Estimating parameters from data
Part II: the force of infection

Niel Hens

www.simid.be
www.socialcontactdata.org

Trento, 2019
Material based on:

- Hens et al. (Springer-Verlag, 2012)
- R-code and data is available at www.simid.be
- Other reading material
  - Vynnycky and White [2010]
  - Bjornstad [2018]
## Rules of Engagement

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

**Table:** Glossary of the most important symbols.
Overview

1. The SIR Model in Endemic Equilibrium
   - The General SIR Model
   - Endemic Equilibrium

2. Serological Data
   - Serological surveys
   - Current Status Data
   - Modelling Current Status Data

3. Estimating the Force of Infection
   - Exploiting the GLM Framework
   - The Early Work
   - The Outbreak in the 90’s
   - The Issue of Monotonicity
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
The SIR Model

- Flow diagram:

```
Birth  \rightarrow  S  \rightarrow  I  \rightarrow  R
\mu  \mu  \mu
\lambda  \nu
```

- Susceptible \rightarrow Infected \rightarrow Immune

- parameters:
  - $\mu$: the death rate
  - $\lambda$: the force of infection
  - $\nu$: the recovery rate
The SIR Model

- Mathematically: partial differential equations
- Each differential equation represents the change (over time and age) in the specific compartment.

\[
\begin{align*}
\frac{\partial S}{\partial a} + \frac{\partial S}{\partial t} &= -\left(\lambda(a, t) + \mu(a)\right)S(a, t), \\
\frac{\partial I}{\partial a} + \frac{\partial I}{\partial t} &= \lambda(a, t)S(a, t) - (\nu + \mu(a))I(a, t), \\
\frac{\partial R}{\partial a} + \frac{\partial R}{\partial t} &= \nu I(a, t) - \mu(a)R(a, t),
\end{align*}
\]

with
- \(N(a, t) = S(a, t) + I(a, t) + R(a, t)\): the population size
- \(S(0, T) = B(t)\): the number of births all susceptible to infection
Endemic Equilibrium

- Focus on the setting of endemic equilibrium
- Endemic equilibrium = time homogeneity:
- Only one time scale: time = age.
- Following one cohort of individuals from birth to death.
- Main assumption: there is no change over time.
The Lexis diagram
The force of infection, \( \lambda(a, t) = \lambda(a) \), is independent of time.

Differential equations:

\[
\begin{align*}
\frac{dS}{da} &= -(\lambda(a) + \mu(a))S(a), & \text{susceptible,} \\
\frac{dI}{da} &= \lambda(a)S(a) - (\nu + \mu(a))I(a), & \text{infected,} \\
\frac{dR}{da} &= \nu I(a) - \mu(a)R(a), & \text{Immune.}
\end{align*}
\]

with

- \( N(a) = S(a) + I(a) + R(a) \): the age-specific population size
- \( S(0) = N(0) = B \) (the number of newborns)
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 ($L = \text{life expectancy}$)

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

- 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.
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.
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:

$$\frac{ds}{da} = -\lambda(a)s(a).$$
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 (susceptible, infected and immune): $\lambda = 0.1$ and $\lambda = 0.2$
- What do you expect in terms of $(s(a), i(a), r(a))$?
The SIR Model in Endemic Equilibrium: an Example

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

- What is the expression for the proportion of susceptible individuals for an MSIR model in endemic equilibrium (M=maternal immunity class)? To simplify calculations you can assume that maternal immunity follows the same principles as ‘Type I’ mortality.

- What happens to the proportion of susceptible individuals for an SIR model in endemic equilibrium in case of disease-related mortality?
Data

- Serological data:
  - focus: measuring antigen-specific antibody levels in the blood.
  - goal: often quantifying population immunity
  - cross-sectional surveys

- Data (all pre-vaccination):
  - Hepatitis A (Bulgaria)
  - Parvovirus B19 (Belgium)
  - Rubella (UK)
  - Varicella Zoster Virus (Belgium)
### Data

**Table**: Summary of the serological data sets.

<table>
<thead>
<tr>
<th>Infection</th>
<th>Main Transmission Route</th>
<th>Time Frame</th>
<th>Country</th>
<th>Age range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hepatitis A</td>
<td>orofecal</td>
<td>1964</td>
<td>Bulgaria</td>
<td>1-86</td>
</tr>
<tr>
<td>Parvovirus B19</td>
<td>airborne</td>
<td>2001-2003</td>
<td>Belgium</td>
<td>0-82</td>
</tr>
<tr>
<td>Rubella</td>
<td>airborne</td>
<td>1986-1987</td>
<td>UK</td>
<td>1-44</td>
</tr>
<tr>
<td>VZV</td>
<td>airborne</td>
<td>2001-2003</td>
<td>Belgium</td>
<td>0-82</td>
</tr>
</tbody>
</table>
Hepatitis A: Bulgaria

- Acute inflammatory disease of the liver.
- Transmission: orofecal
- Infectiousness: two weeks before symptom onset
- Keiding [1991]
Parvovirus B19: Belgium

- B19 infection causes the so-called ‘fifth disease’, a mild rash illness (‘slapped-cheek’ rash)
- Transmission: respiratory droplets
- Infectious (± 6 days) during the incubation period (± 14 days)
- Disease burden: for pregnant women there is a potential for the fetus to have severe anemia, possibly leading to miscarriage.
- Mossong et al. [2008], Hens et al. [2008]
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
- Hens et al. [2008]
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(antibody titers) versus age (left panel); a histogram of the log(antibody titers) ignoring age (right panel).
From Serology to Current Status Data

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

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

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

- There is an alternative approach: mixture models [see e.g. Gay, 1996, Bollaerts et al., 2012]
From Serology to Current Status Data

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

**Figure**: Belgian data on parvovirus B19.
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.
Rubella: UK

UK data on rubella

Figure: Age-specific proportion positive samples of rubella based on cross-sectional survey in the UK.
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).
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 loglikelihood (parameter vector $\theta$):
  \[ L(\theta) = \sum_{i=1}^{N} \left\{ Y_i \log(\pi(a_i)) + (1 - Y_i) \log(1 - \pi(a_i)) \right\}. \]
Estimating the Force of Infection

Binomial likelihood

- Direct Parametrization

\[
\ell(\theta) = \sum_{i=1}^{N} Y_i \log(1 - e^{- \int_{0}^{a_i} \lambda(u) du}) + (1 - Y_i) \log(e^{- \int_{0}^{a_i} \lambda(u) du}).
\]

\[\hat{\lambda}(a)\]

- Indirect Parametrization

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

\[\hat{\lambda}(a) = \hat{\pi}'(a)/(1 - \hat{\pi}(a))\]
Estimating the Force of Infection in R

- Use generalized linear models
  
  \texttt{ glm(...) }

- Define your own (log)likelihood and maximize it:
  
  \texttt{ loglik=function(theta){}
  
  \hspace{1cm} \ldots
  
  \text{ return(totalloglik) }
  
  \}

  \texttt{ mle(loglik,...) }
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:
    \[
    \log\left(\frac{\pi(a)}{1 - \pi(a)}\right).
    \]
  - A complementary log-log link:
    \[
    \log\left(-\log(1 - \pi(a))\right).
    \]
  - A log link:
    \[
    -\log(1 - \pi(a)).
    \]
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 cumulative 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 a binary response, the force of infection is

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

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

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

“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, JASA, 1934)
... 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.

```r
> fit.muench<-glm(NEG/NTOT ~ -1+AGE,
                     family = binomial(link=log),
                     data = Kei1)
```
Software: Muench’s model in R (1)

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

```r
> 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()`.

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

# 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))
```
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 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 $\pi(0) = 0$:

```r
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.

```r
> 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
```
Hepatitis A in Bulgaria: Griffiths’ Model
Griffiths’ Model 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 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 Linear 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
  \[ \lambda(a) = \sum \theta_i a^{i-1}. \]

- 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.
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:

  ```r
  > fit.grenfell <- glm(Kei1$NEG/Kei1$NTOT ~ age1+age2+age3,
  >                       family = binomial(link=log))
  ```
Software: Grenfell and Anderson’s Model

- R output.

```r
> 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)  0.314e-01  4.063e-01   0.816   0.415
age1         0.012e-03  6.692e-02   0.135   0.893
age2        -0.903e-03  2.680e-03  -0.344   0.730
age3         0.364e-05  2.762e-05   0.132   0.895

AIC: 45.755
```
Hepatitis A in Bulgaria: Grenfell and Anderson’s Model
The Piecewise Constant Force of Infection

- Becker [1989]:
  - nonlinear shape:
    \[ \lambda(a) = \mu \alpha a^{\alpha - 1}, \]
  - piecewise constant:
    \[ \lambda(a) = \theta_i \text{ for } a_{i-1} \leq a < a_i \]
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)
Becker’s Model Applied

Becker applied to Rubella in the UK:
Assumptions

Estimating the force of infection from serological data holds under the assumptions of

- endemic equilibrium/time homogeneity and
- lifelong immunity and
- antibody titres are a good marker for immunity
The Outbreak in the 90's

Estimating the Force of Infection


Statistics in Medicine

Semi-parametric estimation of age-time specific infection incidence from serial prevalence data

Nico Nagelkerke, Siem Heisterkamp, Martien Borgdorff, Jaap Broekmans, and Hans van Houwelingen

1National Institute of Public Health and the Environment, Bilthoven, The Netherlands
2Department of Public Health, Erasmus University, Rotterdam, The Netherlands
3Department of Medical Microbiology, University of Manitoba, Winnipeg, Canada
4Royal Netherlands Tuberculosis Association, The Hague, The Netherlands
5Department of Medical Statistics, University of Leiden, Leiden, The Netherlands


Age-specific Incidence and Prevalence: a Statistical Perspective

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), while prevalence is the frequency in the population (of diseased people). From a statistical point of view it is useful to understand incidence and prevalence in the parameter space, incidences 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 in age direction and recruiting individuals from a Poisson process in the time direction. Resulting distributions in the cross-sectional population allow a rigorous discussion of interplay between age-specific incidence and prevalence as well as of the statistical ana...
The Outbreak in the 90’s

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

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

- Otherwise: the estimated force of infection is negative.
The Issue of Monotonicity

- estimating the force of infection
  - relies on the endemic equilibrium assumption
  - untestable in case of one cross-sectional sample [Keiding, 1991]
  - should result in a positive estimate
  - or equivalently a monotonically increasing prevalence
The Issue of Monotonicity

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

  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]:
  \[ \sum_i \theta_i a_i^{i-1} \geq 0, \forall a, \text{ and } \theta_0 = 0 \]
  \[ \text{ex: } \theta_1 a + \theta_2 a^2 \rightarrow \theta_1 + 2\theta_2 a \geq 0 \]
- 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: 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]:

---

doi:10.1017/S0950268809990781

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

N. HENS¹,²*, M. AERTS¹, C. FAES¹, Z. SHKEDY¹, O. LEJEUNE², P. VAN DAMME³
AND P. BEUTELS²
A guide to modelling the force of infection

- What to do with that 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.
References


