

$$\frac{\pi'(a)}{1 - \pi(a)} = \frac{\lambda e^{-\lambda a}}{e^{-\lambda a}} = \lambda.$$

## Software: Muench's model in R (1)

- A model with constant force of infection.

```
> fit.muensch<-glm(NEG/NTOT ~ -1+AGE,  
                     family = binomial(link=log),  
                     data = Kei1)
```

## Software: Muench's model in R (1)

- $\hat{\lambda} = 0.052.$

```
> summary(fit.muench)
```

Call:

```
glm(formula = NEG/NTOT ~ -1 + AGE, family = binomial(link = log),  
    data = Kei1)
```

Coefficients:

|                | Estimate  | Std. Error | z value | Pr(> z ) |        |        |     |   |
|----------------|-----------|------------|---------|----------|--------|--------|-----|---|
| AGE            | -0.051856 | 0.008321   | -6.232  | 4.6e-10  | ***    |        |     |   |
| ---            |           |            |         |          |        |        |     |   |
| Signif. codes: | 0         | ***        | 0.001   | **       | 0.01 * | 0.05 . | 0.1 | 1 |

AIC: 39.321

## Software: Muench's model in R (2)

- We maximize the likelihood using the function `mle()`.

```
# The library for maximum likelihood estimation
library(stats4)

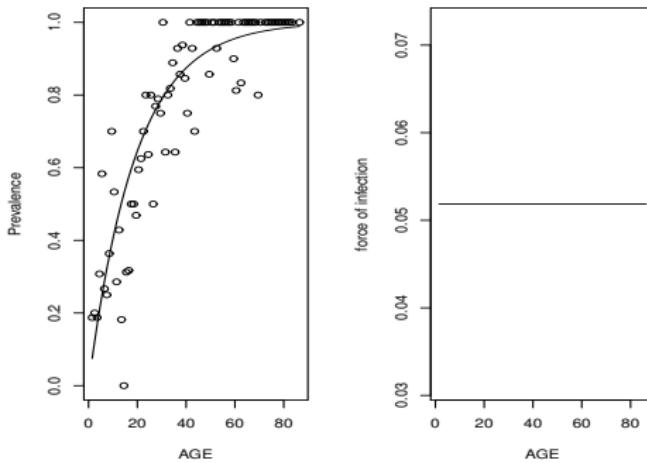
# Read in the data
b19<-read.table("VZV-B19-BE.dat",header=T)
age<-b19$age[!is.na(b19$parvores)&b19$age>0.5]
resp<-b19$parvores[!is.na(b19$parvores)&b19$age>0.5]
y<-resp[order(age)]
a<-age[order(age)]

# Model of Muench
muench<-function(theta){
  p<-1-exp(-theta*a)
  ll<-y*log(p)+(1-y)*log(1-p)
  return(-sum(ll))
}
fit<-mle(muench,start=list(theta=0.2))
```

└ Estimating the force of infection

  └ The early work

## Hepatitis A in Bulgaria: Muench's Model



## Age-dependent force of infection: Griffiths (1974)

- Griffiths (1974) proposed a model for measles in which the force of infection increases linearly with age.
- Specifically, Griffiths (1974) suggested

$$\lambda(a) = \begin{cases} \theta_1(a + \theta_0) & a > \tau, \\ 0 & a \leq \tau. \end{cases}$$

- In this model the force of infection is zero between  $0 - \tau$  years which corresponds to the maternal antibody period.
- Note that, since Griffiths (1974) specified  $\tau$  as a parameter in the model, Griffiths' model should be interpreted as a changepoint model.

## The prevalence for a linear force of infection

- Griffiths (1974) mentioned that his model for the prevalence corresponds to a model in which the linear predictor is a quadratic function of age.
- This means that if:

$$\lambda(a) = \theta_1 + \theta_2 a$$

then

$$\pi(a) = 1 - \exp \left\{ - \left( \theta_0 + \theta_1 a + \frac{1}{2} \theta_2 a^2 \right) \right\}.$$

- Griffiths (1974) applied the model to the first 10 years of age.

## Software: Griffiths' model in R (1)

- A model with linear force of infection.
- A GLM with log link and the restriction  $\pi(0) = 0$ :

```
age1<-Kei1$AGE  
age2<-Kei1$AGE^2  
fit.griffiths<-glm(Kei1$NEG/Kei1$NTOT~-1+age1+age2,  
                    family = binomial(link=log))
```

## Software: Griffiths' model

- R output.

```
> summary(fit.griffiths)
Call:
glm(formula = Kei1$NEG/Kei1$NTOT ~ -1 + age1 + age2,
     family = binomial(link = log))

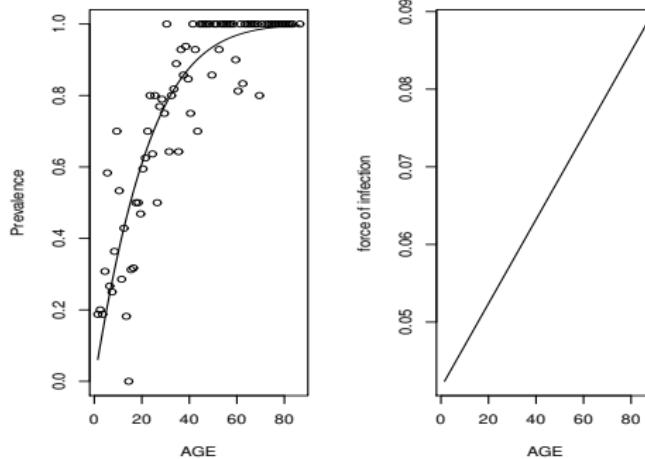
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
age1 -0.0415572  0.0183284 -2.267   0.0234 *
age2 -0.0002707  0.0004581 -0.591   0.5545
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

AIC: 40.411
```

└ Estimating the force of infection

  └ The early work

## Hepatitis A in Bulgaria: Griffiths' model



- └ Estimating the force of infection

- └ The early work

## Griffiths' Model in R (2)

```
# Model of Griffiths
griffiths<-function(theta0,theta1){
  p<-1-exp(-theta0*a[a<=10]-theta1*a[a<=10]^2)
  ll<-y[a<=10]*log(p)+(1-y[a<=10])*log(1-p)
  return(-sum(ll))
}
fit<-mle(griffiths,start=list(theta0=0.052,theta1=-0.00006))
```

## First attempt for a nonparametric estimate

- Interestingly, Griffiths (1974) justified his choice of a linear force of infection by using a nonparametric estimate for the force of infection.
- The estimate

$$\lambda(a) = \Delta\pi/(1 - \pi(a)),$$

was plotted against age and showed the linear trend of the force of infection.

## Flexibility: Grenfell & Anderson

- Grenfell and Anderson (1985) extended the model further and used polynomial functions to model the force of infection.
- The advantage of higher order polynomials is their flexible curve shapes. Grenfell and Anderson (1985) did not restrict the force of infection to be constant or linear but gave the data to lead the results.

## The prevalence and polynomial force of infection

- Grenfell and Anderson (1985)'s model assumes that the prevalence is given by

$$\pi(a) = 1 - e^{-\sum \theta_i a^i},$$

- which implies that the force of infection is

$$\lambda(a) = \sum \theta_i i a^{i-1}.$$

- Note that within the framework of generalized linear models (McCullagh and Nelder, 1989) for binary responses, the model of Grenfell and Anderson (1985) can be fitted using a log link function.
- In this case, the force of infection is simply the first derivative of the linear predictor.

└ Estimating the force of infection

  └ The early work

## The framework: serological data

- Grenfell and Anderson (1985) were the first to use serological data for the estimation of the force of infection.
- They proposed to choose the model which minimizes the deviance since this model has the best goodness-of-fit to the data.

## Software: Grenfell and Anderson's Model in R (1)

- A model with linear force of infection.
- A GLM with log link:

```
> fit.grenfell<-glm(Kei1$NEG/Kei1$NTOT ~ age1+age2+age3,  
family = binomial(link=log))
```

## Software: Grenfell and Anderson's Model

- R output.

```
> summary(fit.grenfell)
glm(formula = Kei1$NEG/Kei1$NTOT ~ age1 + age2 + age3,
    family=binomial(link = log))
```

Coefficients:

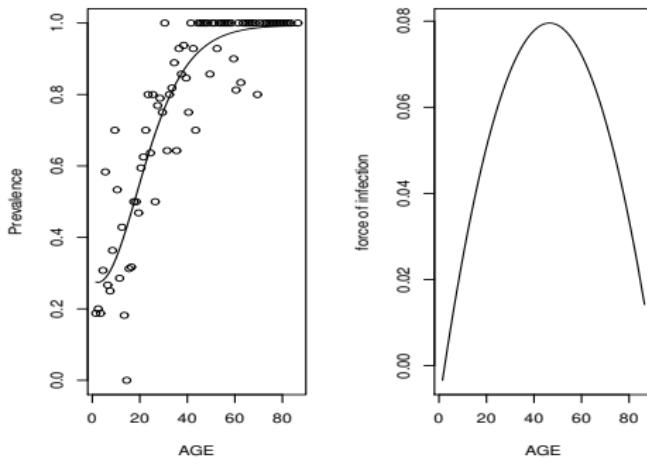
|             | Estimate   | Std. Error | z value | Pr(> z ) |
|-------------|------------|------------|---------|----------|
| (Intercept) | -3.314e-01 | 4.063e-01  | -0.816  | 0.415    |
| age1        | 9.012e-03  | 6.692e-02  | 0.135   | 0.893    |
| age2        | -1.903e-03 | 2.680e-03  | -0.710  | 0.478    |
| age3        | 1.364e-05  | 2.762e-05  | 0.494   | 0.622    |

AIC: 45.755

└ Estimating the force of infection

  └ The early work

## Hepatitis A in Bulgaria: Grenfell and Anderson's model



# The piecewise constant force of infection

- Becker (1989):
  - nonlinear shape:
  - piecewise constant:

$$\lambda(a) = \mu\alpha a^{\alpha-1},$$

$$\lambda(a) = \theta_i \text{ for } a_{i-1} \leq a < a_i$$

- └ Estimating the force of infection

- └ The early work

## Becker's model in R

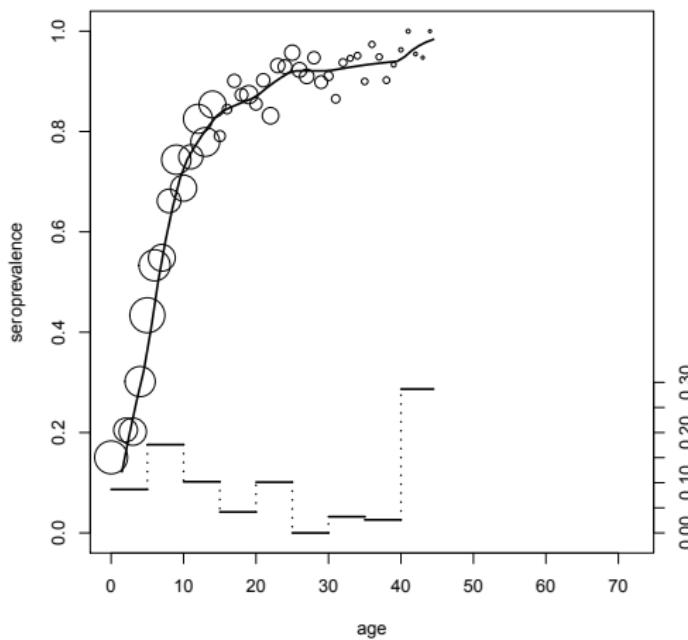
```
# Becker 1989: piecewise constant FOI (ensuring positivity)
breakpoints<-seq(0,45,5,include.lowest=T)
pcwrate.fitter(Rub1$POS,Rub1$AGE,Rub1$NTOT,breakpoints)
```

- Estimating the force of infection

- The early work

## Becker's model applied

Becker applied to Rubella in the UK:



└ Estimating the force of infection

  └ The early work

## Assumptions

Estimating the force of infection from serological data holds when:

- endemic equilibrium/time homogeneity and
- lifelong immunity and
- antibody titers are a good marker for immunity

# The outbreak in the 90's

STATISTICS IN MEDICINE  
*Statist. Med.* **18**, 307–320 (1999)

*Appl. Statist.* (2001)  
**50**, Part 3, pp. 251–292

## SEMI-PARAMETRIC ESTIMATION OF AGE-TIME SPECIFIC INFECTION INCIDENCE FROM SERIAL PREVALENCE DATA

### Estimation for infecti serologica

NICO NAGELKERKE<sup>1, 2, 3\*</sup>, SIEM HEISTERKAMP,<sup>1</sup> MARTIEN BORGDOORFF<sup>4</sup>,  
 JAAP BROEKMAN<sup>5</sup> AND HANS VAN HOUWELINGEN<sup>3</sup>

<sup>1</sup>National Institute of Public Health and the Environment, Bilthoven, The Netherlands

<sup>2</sup>Department of Public Health, Erasmus University, Rotterdam, The Netherlands

<sup>3</sup>Department of Medical Microbiology, University of Manitoba, Winnipeg, Canada

<sup>4</sup>Royal Netherlands Tuberculosis Association, The Hague, The Netherlands

<sup>5</sup>Department of Medical Statistics, University of Leiden, Leiden, The Netherlands

C. P. Farrington and M.

*The Open University, Mil-*

and N. J. Gay

*Communicable Disease S*

[Read before The Royal Sta  
 Professor D. A. Lievesley, ir

**Summary.** The basic reprodt  
 infections generated by a sin

*J. R. Statist. Soc. A* (1991)  
**154**, Part 3, pp. 371–412

## Age-specific Incidence and Prevalence: a Statistical Perspecti

By NIELS KEIDING†

*University of Copenhagen, Denmark*

[Read before The Royal Statistical Society on Wednesday, February 6th, 1991,  
 the President, Professor P. G. Moore, in the Chair]

## SUMMARY

In epidemiology *incidence* denotes the rate of occurrence of new cases (of disease), *v prevalence* is the frequency in the population (of diseased people). From a statistical p view it is useful to understand incidence and prevalence in the parameter space, incid as intensity (hazard) and prevalence as probability, and to relate observable quantities these via a statistical model. In this paper such a framework is based on modelling individual's dynamics in the Lexis diagram by a simple three-state stochastic process it age direction and recruiting individuals from a Poisson process in the time direction. resulting distributions in the cross-sectional population allow a rigorous discussion o

└ Estimating the force of infection

  └ The outbreak in the 90's

## The outbreak in the 90's

- The main issue: estimation under order restrictions.
- The prevalence  $\pi(a_i)$  should be estimated under the restrictions:

$$\pi(a_1) \leq \pi(a_2) \leq \dots \leq \pi(a_K).$$

- Otherwise: the estimated force of infection is negative.

└ Estimating the force of infection

  └ The outbreak in the 90's

## Farrington (1990)

- Becker (1989); Farrington (1990); Farrington et al. (2001) proposed nonlinear models for the force of infection.
- In Farrington (1990) the model for the force of infection is defined by

$$\lambda(a) = (\alpha_1 a - \alpha_3)e^{-\alpha_2 a} + \alpha_3.$$

- In order to ensure that the force of infection satisfies  $\lambda(a_i) \geq 0$ ,  $i = 1, 2, \dots, n$ , Farrington (1990) constrained the parameter space to be nonnegative ( $\alpha_j \geq 0$ ,  $j = 1, 2, 3$ ).

└ Estimating the force of infection

  └ The outbreak in the 90's

## The prevalence

- The model of Farrington assumes that the force of infection is zero at birth ( $\lambda(0) = 0$ ) and then rises to a peak in a linear fashion followed by an exponential decrease.
- The peak is reached at an age corresponding to the maximum contact rate of susceptibles with infectious individuals.
- The parameter  $\alpha_3$  is called the long term residual value of the force of infection.
- If  $\alpha_3 = 0$ , the force of infection decreases to 0 as  $a$  tends to infinity.

└ Estimating the force of infection

  └ The outbreak in the 90's

## Example

### ■ The prevalence

$$\pi(a) = 1 - \exp \left\{ \frac{\alpha_1}{\alpha_2} a e^{-\alpha_2 a} + \frac{1}{\alpha_2} \left[ \frac{\alpha_1}{\alpha_2} - \alpha_3 \right] [e^{-\alpha_2 a} - 1] - \alpha_3 a \right\}.$$

└ Estimating the force of infection

  └ The outbreak in the 90's

## Farrington's model in R

```
# Farrington 1990
farrington<-function(apareta,bpareta,cpareta){
  apar<-exp(aeta) # to ensure positivity
  bpar<-exp(beta) # to ensure positivity
  cpar<-exp(ceta) # to ensure positivity
  p<-1-exp(apar/bpar*a*exp(-bpar*a)+1/bpar*(apar/bpar-cpar)
            *(exp(-bpar*a)-1)-cpar*a)
  ll<-y*log(p)+(1-y)*log(1-p)
  return(-sum(ll))}

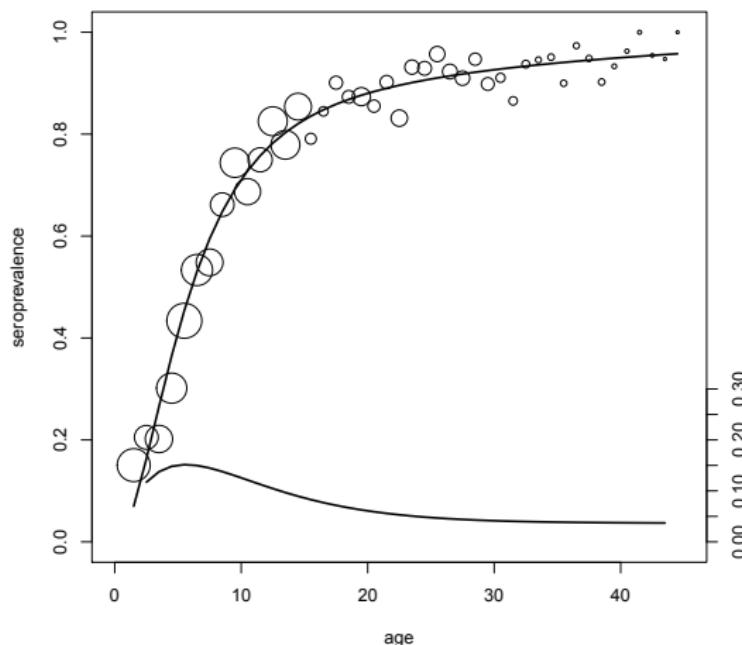
fit<-mle(farrington,start=list(apareta=-2.7,bpareta=-1.6,cpareta=-2))
```

- Estimating the force of infection

- The outbreak in the 90's

## Farrington's model applied

Farrington applied to Rubella in the UK:



## Finally, a nonparametric method

- The first attempt of Griffiths (1974) to estimate the force of infection non-parametrically followed by Farrington (1990), who used a smoothed version of the Griffiths $\frac{1}{2}$  estimator.
- However, both can lead to a negative estimate for the force of infection.
- Keiding (1991) proposed a two step approach in which in the first step the prevalence is estimated by isotonic regression (Barlow et al., 1972; Robertson et al., 1988) and in the second step a kernel smoother is used in order to estimate the force of infection.

- └ Estimating the force of infection
  - └ The outbreak in the 90's

## Keiding (1991)

- In the first step, Keiding (1991) proposed to estimate the prevalence using an isotonic regression estimate of the observed prevalence.
- This can be done by applying the pool adjacent violator algorithm (PAV) to the data. In this case, the nonparametric maximum likelihood (NPMLE) for the prevalence is a step function.
- The prevalence

$$\pi(a) = 1 - e^{-\Lambda(a)}.$$

- The force of infection  $\lambda(a)$ :

$$\Lambda(a) = \int_0^a \lambda(u) du.$$

└ Estimating the force of infection

  └ The outbreak in the 90's

## Isotonic regression: the PAV algorithm

- For  $a_i > a_{i-1}$ , if  $\pi(a_i) < \pi(a_{i-1})$  then:

$$\pi(a_i) = \pi(a_{i-1}) = \frac{\pi(a_i) + \pi(a_{i-1})}{2}$$

- We pool the observed prevalence whenever the order is violated.

└ Estimating the force of infection

  └ The outbreak in the 90's

## The force Of infection

- In the second step the force of infection, assumed to be a smooth function of age, is estimated by

$$\hat{\lambda}(a) = \frac{1}{h} \int_{a-h}^{a+h} K\left(\frac{x-a}{h}\right) \frac{d\hat{\pi}(x)}{1-\hat{\pi}(x^-)},$$

where  $K$  is a kernel function,  $h$  is the bandwidth and  $\hat{\pi}(u)$  is the isotonic regression of the observed prevalence.

- Keiding's method requires the choice of a bandwidth.

└ Estimating the force of infection

  └ The outbreak in the 90's

## Keiding's model in R

- Isotonic regression in R: `monoreg()`, `pava()`, `isoreg()`.

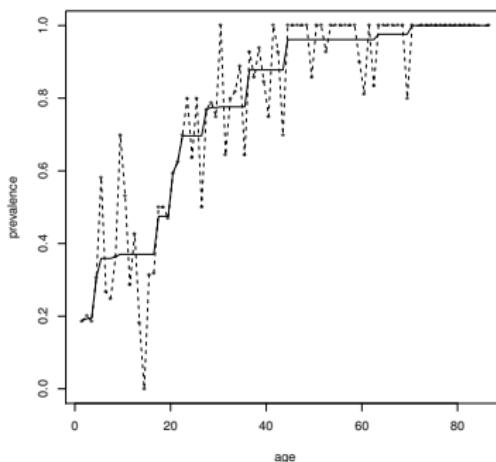
```
xi<-Kei1$AGE  
wi<-Kei1$NTOT  
yi<-Kei1$POS/Kei1$NTOT  
par(mfrow=c(1,1))  
plot(xi,yi,pch="*",xlab="age",ylab="prevalence")  
lines(xi,yi,lty=2)  
iso1<-pava(yi,wt=wi)  
lines(xi,iso1)
```

└ Estimating the force of infection

  └ The outbreak in the 90's

## Example: Keiding's model for hepatitis A

- Prevalence: Isotonic regression for hepatitis A in Bulgaria.

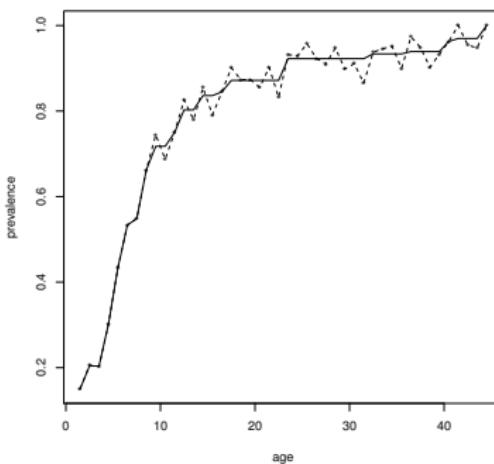


└ Estimating the force of infection

  └ The outbreak in the 90's

## Example: Keiding's model for Rubella in the UK

- Prevalence: Isotonic regression for rubella in the UK.



└ Estimating the force of infection

└ The outbreak in the 90's

# Keiding's model in R

## ■ Force of infection.

$$\hat{\lambda}(a) = \frac{1}{h} \int_{a-h}^{a+h} K\left(\frac{x-a}{h}\right) \frac{d\hat{\pi}(x)}{1 - \hat{\pi}(x^-)},$$

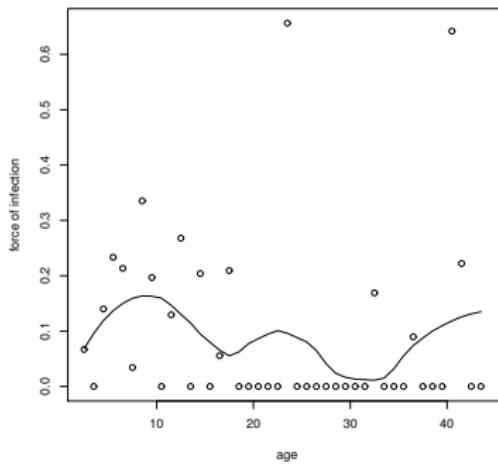
```
> d.iso1<-diff(iso1)
> foi.iso<-d.iso1/(1-isol[-c(1)])
> plot(xi[-c(1)],foi.iso)
> foi.iso1<-foi.iso[-c(43)]
> fit.foi<-loess(foi.iso1~xi[-c(1,44)],span=0.5)
> lines(xi[-c(1,44)],fit.foi$fit)
```

└ Estimating the force of infection

  └ The outbreak in the 90's

## Example: Keiding's model for Rubella in the UK

### ■ Force of infection for rubella in the UK



└ Estimating the force of infection

  └ The outbreak in the 90's

## Flexible models

- Choice between parametric and semi-or non-parametric models
  - parametric models are more powerful
  - semi- or non-parametric models allow the data to speak for themselves
- Example:
  - fractional polynomials
  - P-splines with shape-constraints

└ Estimating the force of infection

└ The outbreak in the 90's

# First order fractional polynomials

- Linear predictor:  $g(\pi(a)) = \eta(a)$ . With

$$\eta_1(a; \theta, p) = \theta_0 + \theta_1 H(a).$$

- $H(a)$  is a transformation given by

$$H(a) = \begin{cases} a^p & \text{if } p \neq 0 \\ \log(a) & \text{if } p = 0 \end{cases}$$

- $\eta(a)$ :

$$\eta(a) = \begin{cases} \theta_0 + \theta_1 a^p & \text{if } p \neq 0 \\ \theta_0 + \theta_1 \log(a) & \text{if } p = 0 \end{cases}$$

## Higher order fractional polynomials

- For a given degree  $m$  and a variable  $a > 0$ , fractional polynomials (Royston and Altman, 1994) are defined as

$$\eta_m(a; \theta, p) = \sum_{i=0}^m \theta_i H_i(a),$$

- $\theta = (\theta_0, \dots, \theta_m)$ : regression parameters,
- $p = (p_1, \dots, p_m)$ : vector of powers  $p_1 \leq \dots \leq p_m$ , positive or negative integers/fractions and  $H_i(a)$  a transformation:

$$H_i(a) = \begin{cases} a^{p_i} & \text{if } p_i \neq p_{i-1} \\ H_{i-1}(a) \times \log a & \text{if } p_i = p_{i-1} \end{cases}$$

with  $p_0 \equiv 0$  and  $H_0 \equiv 1$ .

- $p_i \in \{-2, -1, -0.5, 0, 0.5, 2, \dots, \max(3, m)\}$ ,  $i = 1, \dots, m$ .

## Model selection

- The deviance criterion:  $D(m, p) = -2\ell(\theta, p, m)$  assesses goodness of fit,
- Model selection for the best FP is based on the deviance criterion:
  - FPs of the same degree  $m$  are chosen by minimizing the deviance criterion or equivalently using the AIC-criterion,
  - $D(m, p) - D(m + 1, p) > \chi^2_{2;0.90}$  is used to decide for a  $m + 1$  degree FP,
- Usually, FPs of degree  $m = 2$  are considered sufficiently flexible

└ Estimating the force of infection

└ The outbreak in the 90's

# Fractional polynomials

- Existing models as a fractional polynomial:

| Publication                  | Fractional polynomial                 | Link function | Force of infection |
|------------------------------|---------------------------------------|---------------|--------------------|
| Muench (1934)                | $\eta(m = 1, p = 0, \theta_1 = 1)$    | cloglog       | constant           |
| Griffiths (1974)             | $\eta(m = 1, p = 0, \theta_1 = 2)$    | cloglog       | linear             |
| Grenfell and Anderson (1985) | $\eta(m = k, p_i = i)$                | log           | polynomial         |
| Becker (1989)                | $\eta(m = 1, p = 0, \theta_1 \neq 0)$ | cloglog       | monotone           |

└ Estimating the force of infection

  └ The outbreak in the 90's

## Non-and semi-parametric models

- Several nonparametric smoothers, relaxing on assumptions typically made using parametric models, have been proposed:
  - Keiding (1991) isotonic regression (Barlow et al., 1972): *kernel smoother*
  - Rossini and Tsiatis (1996): *semi-parametric approach (step function)*
  - Shkedy et al. (2003): *local polynomials*
  - Shiboski (1998); Hens et al. (2008): *P-splines*

└ Estimating the force of infection

  └ The outbreak in the 90's

## Non-and semi-parametric models

- In the generalized linear model framework
- a  $p$ -th degree P-spline is given by taking the linear predictor

$$\eta(x) = \theta_0 + \theta_1 x + \dots + \theta_p x^p + \sum_{k=1}^K b_k (x - \kappa_k)_+^p,$$

where

- $\kappa_k, k = 1, \dots, K$  are the knots which together with  $p$  determine the smoothness,
- $u_+ = u$  if  $u \geq 0$  and  $u_+ = 0$  if  $u < 0$ .

# Non-and semi-parametric models

## ■ P-splines

- fitting a P-spline as such will result in an unsMOOTH/rough fit
  - therefore one can put a restriction on  $\{b_k\}_k$ :
    - one penalizes roughness of the fit by putting a constraint on  $\{b_k\}_k$
- $$b_k \sim \mathcal{N}(0, \sigma_b^2)$$
- so using a mixed model approach one can fit penalized splines.
  - the splines here are truncated power splines
  - other splines exist: radial splines, ...

(Ruppert, Wand and Carroll, 2003)

# Non-and semi-parametric models in R

- Overview of the different smoothing methods and their basis function, knot selection, penalty and smoothing parameter selection method.

| Method                                                   | Basis Function                       | Knots                                | Penalty                  | Smoothing Parameter Selection |
|----------------------------------------------------------|--------------------------------------|--------------------------------------|--------------------------|-------------------------------|
| Smoothing Splines<br>( <i>gam-library</i> )              | cubic                                | all grid points                      | second order derivative  | manual                        |
| B-splines<br>( <i>pspline.fit</i> )                      | differences of truncated polynomials | user-defined ( <i>ps.intervals</i> ) | difference penalty       | manual                        |
| Cubic Regression Splines<br>( <i>mgcv-library</i> )      | cubic                                | user-defined ( <i>mgcv-default</i> ) | second order derivative  | automated                     |
| Thin-plate Regression Splines<br>( <i>mgcv-library</i> ) | thin plate                           | none                                 | eigenvalue decomposition | automated                     |

- └ Estimating the force of infection
  - └ The outbreak in the 90's

## P-splines

```
source("monotone psplinefit.R")
fit<-mpspline.fitter(response=y,x.var=a,ps.intervals=20,degree=3,
order=2,link="logit",family="binomial",alpha=50,kappa=0)
```

└ Estimating the force of infection

└ The outbreak in the 90's

## Applied to Parvovirus B19

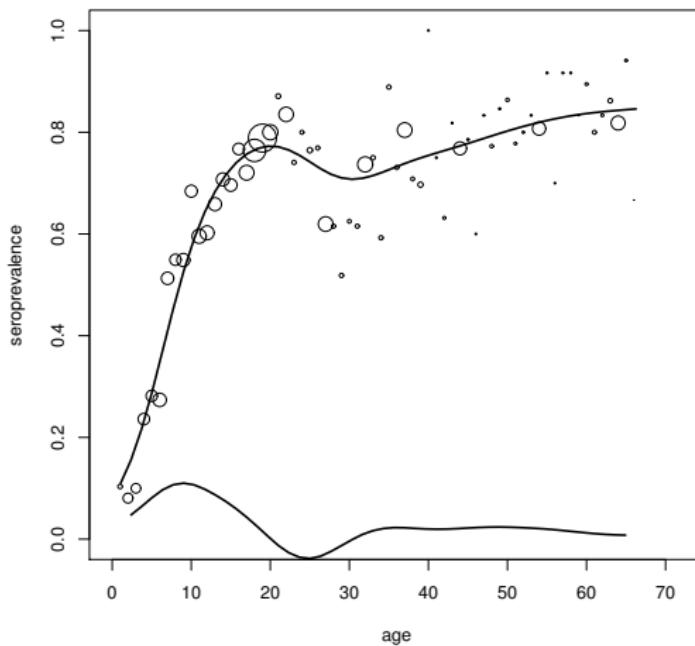


Figure: P-splines applied to seroprevalence data on parvovirus B19 in Belgium.

## The issue of monotonicity

- estimating the force of infection
  - relies on the steady state assumption
  - untestable in case of one cross-sectional sample (Keiding, 1991)
  - should result in a positive estimate
  - or equivalently a monotonically increasing prevalence estimate

└ Estimating the force of infection

  └ The issue of monotonicity

## The issue of monotonicity

- monotonicity:  $\forall a : \lambda(a) \geq 0 \Leftrightarrow \pi'(a) \geq 0$

several options:

- monotone functions
  - constrained optimization
  - select only monotone fits
  - smooth then constrain
- there is no infection at birth so:  $\pi(0) = 0$

## The issue of monotonicity: parametric models

- Muench (1934):  $\theta \geq 0$
- Griffiths (1974):  $\theta_1 \geq 0$  and  $\theta_0 \geq -\tau$
- Grenfell and Anderson (1985):

$$\sum_i \theta_i i a^{i-1} \geq 0, \forall a, \text{ and } \theta_0 = 0$$

ex:  $\theta_1 a + \theta_2 a^2 \rightarrow \theta_1 + 2\theta_2 a \geq 0$

- Becker (1989):
  - $\mu, \alpha \geq 0$
  - $\theta_i \geq 0, \forall a, a_{i-1} \leq a < a_i$
- Farrington (1990):  $\alpha_1, \alpha_3 \geq 0$

## The issue of monotonicity: parametric models in R

- constrained optimization - monotone function:
  - use a reparametrization  $\zeta = \exp(\tilde{\zeta})$
  - optimize for  $\tilde{\zeta}$

## The issue of monotonicity: non/semi-parametric models

- P-splines:
  - computationally efficient methods for constrained optimization
  - a 'smooth then constrain'-methodology
  - extra penalty for non-monotone behaviour
- the 'Pool Adjacent Violator'-algorithm (Barlow et al., 1972)

## The issue of monotonicity: B19

- Bollaerts et al. (2006): P-spline regression with shape constraints

```
source("monotone psplinefit.R")
fit<-mpspline.fitter(response=y,x.var=a,ps.intervals=20,degree=3,
                      order=2,link="logit",family="binomial",alpha=50,kappa=1E5)
```

└ Estimating the force of infection

└ The issue of monotonicity

## The issue of monotonicity: B19

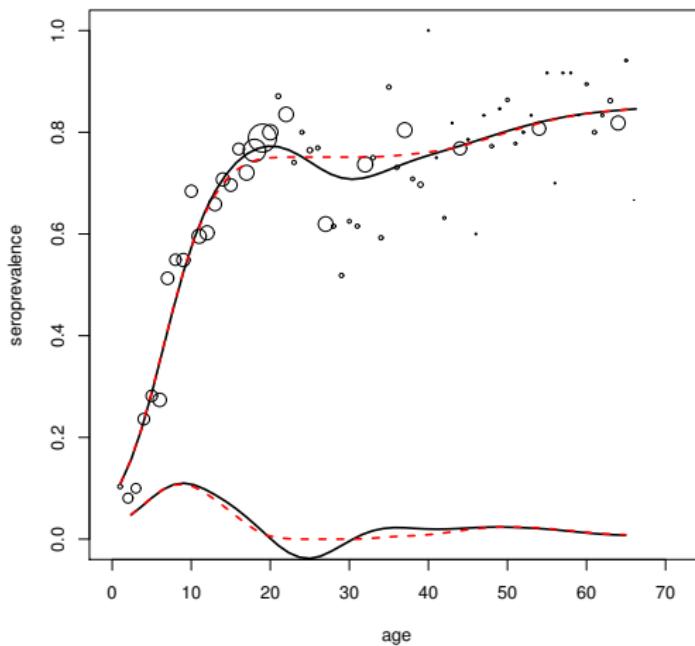


Figure: P-splines applied to seroprevalence data on parvovirus B19 in Belgium:  
black line: unconstrained fit; red line: constrained fit.

# Historical overview

- Historical perspective Hens et al. (2010):

*Epidemiol. Infect.* (2010), 138, 802–812. © Cambridge University Press 2009  
doi:10.1017/S0950268809990781

---

## Seventy-five years of estimating the force of infection from current status data

---

N. HENS<sup>1,2\*</sup>, M. AERTS<sup>1</sup>, C. FAES<sup>1</sup>, Z. SHKEDY<sup>1</sup>, O. LEJEUNE<sup>2</sup>, P. VAN DAMME<sup>2</sup>  
AND P. BEUTELS<sup>2</sup>

# A guide to modelling the force of infection

## ■ What to do with so many different models?

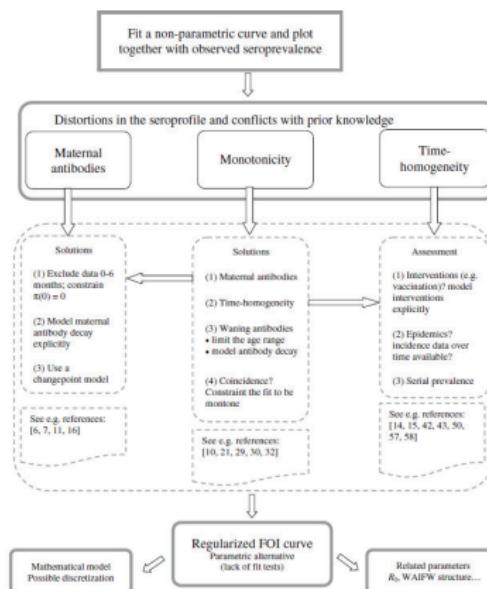


Fig. 3. Flow chart of a practical guide to estimate the force of infection (FOI) from seroprevalence data with reference to the literature on what to do and how to do it.

## Part III

Estimating Infectious Disease  
Parameters: transmission parameters

# Overview

## 8 The Mass Action Principle and $R_0$

- Concepts
- Some maths

## 9 Who Acquires Infection From Whom?

- The traditional approach
- A parsimonious contact surface

## 10 The social contact approach

- The social contact hypothesis
- Refinements
- Parvovirus B19: Conclusion

## 11 Summary

# The force of infection and the mass action principle

The force of infection  $\lambda(a)$  can be approximated by

$$\lambda(a) = D \int_0^{\infty} \beta(a, a') \lambda(a') S(a') da',$$

(Anderson and May, 199)

- where
  - $\beta(a, a')$ : the transmission rate:  
'the per capita rate at which an individual of age  $a'$  makes effective contact with a person of age  $a$  per year'
  - $S(a')$ : the number of susceptible persons of age  $a'$
  - $D\lambda(a')S(a')$ : the number of infectious individuals of age  $a'$

## The force of infection and the mass action principle

The force of infection  $\lambda(a)$  can be approximated by

$$\lambda(a) = \int_0^{\infty} \beta(a, a') I(a') da',$$

(Anderson and May, 199)

- assumptions:
  - endemic equilibrium
  - short mean duration of infectiousness  $D$
  - infectious and susceptible individuals mix completely

# An MSIR-model

- Assume an MSIR-model:
  - first phase: maternally derived immunity
  - second phase: susceptible to infection
  - third phase: infected & infectious
  - fourth phase: immune/recovered
- Further:
  - closed population of size  $N$
  - demographic and endemic equilibrium
- Cohort model

## An MSIR-model

- The proportion susceptibles is given by

$$s(a) = \left[ \int_0^a \gamma(u) \exp \left( \int_0^u \lambda(v) - \gamma(v) dv \right) du \right] \exp \left( - \int_0^a \lambda(u) du \right),$$

where

- $\gamma(a)$  is the maternal antibody decay rate
- $\lambda(a)$  is the force of infection

## An MSIR-model with type I maternal antibodies

- Without loss of generality, we assume a prompt loss of maternal immunity at age  $A$ :

$$m_{pi}(a) = \exp\left(-\int_0^a \gamma(u)du\right) = \begin{cases} 1, & \text{if } a \leq A \\ 0, & \text{if } a > A, \end{cases} \quad (1)$$

- This is known as 'Type I maternal antibody assumption':  
"newborns are protected by maternal antibodies until age  $A$  after which they instantaneously move to the susceptible class"
- In practice,  $A = 0.5$  years (6 months) is a common choice

## An MSIR-model with type I maternal antibodies

- The proportion of susceptibles becomes

$$s(a) = \exp\left(-\int_A^a \lambda(u)du\right), \quad \text{if } a > A, \quad (2)$$

where

- $\lambda(a)$  denotes the age-specific force of infection
- $s(a) = 0$  if  $a \leq A$ .

└ The Mass Action Principle and  $R_0$

└ Concepts

## So far . . .

- $\lambda(a)$  can be estimated
- $s(a')$  can be estimated
- $S(a')$  ?

## Estimating the mortality rate

- fixed total population size  $N$
- age-specific mortality rate  $\mu(a)$
- under demographic equilibrium, we have

$$\frac{dN(a)}{da} = -\mu(a)N(a),$$

and thus

$$N(a) = N(0) \exp \left\{ - \int_0^a \mu(u) du \right\}$$

## Estimating the mortality rate

- Define the survivor function

$$m(a) = \exp\left\{-\int_0^a \mu(u)du\right\}.$$

- The age density is then given by

$$L^{-1} \exp\left\{-\int_0^a \mu(u)du\right\} = L^{-1} m(a),$$

where  $L$  is the life expectancy

$$L = \int_0^{+\infty} m(u)du$$

## Estimating the mortality rate in R

- data on mortality from Belgium anno 2006:
  - the number of deaths  $ND$ ,
  - per integer age value  $AGE$  and
  - the population size  $PS$
- Poisson model with log-link and offset term

$$\log(ND) = \log(PS) + f(AGE)$$

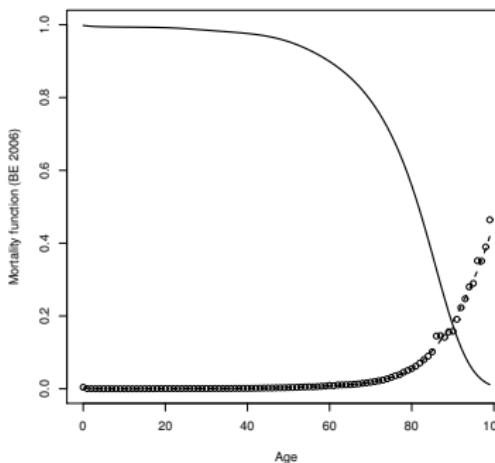
- maximize the corresponding likelihood to estimate  $f$
- ensuring flexibility can be done using a thin plate regression spline

# Estimating the mortality rate

- *gam*-function in the R-package *mgcv*:

```
library(mgcv)
demfit<-gam(ND~s(AGE),offset=log(PS),family="poisson",link="log")
muy<-predict(demfit,type="response")
my<-exp(-cumsum(muy))
L<-sum(my)
```

- The estimated mean life expectancy  $\hat{L} = 78.8$  years.
- Figure: The estimated survivor function (solid line) and mortality rate (dashed line) with observed death rate (dots) for Belgium anno 2006.



# The mass action principle

- $S(a')$  can be expressed as

$$S(a') = L^{-1} N s(a') m(a'),$$

where

- $s(a')$ : the proportion susceptible
- $m(a')$ : the survivor function at age  $a'$
- Rewrite the mass action principle under Type I maternal antibodies as

$$\lambda(a) = \frac{ND}{L} \int_A^{\infty} \beta(a, a') \lambda(a') s(a') m(a') da',$$

for  $a \geq A$ .

Diekmann et al. (1990); Farrington et al. (2001)

# The next generation operator

- The next generation operator:

“the operator producing the next generation of infected individuals”

$$G[](a) = \frac{ND}{L} m(a) \beta(a, a').$$

- The basic reproduction number  $R_0$ :

“the average number of secondary infections produced by a typical infectious individual during his entire infectious period when introduced into a completely susceptible population”

$$\max(\text{eigenvalue}(G[](a))).$$

## The next generation operator

- The effective reproduction number  $R$ :

“the average number of secondary infections produced by a typical infectious individual during his entire infectious period when introduced into a population”

$$\max(\text{eigenvalue} \left( \frac{ND}{L} m(a)s(a)\beta(a, a') \right)).$$

## Some Maths . . .

- $R_0$  is the dominant eigenvalue of the next generation operator:

$$R_0 \ell(a') = \frac{ND}{L} \int_A^{+\infty} \ell(a) m(a) \beta(a, a') da,$$

where  $\ell(a)$  denotes the leading left eigenfunction of  $G[](a)$ .

- After some calculus we obtain:

$$R_0 = \frac{\int_A^{\infty} \ell(a) \lambda(a) m(a) da}{\int_A^{\infty} \ell(a) \lambda(a) s(a) m(a) da}.$$

- Given  $\lambda(a)$ ,  $R_0$  can only be estimated when  $\ell(a)$  is identifiable.
- $\ell(a)$  depends on  $\beta(a, a')$  which is unknown.

## Basic mixing assumptions

- Homogeneous mixing:

$$\beta(a, a') = \beta \quad \forall a, a'$$

Consequently  $R_0 = L/A^*$  where

- $L$  is the life expectancy since loss of maternal immunity
- $A^*$  is the average time to removal by infection or death since loss of maternal immunity

## Basic mixing assumptions

- Separable mixing:

$$\exists u, v : \beta(a, a') = u(a)v(a'),$$

Consequently  $I(a') \propto v(a')$  and  $\lambda(a) \propto u(a)$  but not sufficient for  $R_0$  to be identifiable.

- Proportional mixing:

$$\exists u : \beta(a, a') = u(a)u(a').$$

$$R_0 \approx \frac{\int_A^\infty \lambda(a)^2 m(a) da}{\int_A^\infty \lambda(a)^2 s(a)m(a) da}.$$

(Dietz and Schenzle, 1985; Hethcote and Van Ark, 1987)

└ The Mass Action Principle and  $R_0$

└ Some maths

## Basic mixing assumptions

- Susceptibility-dependent mixing:

$$\exists u : \beta(a, a') \propto u(a)$$

$$R_0 \approx \frac{\int_A^\infty \lambda(a)m(a)da}{\int_A^\infty \lambda(a)s(a)m(a)da}.$$

## Example: Parvovirus B19 in Belgium

| Mixing                            | $R_0$ | bootstrap-conf.int. |
|-----------------------------------|-------|---------------------|
| Homogeneous                       | 4.40  | (4.17,4.64)         |
| Proportional - gamma-shaped FOI   | 1.70  | (1.65,1.76)         |
| Susceptibility - gamma-shaped FOI | 1.91  | (1.85,1.99)         |
| Proportional - spline-based FOI   | 1.74  | (1.66,1.99)         |
| Susceptibility - spline-based FOI | 2.03  | (1.87,2.32)         |

- These mixing assumptions do not hold for parvovirus B19.

Hens et al. (2008)

- More general mixing patterns should be envisaged.

└ Who Acquires Infection From Whom?

└ The traditional approach

## Imposing Mixing Patterns

- The Mass Action Principle can be discretized into:

$$\lambda_i = \frac{ND}{L} \sum_{j=1}^J \beta_{ij} \frac{\lambda_j}{\lambda_j + \mu_j} (s_{j-1} - s_j),$$

where

- $s_j = \exp \left( - \sum_{k=1}^j (\lambda_k + \mu_k) (a_{[k+1]} - a_{[k]}) \right)$  represents the proportion susceptible individuals in age-class  $j$ .
- $\lambda_i$ : FOI for age class  $i = 1, \dots, J$
- $D$ : the mean duration of infectiousness
- $\beta_{ij}$ : per capita rate at which an individual of age class  $j$  makes an effective contact with a person of age class  $i$ , per year.

└ Who Acquires Infection From Whom?

  └ The traditional approach

## Imposing mixing patterns

- The transmission matrix  $\beta_{ij}$  with dimension  $J \times J$ : WAIFW-matrix
- System of  $J$  equations with  $J^2$  unknown parameters: not identifiable
- $J$  usually small: 3 to 6
- age-classes  $\sim$  mixing groups
- imposed WAIFW-matrix  $\rightarrow$  high impact on the estimate of  $R_0$ .

[Greenhalgh and Dietz \(1994\)](#)

└ Who Acquires Infection From Whom?

└ The traditional approach

## Imposing mixing patterns: Varicella Zoster Virus

Six structures:  $W_1, W_2, W_3, W_4, W_5, W_6$ :

$$W_1 = \begin{pmatrix} \beta_1 & \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 \\ \beta_6 & \beta_2 & \beta_6 & \beta_6 & \beta_6 & \beta_6 \\ \beta_6 & \beta_6 & \beta_3 & \beta_6 & \beta_6 & \beta_6 \\ \beta_6 & \beta_6 & \beta_6 & \beta_4 & \beta_6 & \beta_6 \\ \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_5 & \beta_6 \\ \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 \end{pmatrix}$$

(Anderson and May, 1991; Van Effelterre et al., 2009; Ogunjimi et al., 2009; Goeyvaerts et al., 2010):

└ Who Acquires Infection From Whom?

└ The traditional approach

## Imposing mixing patterns: Varicella Zoster Virus

Six structures:  $W_1, W_2, W_3, W_4, W_5, W_6$ :

$$W_2 = \begin{pmatrix} \beta_1 & \beta_1 & \beta_3 & \beta_4 & \beta_5 & \beta_6 \\ \beta_1 & \beta_2 & \beta_3 & \beta_4 & \beta_5 & \beta_6 \\ \beta_3 & \beta_3 & \beta_3 & \beta_4 & \beta_5 & \beta_6 \\ \beta_4 & \beta_4 & \beta_4 & \beta_4 & \beta_5 & \beta_6 \\ \beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_6 \\ \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 \end{pmatrix}$$

(Anderson and May, 1991; Van Effelterre et al., 2009; Ogunjimi et al., 2009; Goeyvaerts et al., 2010):

└ Who Acquires Infection From Whom?

└ The traditional approach

## Imposing mixing patterns: Varicella Zoster Virus

Six structures:  $W_1, W_2, W_3, W_4, W_5, W_6$ :

$$W_3 = \begin{pmatrix} \beta_1 & \beta_1 & \beta_1 & \beta_4 & \beta_5 & \beta_6 \\ \beta_1 & \beta_2 & \beta_3 & \beta_4 & \beta_5 & \beta_6 \\ \beta_1 & \beta_3 & \beta_3 & \beta_4 & \beta_5 & \beta_6 \\ \beta_4 & \beta_4 & \beta_4 & \beta_4 & \beta_5 & \beta_6 \\ \beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_6 \\ \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 \end{pmatrix}$$

(Anderson and May, 1991; Van Effelterre et al., 2009; Ogunjimi et al., 2009; Goeyvaerts et al., 2010):

└ Who Acquires Infection From Whom?

└ The traditional approach

## Imposing mixing patterns: Varicella Zoster Virus

Six structures:  $W_1, W_2, W_3, W_4, W_5, W_6$ :

$$W_4 = \begin{pmatrix} \beta_1 & \beta_1 & \beta_1 & \beta_1 & \beta_1 & \beta_1 \\ \beta_2 & \beta_2 & \beta_2 & \beta_2 & \beta_2 & \beta_2 \\ \beta_3 & \beta_3 & \beta_3 & \beta_3 & \beta_3 & \beta_3 \\ \beta_4 & \beta_4 & \beta_4 & \beta_4 & \beta_4 & \beta_4 \\ \beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_5 \\ \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 \end{pmatrix}$$

(Anderson and May, 1991; Van Effelterre et al., 2009; Ogunjimi et al., 2009; Goeyvaerts et al., 2010):

└ Who Acquires Infection From Whom?

└ The traditional approach

## Imposing mixing patterns: Varicella Zoster Virus

Six structures:  $W_1, W_2, W_3, W_4, W_5, W_6$ :

$$W_5 = \begin{pmatrix} \beta_1 & \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 \\ \beta_6 & \beta_2 & \beta_6 & \beta_6 & \beta_6 & \beta_6 \\ \beta_6 & \beta_6 & \beta_3 & \beta_6 & \beta_6 & \beta_6 \\ \beta_6 & \beta_6 & \beta_6 & \beta_4 & \beta_6 & \beta_6 \\ \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_5 & \beta_6 \\ \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_5 \end{pmatrix}$$

(Anderson and May, 1991; Van Effelterre et al., 2009; Ogunjimi et al., 2009; Goeyvaerts et al., 2010):

└ Who Acquires Infection From Whom?

└ The traditional approach

## Imposing mixing patterns: Varicella Zoster Virus

Six structures:  $W_1, W_2, W_3, W_4, W_5, W_6$  :

$$W_6 = \begin{pmatrix} \beta_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \beta_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \beta_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & \beta_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & \beta_5 & 0 \\ 0 & 0 & 0 & 0 & 0 & \beta_6 \end{pmatrix}.$$

(Anderson and May, 1991; Van Effelterre et al., 2009; Ogunjimi et al., 2009; Goeyvaerts et al., 2010):

└ Who Acquires Infection From Whom?

└ The traditional approach

## Regular and non-regular configurations

- Goal: estimating transmission parameters:

$$\boldsymbol{\beta} = (\beta_1, \dots, \beta_6)^T.$$

- There exists a unique matrix  $D(\boldsymbol{\lambda})$  such that  $\boldsymbol{\lambda} = D(\boldsymbol{\lambda})\boldsymbol{\beta}$ .
- Conditional on  $D(\boldsymbol{\lambda})$  being invertible, the parameter vector  $\boldsymbol{\beta}$  can be estimated from a piecewise constant FOI:

$$\hat{\boldsymbol{\beta}} = D(\hat{\boldsymbol{\lambda}})^{-1}\hat{\boldsymbol{\lambda}}.$$

- $D(\hat{\boldsymbol{\lambda}})^{-1}\hat{\boldsymbol{\lambda}} \geq 0 \rightarrow W$  regular configuration for the data.
- The deviance/likelihood is identical for all such configurations.
- Non-regular configurations: constraints  $D(\hat{\boldsymbol{\lambda}})^{-1}\hat{\boldsymbol{\lambda}} \geq 0$ .

- └ Who Acquires Infection From Whom?

- └ The traditional approach

## Regular and non-regular configurations

- Derivation of  $D(\lambda)$  for  $W_3$

$$\Psi_j = \frac{ND}{L} \frac{\lambda_j}{\lambda_j + \mu_j} (s_{j-1} - s_j)$$

for  $j = 1, \dots, J$ .

- $\lambda = D(\lambda)\beta$  can be rewritten as  $\lambda = W_3 \Psi^T$ .
- $D(\lambda)$ :

$$\left( \begin{array}{cccccc} \sum_{j=1}^3 \Psi_j & 0 & 0 & \Psi_4 & \Psi_5 & \Psi_6 \\ \Psi_1 & \Psi_2 & \Psi_3 & \Psi_4 & \Psi_5 & \Psi_6 \\ \Psi_1 & 0 & \Psi_2 + \Psi_3 & \Psi_4 & \Psi_5 & \Psi_6 \\ 0 & 0 & 0 & \sum_{j=1}^4 \Psi_j & \Psi_5 & \Psi_6 \\ 0 & 0 & 0 & 0 & \sum_{j=1}^5 \Psi_j & \Psi_6 \\ 0 & 0 & 0 & 0 & 0 & \sum_{j=1}^6 \Psi_j \end{array} \right)$$

- └ Who Acquires Infection From Whom?
  - └ The traditional approach

## Imposing mixing patterns: extension

- Until now: piecewise constant FOI
- Fractional polynomials  $\eta = \eta_m(a, \beta, p)$ :

$$\Psi_j = \frac{(1 + e^{\eta(a_0)})(e^{\eta(a_j)} - e^{\eta(a_{j-1})})}{(1 + e^{\eta(a_{j-1})})(1 + e^{\eta(a_j)})}.$$

Van Effelterre et al. (2009)

- Any function: sole condition is identifiability - numerical approximation.

└ Who Acquires Infection From Whom?

  └ The traditional approach

## An example: Parvovirus B19 in Belgium

- Consider
  - survivor function for Belgium anno 2006: population size 10 511 382
  - $D = 6/365$  years
  - Type I maternal antibodies with age  $A = 0.5$ .
  - Six age classes following the schooling system in Belgium:  $(0.5, 2)$ ,  $[2, 6)$ ,  $[6, 12)$ ,  $[12, 19)$ ,  $[19, 31)$ ,  $[31, +\infty)$ .

- └ Who Acquires Infection From Whom?
  - └ The traditional approach

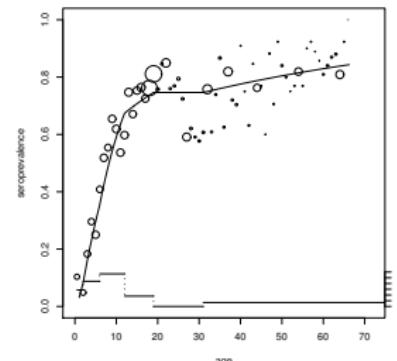
## R-code

- R-code to fit a piecewise constant FOI while ensuring a positive FOI
  - input: vector of binary values indicator of past infection; age-values or: number of individuals; age-values; and number of ind. with past infection

```
foi.fit<-pcwrate.fitter(y.var=y,x.var=a,breaks=breakpoints)
```

- $\hat{\lambda}^{\text{ML}} = (0.0576 \ 0.0878 \ 0.1138 \ 0.0359 \ 0.0000 \ 0.0137)^T$ .

Figure: The piecewise constant FOI and corresponding prevalence for parvovirus B19 in Belgium. The dots are the observed seroprevalence per integer age-value with size proportional to the samples taken.



└ Who Acquires Infection From Whom?

  └ The traditional approach

## R-code

- assessment of regularity and invertibility of  $D(\hat{\lambda})$ :

```
waifw<-waifw.6parms(foihat=foihat,muhat=muhat,  
breaks=breakpoints,N=N,D=D,Lmax=Lmax)
```

- fitting non-regular matrices:

```
waifw.6parms.irr<-waifw.fitter(a=a,y=y,waifw.choice,  
breaks=breakpoints,N=N,D=D,Lmax=Lmax)
```

- └ Who Acquires Infection From Whom?

- └ The traditional approach

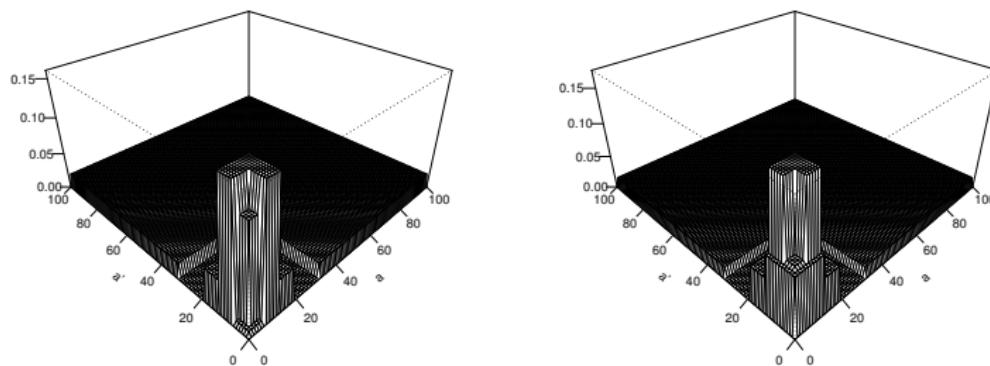
## Results

**Table:** Overview of WAIFW-matrices, AIC-values and estimated basic reproduction numbers for the Belgian parvovirus B19.

| WAIFW | Invertible     | Regular   | $\hat{R}_0$ | AIC     |
|-------|----------------|-----------|-------------|---------|
| 1     | Not Invertible | -         | -           | -       |
| 2     | Invertible     | Irregular | 2.00        | 3454.44 |
| 3     | Invertible     | Irregular | 1.89        | 3463.53 |
| 4     | Invertible     | Regular   | 2.33        | 3452.11 |
| 5     | Invertible     | Regular   | 14.80       | 3452.11 |
| 6     | Not Invertible | -         | -           | -       |

## └ Who Acquires Infection From Whom?

## └ The traditional approach



**Figure:** Perspective plots of the irregular WAIFW matrices  $W_2$  and  $W_3$  fitted under constrained optimization for parvovirus B19 in Belgium.

## └ Who Acquires Infection From Whom?

## └ The traditional approach

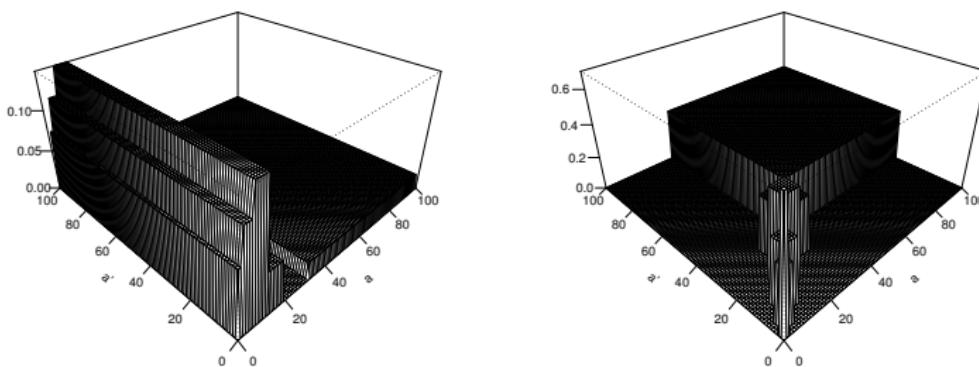


Figure: Perspective plots of the regular WAIFW matrices  $W_4$  and  $W_5$  for parvovirus B19 in Belgium.

- └ Who Acquires Infection From Whom?

- └ A parsimonious contact surface

## A parsimonious contact surface

- low-dimensional bivariate parametric model  
[Farrington and Whitaker \(2005\)](#)

- let  $u = (a + a')/\sqrt{2}$  and  $v = (a - a')/\sqrt{2}$ :

$$\beta(a, a') = \kappa(\gamma(u) \times b(v|u) + \delta),$$

where

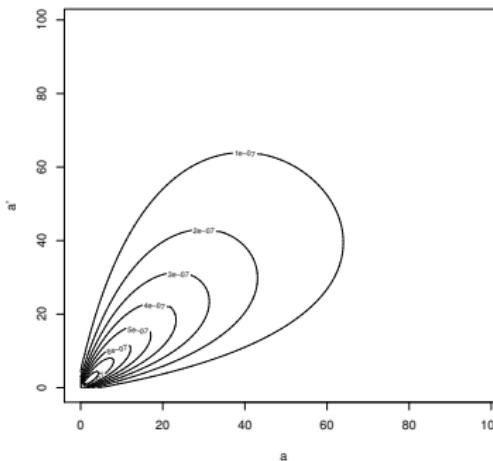
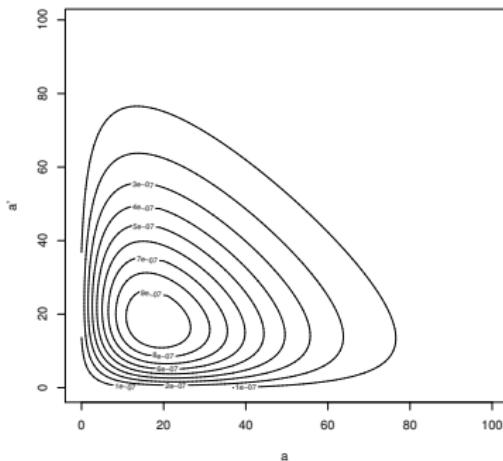
$$\begin{aligned}\gamma(u; \mu, \nu) &= c^{-1} u^{\nu-1} \exp\left(-\frac{\nu u}{\sqrt{2}\mu}\right), \\ b(v|u; \alpha, \beta) &= \frac{(u+v)^{\alpha-1}(u-v)^{\beta-1}}{u^{\alpha+\beta-2}},\end{aligned}$$

with  $c = \{\sqrt{2}\mu(1 - 1/\nu)\}^{\nu-1} e^{1-\nu}$ ,  $\alpha = \beta$ ;  $\gamma = (\alpha + \beta + 1)^{-1}$  and  $\sigma = \nu^{-2}$

- five parameters  $\kappa, \mu, \sigma, \gamma$  and  $\delta$  to estimate.

- └ Who Acquires Infection From Whom?
  - └ A parsimonious contact surface

## A parsimonious contact surface



**Figure:** Contourplot of two surface examples. For the first surface (left panel) the parameters were chosen as ( $\kappa = 1e - 6$ ,  $\mu = 25$ ,  $\sigma = 1$ ,  $\gamma = 0.1$ ,  $\delta = 0$ ), whereas for the second surface (right panel) parameters were chosen as ( $\kappa = 1e - 6$ ,  $\mu = 25$ ,  $\sigma = 0.5$ ,  $\gamma = 0.2$ ,  $\delta = 0$ ).

└ Who Acquires Infection From Whom?

└ A parsimonious contact surface

## R-code

- R-code

```
# Function code to fit surface model  
surface.fitter(a=a,y=y,muy=muy,Lmax=Lmax,N=N,D=D,plots="TRUE")
```

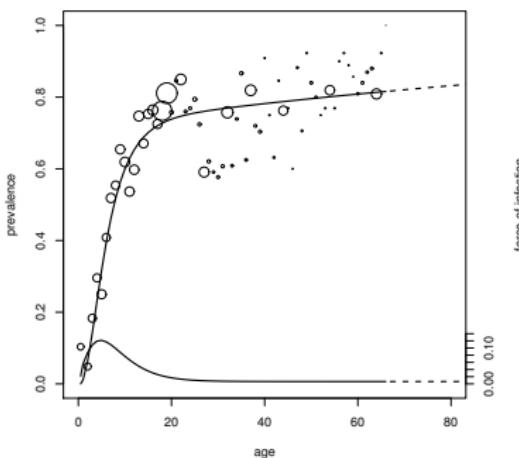
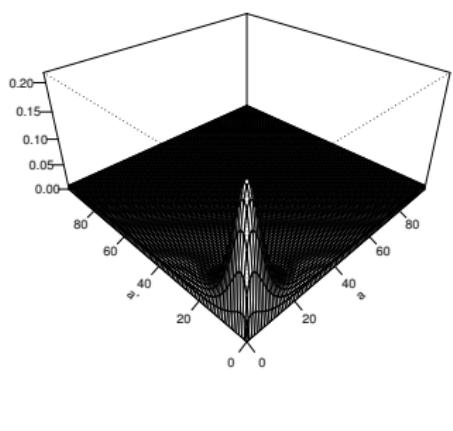
- Parvovirus B19:

- sensitive for starting values
- AIC=3461.00 (slightly higher than WAIFW)
- $\hat{R}_0 = 1.76$
- ML-estimates:  
 $\hat{\kappa} = 9.6e - 5$ ,  $\hat{\mu} = 7.65$ ,  $\hat{\sigma} = 0.50$ ,  $\hat{\gamma} = 0.12$  and  $\hat{\delta} = 0.04$ .
- Peak at 5 years of age

└ Who Acquires Infection From Whom?

└ A parsimonious contact surface

## An example: Parvovirus B19 in Belgium



**Figure:** Perspective plot of the fitted transmission surface (left panel) and resulting prevalence fit (right panel) for parvovirus B19 in Belgium.

└ Who Acquires Infection From Whom?

└ A parsimonious contact surface

## Topics in estimating WAIFW matrices

- Model Selection:

- AIC or BIC
- Bayes factors: [Farrington et al. \(2001\)](#)

- Lifelong Immunity

- $R_0$  can be approximated without estimating  $\beta(a, a')$  given data on a SIS infection with similar infection route is available
- age-specific incidence data on a newly emerging pathogen can provide the essential information to estimate  $R_0$  without estimating  $\beta(a, a')$ .

replacing  $I(a)$  by  $\lambda_{SIS}(a)$  or the age-specific relative incidence:

$$R_0 \approx \frac{\int_0^\infty \lambda_{SIS}(a) \lambda_{SIR} \exp\{-\int_0^a \mu(u) du\} da}{\int_0^\infty \lambda_{SIS}(a) \lambda_{SIR} \exp\{-\int_0^a \lambda_{SIR}(u) + \mu(u) du\} da}.$$

## Social Contact Survey

## Alternative approach:

Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents.

## Objectives

- Disentangle contact behaviour from transmission process Beutels et al. (2006)
  - Get insights in predictiveness of social contact data Mossong et al. (2008)
  - Get new insights in the transmission process Hens et al. (2009)

## Belgian social contact survey

- Part of POLYMOD project
- Period March - May 2006
- 750 participants, selected through random digit dialing
- Diary-based questionnaire
- Two main types of contact: non-close and close contacts
- Specific questions about contacts: location, duration, ...
- Total of 12775 contacts ( $\approx$  16 contacts per person per day)

└ The social contact approach

└ The social contact hypothesis

## The Social Contact Hypothesis

- social contact data: empirical evidence for mixing patterns  
[\(Wallinga et al., 2006\)](#)
- the social contact hypothesis:

$$\beta(a, a') = q \cdot c(a, a'),$$

with  $q$  a constant proportionality factor and  $c(a, a')$ , the per capita rate at which an individual of age  $a'$  makes contact with a person of age  $a$ , per year.

- └ The social contact approach
  - └ The social contact hypothesis

## Estimating contact rates

- $Y_{ij}$ : the number of contacts in age class  $j$  during one day as reported by a respondent in age class  $i$  ( $i, j = 1, \dots, J$ )
- $m_{ij} = E(Y_{ij})$ : the mean number of contacts in age class  $j$  during one day as reported by a respondent in age class  $i$  ( $i, j = 1, \dots, J$ )

→ 'social contact matrix'

- contact rates

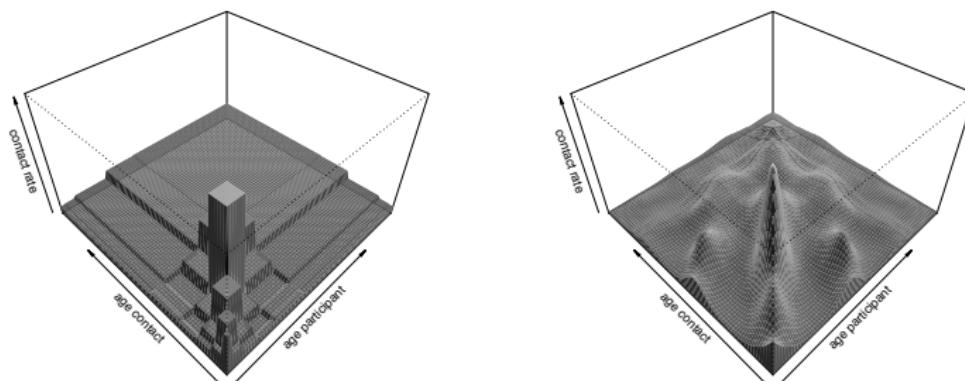
$$c_{ij} = 365 \cdot \frac{m_{ji}}{w_i},$$

where  $w_i$  denotes the population size in age class  $i$

- reciprocal nature of contacts:

$$m_{ij} w_i = m_{ji} w_j,$$

## Estimating contact rates for Belgium



**Figure:** Perspective plots of the estimated contact rates  $c_{ij}$  obtained with maximum likelihood estimation following Wallinga et al. (2006) (left panel) and Goeyvaerts et al. (2010) (right panel). The X- and Y-axis represent age of the respondent and age of the contact, respectively. Source: Goeyvaerts et al. (2010)

The social contact approach

The social contact hypothesis

## An example: Parvovirus B19

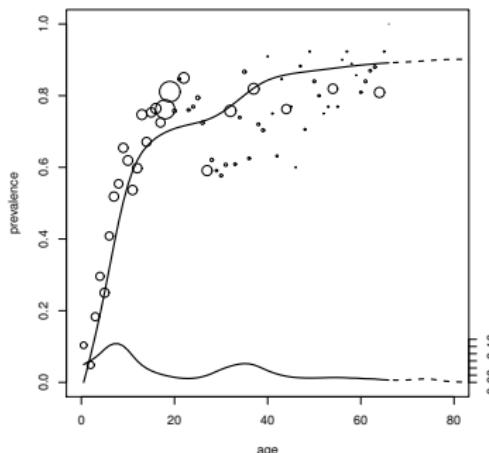
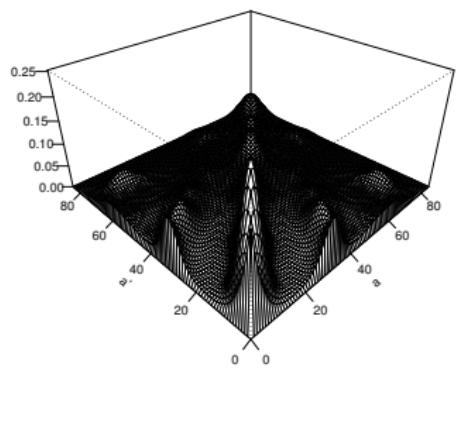
**Table:** Overview of various types of contacts contrasted to the Belgian parvovirus B19 serology using the social contact hypothesis.

| Contact type                | $\hat{q}$ | $\hat{R}_0$ | AIC     |
|-----------------------------|-----------|-------------|---------|
| All contacts                | 0.0320    | 3.99        | 3812.52 |
| Close contacts              | 0.0474    | 3.20        | 3637.01 |
| Close contacts > 15 minutes | 0.0536    | 2.84        | 3581.78 |
| Close contacts > 1 hour     | 0.0708    | 2.47        | 3531.63 |
| Close contacts > 4 hours    | 0.1030    | 2.19        | 3489.22 |

The social contact approach

The social contact hypothesis

## An example: Parvovirus B19



**Figure:** Perspective plot of the transmission surface for close contacts lasting at least 4 hours (left panel) together with the fitted FOI and prevalence (right panel).

## Refinements

- contacts with high transmission potential

$$\beta(a, a') = \sum_I q_I \cdot c_I(a, a'),$$

where  $I$  refers to the different stratification level of the contacts

- example: location 'home', 'school', 'work' and 'other'

- $\hat{q}_{ML} = (0.055 \ 0.048 \ 0.000 \ 0.000)^T$

$$\text{s.e.}(\hat{q}_{ML}) = (0.012 \ 0.009 \ 0.011 \ 0.047)^T$$

## Refinements

- age-dependent proportionality of the transmission rates

$$\beta(a, a') = q(a, a') \cdot c(a, a').$$

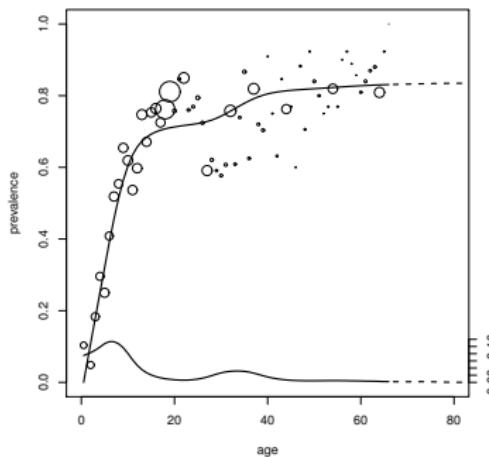
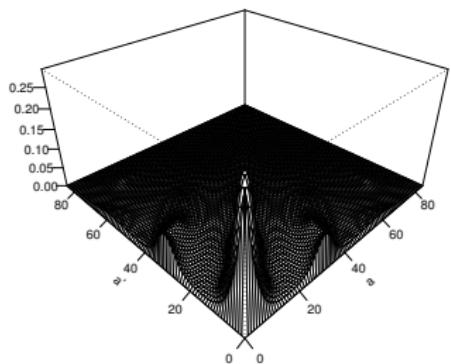
- identifiability needs to be assured:

$$\log\{q(a, a')\} = \gamma_0 + \gamma_1(a + a').$$

- The social contact approach

- Refinements

## An example: Parvovirus B19



**Figure:** Perspective plot of the transmission surface for close contacts lasting at least 4 hours with age-dependent proportionality factor (left panel) together with the fitted FOI and prevalence (right panel).

## Parvovirus B19: Conclusion

**Table:** Overview of the different fitted models to the parvovirus B19 data from Belgium with estimated  $R_0$ , AIC-value and Akaike Weight.

| Structure                              | $\hat{R}_0$ | AIC     | $w_k$ |
|----------------------------------------|-------------|---------|-------|
| WAIFW- $W_2$                           | 2.00        | 3454.44 | 0.131 |
| WAIFW- $W_3$                           | 1.89        | 3463.53 | 0.001 |
| WAIFW- $W_4$                           | 2.33        | 3452.11 | 0.420 |
| WAIFW- $W_5$                           | 14.80       | 3452.11 | 0.420 |
| Mixing surface                         | 1.76        | 3461.00 | 0.005 |
| All contacts                           | 3.99        | 3812.52 | 0.000 |
| Close contacts                         | 3.20        | 3637.01 | 0.000 |
| Close contacts > 15 minutes            | 2.84        | 3581.78 | 0.000 |
| Close contacts > 1 hour                | 2.47        | 3531.63 | 0.000 |
| Close contacts > 4 hours               | 2.19        | 3489.22 | 0.000 |
| Stratification of contacts by location | 2.68        | 3549.32 | 0.000 |
| Age-dependent proportionality factor   | 1.87        | 3457.96 | 0.023 |

# Summary

- We discussed how transmission parameters can be informed using empirical data
- These parameters can be used in mathematical models of infectious diseases

## Part IV

# Integrating Parameter Estimates into Mathematical Models

# Integrating Parameter Estimates into Models

- Static versus dynamic models
  - use the estimated equilibrium force of infection in your system of ODEs
  - use the estimated equilibrium transmission parameters as values in your system of ODEs with WAIFW or social contact hypothesis
- We will illustrate this for the Varicella (Zoster) Virus using a RAS model

The RAS-model consists of the following two-step iteration:  
Assuming one-year age-groups, let  $\{S_i(t), I_i(t), R_i(t)\}$  denote the number of susceptible, infected and recovered individuals of age  $i = 0, \dots, K - 1$  at time  $t$  (in years).

# Integrating Parameter Estimates into Models

**Step 1:** Given initial values

$\{S_i(t), I_i(t), R_i(t)\} = \{S_i(t_0), I_i(t_0), R_i(t_0)\}, i = 0, \dots, K - 1$  we solve the following set of ODEs:

$$\begin{cases} \frac{dS_i(t)}{dt} &= -(\lambda_i(t) + \mu_i)S_i(t), \\ \frac{dI_i(t)}{dt} &= \lambda_i(t)S_i(t) - (\nu + \mu_i)I_i(t), \\ \frac{dR_i(t)}{dt} &= \nu I_i(t) - \mu_i R_i(t), \end{cases}$$

to obtain  $\{S_i(t + 1), I_i(t + 1), R_i(t + 1)\}, i = 0, \dots, K - 1$  after one year.

# Integrating Parameter Estimates into Models

**Step 2:** Individuals are then shifted by one year:

- $\{S_i(t+1), I_i(t+1), R_i(t+1)\} \rightarrow \{S_{i+1}(t+1), I_{i+1}(t+1), R_{i+1}(t+1)\}, i = 0, \dots, K-2$
- all newborns  $B$  are assumed susceptible to infection:  
 $\{S_0(t+1), I_0(t+1), R_0(t+1)\} = \{B, 0, 0\}.$

This process is iterated throughout the time period of interest.

# Integrating Parameter Estimates into Models

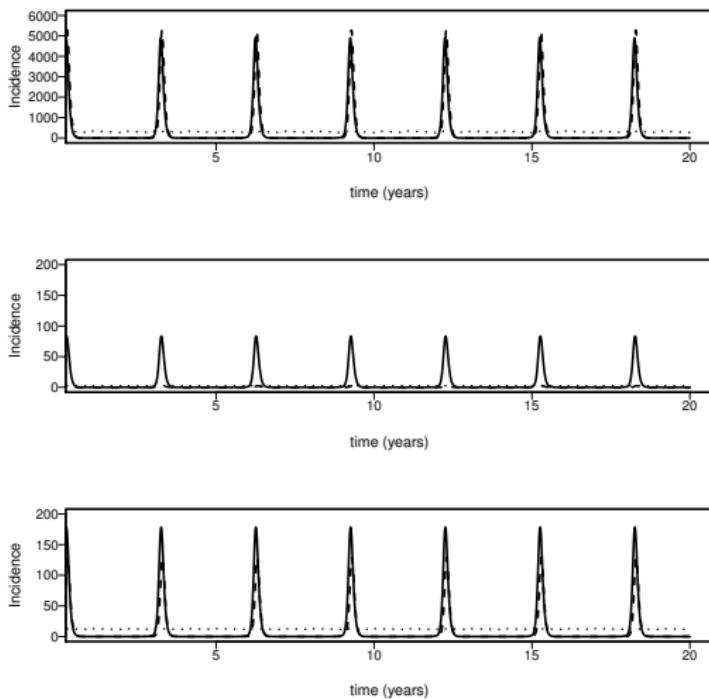
- We illustrate these approaches for varicella ignoring the herpes zoster-component.
- We assume  $\nu = 1/(7/365)$  corresponding to an infectious period of 7 days.
- We use  $\mu_i(t) \equiv \mu_i$  which we estimate from mortality data from Belgium anno 2005.
- First reach endemic equilibrium then turn to a simple vaccination strategy by putting  $S(0) = (1 - p)B$  and  $R(0) = pB$  for specific values of  $p$ .

This corresponds to immunizing newborns at birth with coverage probability  $p$ .

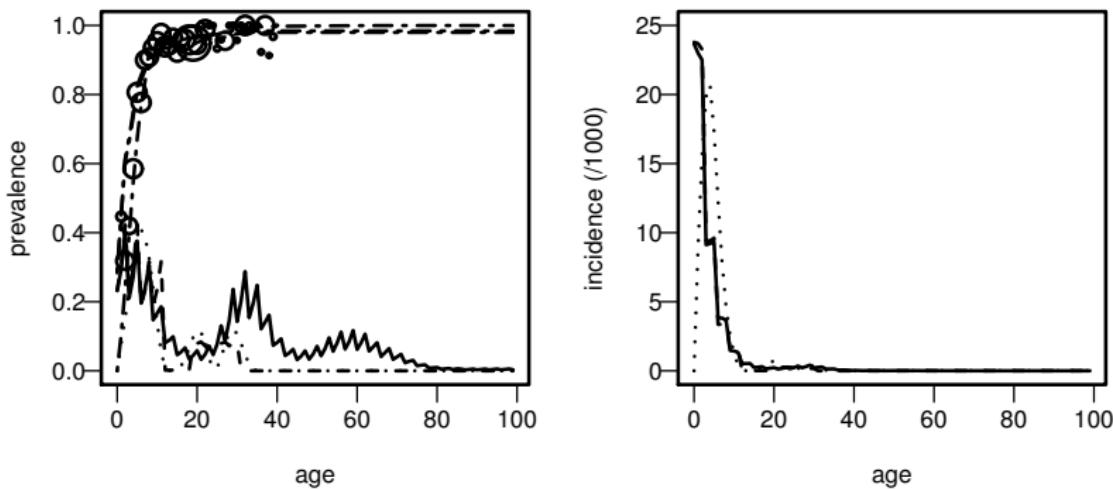
# Integrating Parameter Estimates into Models

- Force of infection:
  - For the static model, we use the FOI estimate based on a penalized spline FOI model
  - For the dynamic WAIFW model, we use WAIFW-matrix W2
  - For the dynamic social contact model we rely on the social contact hypothesis and use close contacts lasting at least 15 minutes

## └ Illustration by Example

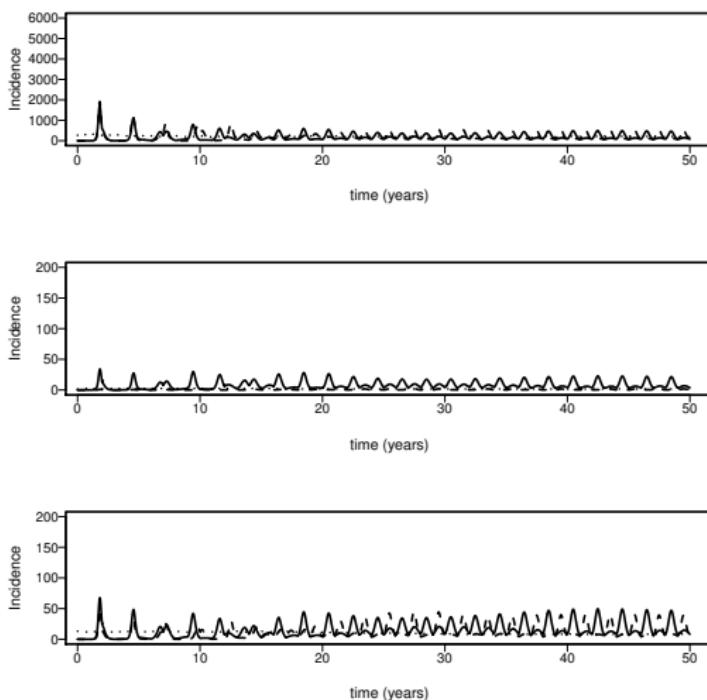


**Figure:** Number of infected individuals over 20 years for age-categories 0-12 years (first panel), 13-18 years (second panel) and 19 years and above (third panel) following the dynamic social contact data model (solid line), the dynamic WAIFW model (dashed line) and the static model (dotted line).



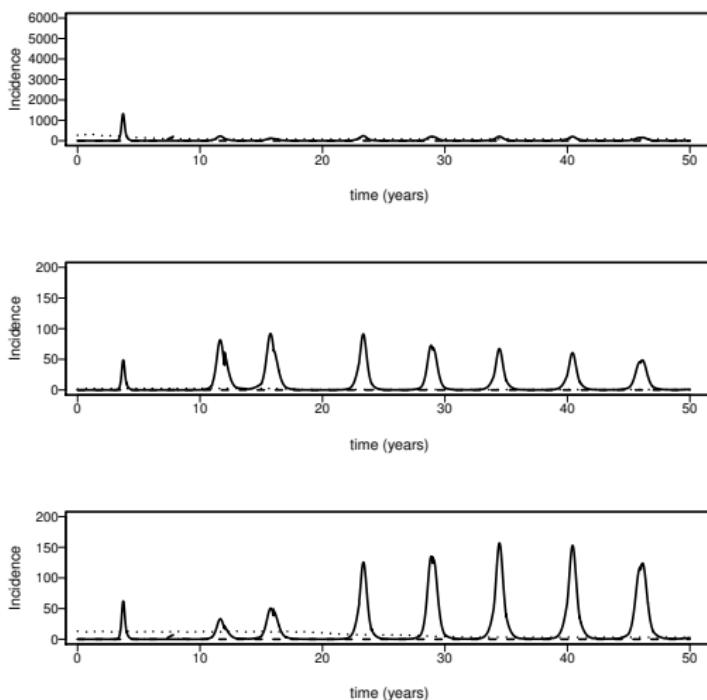
**Figure:** Prevalence, force of infection (left panel) and yearly incidence averaged over a period of three years (right panel) by age following the dynamic social contact data model (solid line), the dynamic WAIFW model (dashed line) and the static model (dotted line). The observed age-specific prevalence for VZV is shown in the left panel with size proportional to the number of samples taken.

- └ Illustration by Example



**Figure:** Impact of introducing an immunization programme for newborns with 33% coverage: Number of infected individuals for age-categories 0-12 years (first panel), 13-18 years (second panel) and 19 years and above (third panel) following the dynamic social contact data model (solid line), the dynamic WAIFW model (dashed line) and the penalized spline model (dotted line). The dynamic behavior is illustrated for the first 50 years after introducing vaccination.

## └ Illustration by Example



**Figure:** Impact of introducing an immunization programme for newborns with 75% coverage: Number of infected individuals for age-categories 0-12 years (first panel), 13-18 years (second panel) and 19 years and above (third panel) following the dynamic social contact data model (solid line), the dynamic WAIFW model (dashed line) and the penalized spline model (dotted line). The dynamic behavior is illustrated for the first 50 years after introducing vaccination.

## Summary

- Results are different depending on the chosen model
- Static model vs dynamic model: herd immunity
- Uncertainty is another key element

These elements have to be taken into consideration for the economic analyses

## Contact details

- Contact details:

- Niel Hens: niel.hens@uhasselt.be

- Acknowledgements:

- Nele Goeyvaerts
  - Ziv Shkedy: ziv.shkedy@uhasselt.be

- More in:

Hens, N., Shkedy, Z., Aerts, M., Faes, C., Van Damme, P. and Beutels, P.  
"Modeling Infectious Disease Parameters Based on Serological and Social  
Contact Data: A Modern Statistical Perspective"  
Springer-Verlag New York. 2012

## Selected Book references

- Becker (1989)
- Anderson and May (1992)
- Capasso (2008)
- Keeling and Rohani (2008)
- Vynnycky and White (2010)
- Halloran et al. (2010)
- Diekmann et al. (2012)
- ...

- Anderson, R. and May, R. (199). *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, Oxford.
- Anderson, R. and May, R. (1992). *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, Oxford.
- Bailey, N. (1975). *The Mathematical Theory of Infectious Diseases and its Applications*. Charles Griffin and Company, London.
- Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D. (1972). *Statistical Inference Under Order Restrictions*. Wiley, New York.
- Becker, N. G. (1989). *Analysis of infectious disease data*. London, Chapman and Hall.
- Beutels, P., Shkedy, Z., Aerts, M., and Van Damme, P. (2006). Social mixing patterns for transmission models of close contact infections: exploring self-evaluation and diary-based data collection through a web-based interface. *Epidemiology and Infection*, 134:1158–1166.
- Bollaerts, K., Eilers, P., and Van Mechelen, I. (2006). Simple and multiple P-spline regression with shape constraints. *British Journal of Mathematical and Statistical Psychology*, 59:451–469.
- Capasso, V. (2008). *Mathematical Structures of Epidemic Systems*. Springer.
- Daley, D. and Gani, J. (1999). *Epidemic Modelling: An Introduction*. Cambridge University Press.
- Diekmann, O., Heesterbeek, H., and Britton, T. (2012). *Mathematical Tools for Understanding Infectious Disease Dynamics*. Princeton University Press.

- Diekmann, O., Heesterbeek, J., and Metz, J. (1990). On the definition and the computation of the basic reproduction ratio  $R_0$  in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28:65–382.
- Dietz, K. and Schenzle, D. (1985). Proportionate mixing models for age-dependent infection transmission. *Journal of Mathematical Biosciences*, 22:117–120.
- Farrington, C., Kanaan, M., and Gay, N. (2001). Estimation of the basic reproduction number for infectious diseases from age-stratified serological survey data. *Applied Statistics*, 50:251–292.
- Farrington, C. and Whitaker, H. (2005). Contact surface models for infectious diseases: estimation from serologic survey data. *Journal of the American Statistical Association*, 100:370 – 379.
- Farrington, C. P. (1990). Modeling forces of infection for measles, mumps and rubella. *Statistics in Medicine*, 9:953–967.
- Goeyvaerts, N., Hens, N., Ogunjimi, B., Aerts, M., Shkedy, Z., Van Damme, P., and Beutels, P. (2010). Estimating infectious disease parameters from data on social contacts and serological status. *Journal of the Royal Statistical Society Series C*, 59:255–277.
- Goeyvaerts, N., Leuridan, E., Faes, C., Van Damme, P., and Hens, N. (2015). Multi-disease analysis of maternal antibody decay using non-linear mixed models accounting for censoring. *Statistics in Medicine*, 34(20):2858–2871.
- Greenhalgh, D. and Dietz, K. (1994). Some bounds on estimates for reproductive ratios derived from the age-specific force of infection. *Mathematical Biosciences*, 124:9–57.

- Grenfell, B. and Dobson, A. (1995). *Ecology of Infectious Disease in Natural Populations*. Cambridge, UK: Cambridge University Press.
- Grenfell, B. T. and Anderson, R. M. (1985). The estimation of age-related rates of infection from case notifications and serological data. *Journal of Hygiene*, 95(2):419–36.
- Griffiths, D. (1974). A catalytic model of infection for measles. *Applied Statistics*, 23:330–339.
- Halloran, E., Longini, I., and Struchiner, C. (2010). *Design and Analysis of Vaccine Studies*. Springer-verlag.
- Hens, N., Aerts, M., Faes, C., Shkedy, Z., Lejeune, O., Damme, P. V., and Beutels, P. (2010). Seventy-five years of estimating the force of infection from current status data. *Epidemiology and Infection*, 138(6):802–812.
- Hens, N., Aerts, M., Shkedy, Z., Theeten, H., Van Damme, P., and Beutels, P. (2008). Modelling multi-sera data: the estimation of new joint and conditional epidemiological parameters. *Statistics in Medicine*, 27:2651–2664.
- Hens, N., Goeyvaerts, N., Aerts, M., Shkedy, Z., Damme, P. V., and Beutels, P. (2009). Mining social mixing patterns for infectious disease models based on a two-day population survey in Belgium. *BMC Infectious Diseases*, 9:5.
- Hethcote, H. (2000). The mathematics of infectious diseases. *SIAM REview*, 42(4):599–653.

- Hethcote, H. and Van Ark, J. (1987). Epidemiological models for heterogeneous populations: proportionate mixing, parameter estimation and immunization programs. *Mathematical Biosciences*, 84:85–118.
- Keeling, M. and Rohani, P. (2008). *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press.
- Keiding, N. (1991). Age-specific incidence and prevalence: A statistical perspective (with discussion). *Journal of the Royal Statistical Society Series A*, 154:371–412.
- Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London, Series A*, 115:700–721.
- Kovac, T., Haber, T., Reeth, F. V., and Hens, N. (2018). Heterogeneous computing for epidemiological model fitting and simulation. *BMC Bioinformatics*, 19(1):101.
- Malthus, T. (1798). *An Essay on the Principle of Population*. Johnson, J.
- McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*. Chapman & Hall.
- Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Scalia Tomba, G., Wallinga, J., Heijne, J., Sadkowska-Todys, M., Rosinska, M., and Edmunds, J. (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Medicine*, 5(3):381–391.
- Muench, H. (1934). Derivation of rates from summation data by the catalytic curve , vol. 29, no. 185 (mar., 1934), pp. 25-38. *Journal of the American Statistical Association*, 29:25–38.

Ogunjimi, B., Hens, N., Goeyvaerts, N., Aerts, M., Damme, P. V., and Beutels, P. (2009). Using empirical social contact data to model person to person infectious disease transmission: an illustration for varicella. *Mathematical Biosciences*, 218(2):80–87.

Robertson, T., Wright, F., and Dykstra, R. (1988). *Order restricted statistical inference*. Wiley.

Rossini, A. and Tsiatis, A. (1996). A semiparametric proportional odds regression model for the analysis of current status data. *Journal of the American Statistical Society*, 91:713–721.

Royston, P. and Altman, D. (1994). Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling. *Applied Statistics*, 43(3):429–467.

Santermans, E., Robesyn, E., Ganyani, T., Sudré, B., Faes, C., Quinten, C., Van Bortel, W., Haber, T., Kovac, T., Van Reeth, F., Testa, M., Hens, N., and Plachouras, D. (2016). Spatiotemporal evolution of ebola virus disease at sub-national level during the 2014 west africa epidemic: Model scrutiny and data meagreness. *PLoS ONE*, 11(1):e0147172.

Shiboski, S. (1998). Generalized additive models for current status data. *Lifetime Data Analysis*, 4:29–50.

Shkedy, Z., Aerts, M., Molenberghs, G., Beutels, P., and Van Damme, P. (2003). Modelling forces of infection by using monotone local polynomials. *Applied Statistics*, 52(4):469–485.

- Van Effelterre, T., Shkedy, Z., Aerts, M., Molenberghs, G., Van Damme, P., and Beutels, P. (2009). Contact patterns and their implied basic reproductive numbers: an illustration for varicella-zoster virus. *Epidemiology and Infection*, 137:48–57.
- Verhulst, P. F. (1836). Notice sur la loi que la population suit dans son accroissement. *Corr. Mat. et Phys.*, 10:113–121.
- Vynnycky, E. and White, R. (2010). *An Introduction to Infectious Disease Modelling*. Oxford University Press.
- Wallinga, J., Teunis, P., and Kretzschmar, M. (2006). Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *American Journal of Epidemiology*, 164:936–944.