Session 8: Applying modelling techniques to analyse (seroprevalence) data

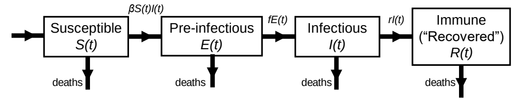
Emilia Vynnycky

Introduction to Infectious Disease Modelling and its Aplications

20th June 2018

Public Health England

General structure of the model set up so far…



Input parameters

- pre-infectious period
- infectious period
- β: the rate at which two specific individuals come into effective contact per unit time…

where $\beta = \dfrac{R_0}{ND}$

N = total population size; D= duration of infectiousness)

BUT…HOW DO WE ESTIMATE R_0???

Answer – through analysing age-stratified data on past exposure to infection to estimate the average force of infection

---

Next steps…

Session 8-9
Analyse infection prevalence data - estimate the average force of infection (λ) by fitting a catalytic model to data and calculate R_0 assuming random mixing

Session 10
Incorporate R_0 in age-structured model of rubella transmission, assuming random mixing and explore effect of vaccination on:
- age-specific proportion susceptible
- infection incidence

Session 11
Use λ to estimate age-dependent contact parameters
Incorporate these parameters in a model to explore effect of age-dependent contact on the impact of vaccination

Session 14
Calculate R_0 using age-dependent contact parameters

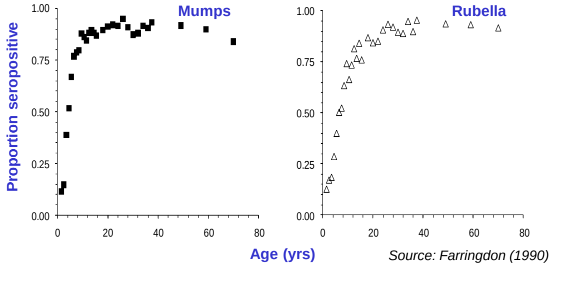
Session 8: Applying modelling techniques to analyse (seroprevalence) data

By the end of this session, you should:

- be able to calculate the average age at infection, proportion susceptible and R_0 using the force of infection, estimated using seroprevalence data
- know how you might use modelling techniques to analyse data on past history of infection
- be able to use graphical and model-free methods to obtain force of infection estimates

---

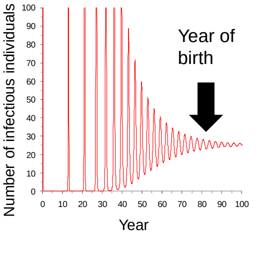
Observed proportion of individuals sero-positive to mumps and rubella antibodies in the UK (late 1980s)



Source: Farringdon (1990)

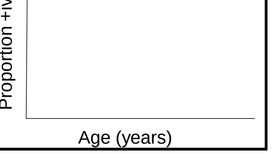
Question: What is the relationship between these age-specific patterns, R_0 and the dynamic predictions from the models studied in block 1???

Predictions of the number of infectious individuals using the measles model from block 1

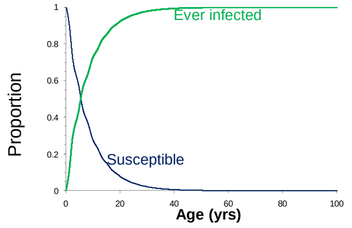


Experiment...

Suppose we track people born in year 80...
How might the proportion who have ever experienced infection (and ∴ seropositive) change as the cohort ages?

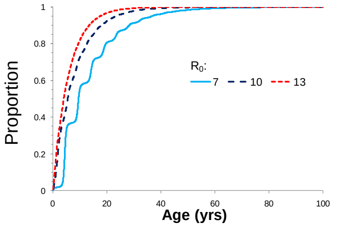


Predictions from the dynamic transmission model from block 1 of the proportion of individuals born in year 80 that should have been infected by different ages



NB if an infection is endemic, these patterns should be similar to those seen in cross-sectional data

R_0 influences the proportion of individuals that have ever been infected (and who should be seropositive) by given ages:



So…we can estimate R_0 by studying age-specific serological data using suitable methods

Methods for analysing seroprevalence data

Seroprevalence data are typically analysed using **catalytic models** to estimate the **average force of infection,** which is then used to calculate:

- the average age at infection,
- the proportion susceptible
- R_0 and herd immunity threshold
- Infection incidence

What do we mean by the average force of infection?

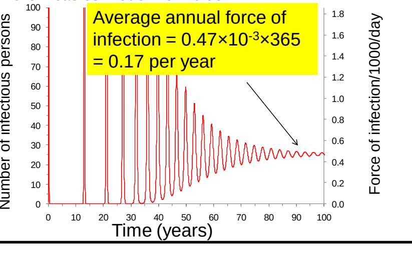
Revision from block 1:

Force of infection = rate at which susceptibles are infected

Note: $\lambda(t) = \beta I(t)$

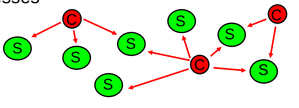
For an endemic infection, the force of infection changes over time, but on **average**, it remains unchanged .

Predictions from the measles model from block 1:



What are catalytic models?

Origin of the term "catalytic models" - Műnch (1959) - analogy from chemical processes



Cases – analogous to a catalyst, exerting a constant force of infection, converting susceptibles into cases, who then become immune

NB analogy is not perfect...

Infection process: Susceptible → (Ever) Infected through contact with infectious persons

Diagrammatically, this can be represented as:

Susceptible *s(a)*	$\xrightarrow{\lambda}$	(Ever) infected *z(a)*

Assumes: a **constant force, λ** on susceptibles per unit time (year)

all individuals are susceptible at birth

Equation for the rate of change in the age-specific proportion susceptible (*s(a)*) and those ever infected (*z(a)*):

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

$$\frac{dz}{da} = \lambda s(a)$$

NB written with respect to age

Infection process: Susceptible → (Ever) Infected through contact with infectious persons

$$\text{Susceptible } s(a) \xrightarrow{\lambda} \text{(Ever) infected } z(a)$$

Recall from lecture on differential equations and/or applying previous knowledge of calculus, the differential equations for this model can be solved to give:

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

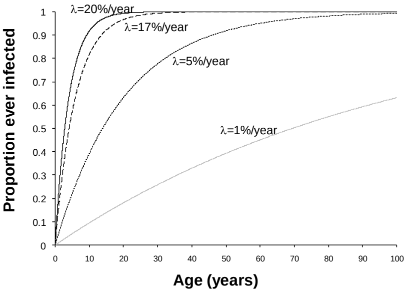
and

Proportion of individuals of age *a* who have been ever infected (*z(a)*)

= 1 – Proportion susceptible at age *a*

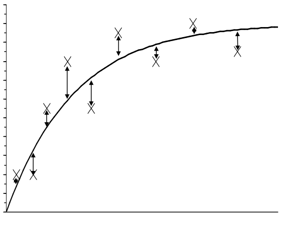
$= 1 - e^{-\lambda a}$

Predictions of the proportion of individuals who have ever been infected (i.e. should be seropositive) for different values for the average annual force of infection



Formal methods for estimating the force of infection

The force of infection is typically estimated formally by fitting model predictions to observed data

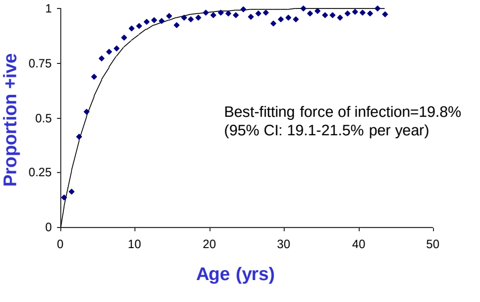


Minimize eg $\sum_i (O_i - E_i)^2$ Least squares

or $\sum_i \frac{(O_i - E_i)^2}{E_i}$ χ^2 statistic

Widely used: "Maximum likelihood" (equivalent to minimizing the loglikelihood deviance – see practical)

Comparison between the mumps data observed in the UK and best-fitting simple catalytic model



What is the difference between a catalytic model and a transmission model?

In transmission models, the force of infection is expressed in terms of the number of infectious individuals in the model, which changes over time, i.e.

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

Catalytic models do not explicitly describe transmission between individuals in the model and the force of infection is taken to be some value which is independent of the size of other compartments in the model

$$\text{Susceptible } s(a) \xrightarrow{\lambda} \text{(Ever) infected } z(a)$$

Methods for analysing seroprevalence data

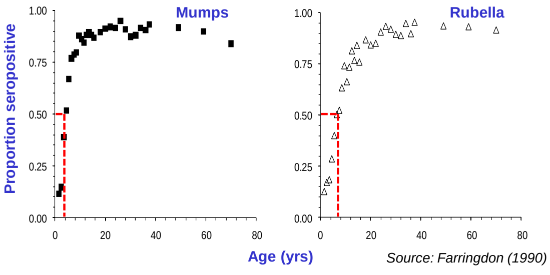
Seroprevalence data are typically analysed using **catalytic models** to estimate the **average force of infection,** which is then used to calculate:

- **the average age at infection,**
- **the proportion susceptible**
- **R_0 and herd immunity threshold**
- **Infection incidence**

Methods for estimating the average age at infection from serological data

- The median age at infection (very quick and crude) – obtained by reading off the age by which 50% of individuals are seropositive

Observed proportion of individuals sero-positive to measles, mumps and rubella antibodies in the UK (late 1980s)



Source: Farringdon (1990)

Methods for estimating the average age at infection from serological data from the average force of infection

- The median age at infection (very quick and crude) – obtained by reading off the age by which 50% of individuals are seropositive
- Using the relationship A≈1/(average force of infection)
  (using the relationship seen in block 1 that the average rate at which something occurs = 1/average time to the event)

NB This expression gives a good approximation to A for different age distributions

Example:
Average annual force of infection for mumps = 0.198/year

=> Average age at infection = 1/0.198 ≈ 5 years

Methods for estimating the average age at infection from serological data from the average force of infection (cont)

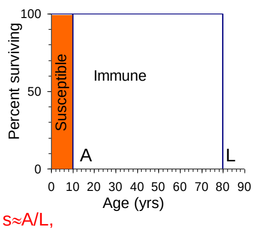
- Using the conventional definition of the average (if the force of infection is age-dependent):

$$A = \frac{\int_0^\infty a\lambda(a)S(a)da}{\int_0^\infty \lambda(a)S(a)da} \quad \text{or} \quad A = \frac{\sum_a a\lambda(a)S(a)}{\sum_a \lambda(a)S(a)}$$

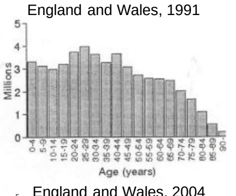
Number of new infections at age a

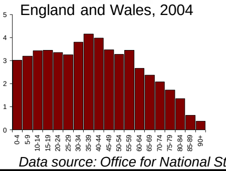
Estimating the proportion susceptible in a population

Rectangular populations:



s≈A/L,
where A= average age at infection,
L = life expectancy

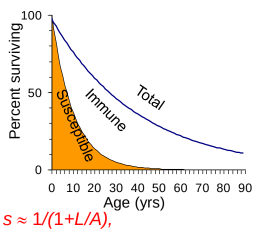




Data source: Office for National Statistics

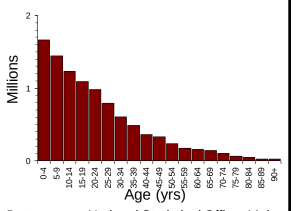
Estimating the proportion susceptible in a population (cont)

Exponential ("type II") populations:



$s \approx 1/(1+L/A)$,
where A= average age at infection,
L = average life expectancy

Malawi: population by age, 1998



Data source: National Statistical Office, Malawi

Estimating R_0

If the force of infection is not age-dependent, and if the infection is endemic:

$$R_0 = 1/\{\text{Proportion susceptible}\}$$

Note: $R_n = R_0 \times$ Proportion susceptible

For an endemic infection, $R_n = 1$

So,

$$R_n = R_0 \times \text{Proportion susceptible} = 1$$

Rearranging this expression gives:

$$R_0 = 1/\ \{\text{Proportion susceptible}\}$$

Rectangular populations: $s \approx A/L \quad \Rightarrow R_0 \approx L/A$

Exponential populations: $s \approx 1/(1+L/A) \Rightarrow R_0 \approx 1+L/A$

Estimates of R_0 for different infections (Anderson and May (1991))

Infection	Location	Time period	R_0
Measles	Cirencester, England	1947-50	13-14
	England and Wales	1950-68	16-18
	Kansas, USA	1918-21	5-6
	Ghana	1960-8	14-15
	Eastern Nigeria	1960-8	16-17
Pertussis	England and Wales	1944-78	16-18
	Ontario, Canada	1912-13	10-11
Chicken pox	Maryland, USA	1913-17	7-8
	Baltimore, USA	1943	10-11
Diphtheria	New York, USA	1918-19	4-5
	Maryland, USA	1908-17	4-5
Mumps	Baltimore, USA	1943	7-8
	England and Wales	1960-80	11-14
Scarlet fever	Maryland, USA	1908-17	7-8
	New York, USA	1918-17	5-6
Rubella	England and Wales	1960-70	6-7
	Poland	1970-7	11-12
	Gambia	1976	15-16
Poliomyelitis	USA	1955	5-6
	Netherlands	1960	6-7

Estimating the (age-specific) incidence of infection

NB This is especially important for infections for which infection at a certain age is associated with complications

e.g. rubella and Congenital Rubella Syndrome

polio and increased risk of paralytic polio for adults

measles and measles encephalitis

etc

Recall from the differential equations sessions that the number of new infections per unit time is given by

$$\lambda(t) \times \text{Number susceptible}$$

Adapting this equations => infection incidence at a given age =

$$\lambda(a) \times \text{Number susceptible at that age}$$

Methods for analysing seroprevalence data

Seroprevalence data are typically analysed using **catalytic models** to estimate the **average force of infection,** which is then used to calculate:

- the average age at infection,
- the proportion susceptible
- R_0 and herd immunity threshold
- Infection incidence

Recap: why do we need the average force of infection?

Estimating the average age at infection:

Applying the formula $A \approx 1/\lambda$ to λ estimated for mumps:
$\Rightarrow A \approx 1/0.198 = 5$ years

Estimating the average proportion susceptible:

Assuming that life expectancy (L) = 70 years, and that the population has a rectangular age distribution
$\Rightarrow s \approx A/L \approx 5/70 = 0.07$

Estimating R_0:
$R_0 = 1/s$ or $L/A \Rightarrow R_0 \approx 1/0.07 \approx 14$

Fine-tuning catalytic models (1)

a) Maternally-derived immunity

$$\begin{array}{c}\text{Maternal}\\\text{protection}\end{array} \rightarrow \begin{array}{c}\text{Susceptible}\\ s(a)\end{array} \xrightarrow{\lambda} \begin{array}{c}\text{(Ever) infected}\\ z(a)\end{array}$$

Assume eg:

i) Maternal immunity is lost at a constant rate μ

(after some calculations…) $s(a) = \dfrac{\mu(e^{-\mu a} - e^{-\lambda a})}{\lambda - \mu}$

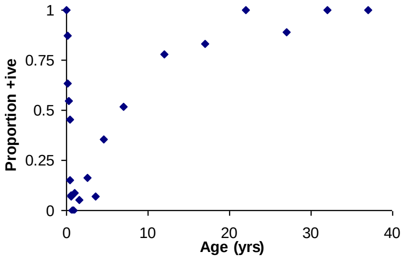
or

ii) Solid immunity to infection during the first 6 months of life:

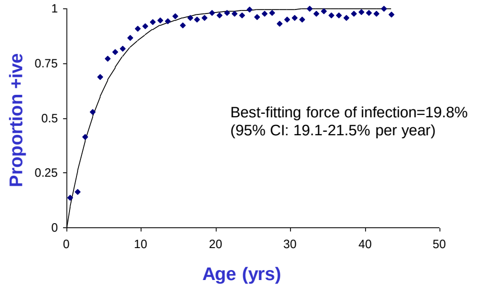
$$s(a) = e^{(-\lambda(a-0.5))}$$

*NB: a-0.5 is the number of years during which individuals of age a **could have been** infected*

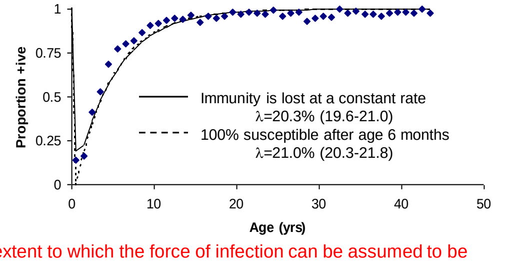
Rubella seroprevalence in Caieiras, Sao Paulo state, Brazil (Azevedo-Neto et al (1994))



Reminder: Comparison between the mumps data observed in the UK and best-fitting simple catalytic model



Comparison between the observed seropositivity to mumps (UK) and the expected prevalence of infection using a simple catalytic model with maternal immunity



The extent to which the force of infection can be assumed to be independent of age can be assessed using graphical methods and logarithms...

Outline of a graphical method for determining whether the force of infection is age-dependent from a dataset

Basic principle: "Natural log i.e. Log to base e

Plot –ln{observed proportion seronegative} for each data point against the age midpoint.

If the resulting plot is a straight line, then:
- the force of infection is the same for all age groups
- the gradient of the line equals the force of infection, λ

Otherwise, the force of infection cannot be assumed to be the same for all age groups...

Why might this work?

Assessing age-dependency in a force of infection (1)

If the force of infection is not age-dependent, then the proportion of individuals of age a who are susceptible is given by:

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

Taking natural logs of both sides of this equation, we obtain the following:

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

We can simplify the right-hand-side of this equation...

Revision of logarithms

Recall from previous mathematical training and maths refresher:

The value of $\log_b x$ is obtained by answering the question:

"To what power must b be raised to get x"

Example

$\log_{10}100 = 2$ since 10 must be raised to the power 2 in order to get 100

$\log_2 8 = 3$ since 2 must be raised to the power 3 in order to get 8

Question: What does $\ln(e^{-\lambda a})$ equal?
$\ln(e^{-\lambda a}) = -\lambda a$ since the power to which "e" must be raised to get $e^{-\lambda a}$ is $-\lambda a$

Assessing age-dependency in a force of infection (1)

If the force of infection is not age-dependent, then the proportion of individuals of age a who are susceptible is given by:

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

Taking natural logs of both sides of this equation, we obtain the following:

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

We can simplify the right-hand-side of this equation:

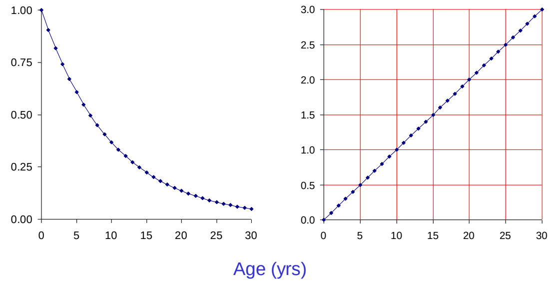
since $\ln(e^{-\lambda a}) = -\lambda a$, we see that

$$-\ln\{s(a)\} = \lambda a$$

This equation is analogous to that of a straight line "y=mx+c" if we replace y by $-\ln\{s(a)\}$, m by λ, x by a and c by 0

Illustration of the relationship between *s(a)* and –ln(*s(a)*)

A. Perfect dataset, $s(a)=e^{-0.1a}$ B. $-\ln(s(a))=0.1a$



=> a graphical check applied to the data may be used to identify whether the force of infection is constant with age

Method for assessing age-dependency in the force of infection (2)

1. Estimate s(a), the proportion susceptible in an age group a as S_a/N_a

   (if S_a =0, replace S_a with 0.5)
2. Calculate the values $-\ln(S_a/N_a)$
3. Plot these values against the midpoints of the age groups

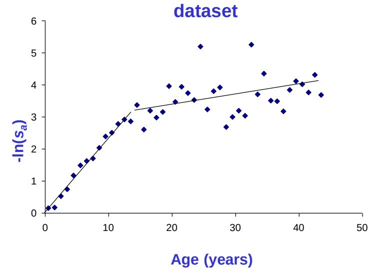
If the force of infection is constant with age,

the plot should approximate to a straight line;

the slope of the line is λ

Otherwise, the force of infection should be assumed to be age-dependent

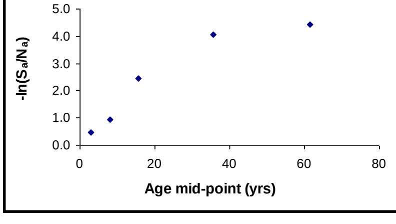
Plot of –ln(Observed proportion seronegative) for the mumps dataset



Conclusion:
λ is constant in the age ranges 0-<15 years and ≥15 years, but differs between these two age groups

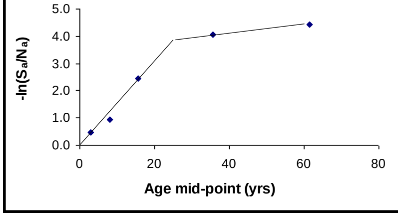
Exercise: Is the force of infection in the population reflected in the following data set age-dependent?

Age (yrs)	N_a	S_a	S_a/N_a	$-\ln(S_a/N_a)$
1-5	52	32	0.615	0.486
6-10	63	25	0.397	0.924
11-20	46	4	0.087	2.442
21-50	58	1	0.017	4.075
51-80	42	0	0.012	4.431



Exercise: Is the force of infection in the population reflected in the following data set age-dependent?

Age (yrs)	N_a	S_a	S_a/N_a	$-\ln(S_a/N_a)$
1-5	52	32	0.615	0.486
6-10	63	25	0.397	0.924
11-20	46	4	0.087	2.442
21-50	58	1	0.017	4.075
51-80	42	0	0.012	4.431



Answer: Maybe – but we need more data among 51-80 year olds…

Estimating an age-specific force of infection

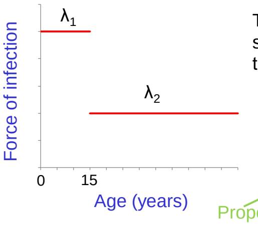
Once we know how the force of infection changes by age, we can estimate it using the following steps:

1. Write down expressions for the age-specific proportion susceptible (assuming an age-dependent force of infection.
2. Fit these expressions to the age-specific serological data.

NB the expressions depend on assumptions about the change in force of infection with age…

Example of an expression for the force of infection (optional)

Suppose we assume that the force of infection differs between the age groups <15 and ≥15 years and equals λ_1 and λ_2



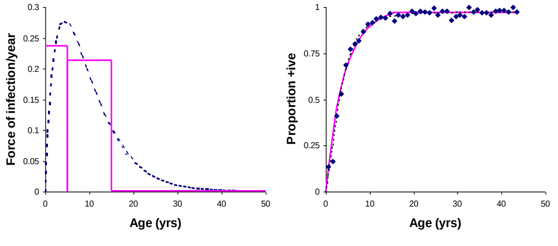
The expression for the proportion susceptible, s(a) differs between those aged <15 and ≥15 years:

$$s(a) = \begin{cases} e^{-\lambda_1 a} & a < 15 \text{ years} \\ e^{-15\lambda_1} e^{-\lambda_2(a-15)} & a \geq 15 \text{ years} \end{cases}$$

Proportion susceptible by age 15 years

Proportion who stay susceptible between age a and 15 years

Fine-tuning catalytic models – the effect of different assumptions about the age-dependency in the force of infection for the mumps data

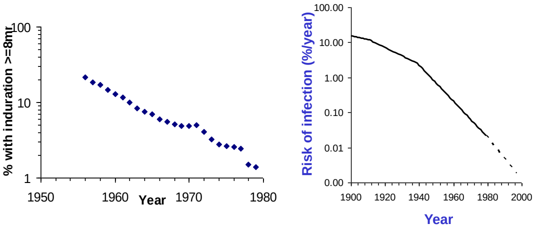


Methods for estimating an age-dependent force of infection are provided in the supplementary questions for this session

Other ways of fine-tuning catalytic models (3)

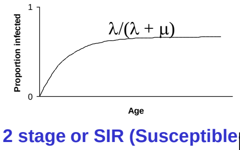
A) Time-dependency in the force of infection

Eg Tuberculin test results in unvaccinated male army recruits – The Netherlands, 1956-79 (Sutherland et al (1983)



Variants of the simple catalytic model

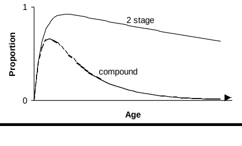
i) Reversible or SIS model: Susceptible $\underset{\mu}{\overset{\lambda}{\rightleftarrows}}$ Infected



Previously applied to cross-sectional tuberculin data, diphtheria, malaria, filariasis

ii) 2 stage or SIR (Susceptible-Infected-Recovered) model: Susceptible → Infected → Recovered

iii) Compound model: Susceptible ⇄ Infected → Recovered



Previously applied to yaws, histoplasmosis (Munch (1959))

Key points from this session

- **Why do we need to study age-specific serological data?** **Answer: to obtain estimates of λ, the average age at infection, proportion susceptible and ultimately R_0**

- If the force of infection is identical for all age groups, the proportion of individuals of age a that is susceptible equals:

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

- We can identify whether λ is age-dependent by plotting $-\ln\{s_a\}$ against the age-midpoint, where s_a=observed proportion seronegative

If the points fall on a straight line, it is reasonable to assume that λ is the same for all ages.

Session 8: Applying modelling techniques to analyse (seroprevalence) data

By the end of this session, you should:

- be able to calculate the average age at infection, proportion susceptible and R_0 using the force of infection, estimated using seroprevalence data
- know how you might use modelling techniques to analyse data on past history of infection
- be able to use graphical and model-free methods to obtain force of infection estimates