

Analysis of seroprevalence data using LSM

Contents

Part I: Fitting the catalytic model to seroprevalence data to estimate the force of infection.....	1
UK.....	1
China.....	3
Maternity.....	4
Age-specific Fol.....	5
Part II: Fitting transmission model to prevalence data to estimate the transmission rate.....	8

Part I: Fitting the catalytic model to seroprevalence data to estimate the force of infection

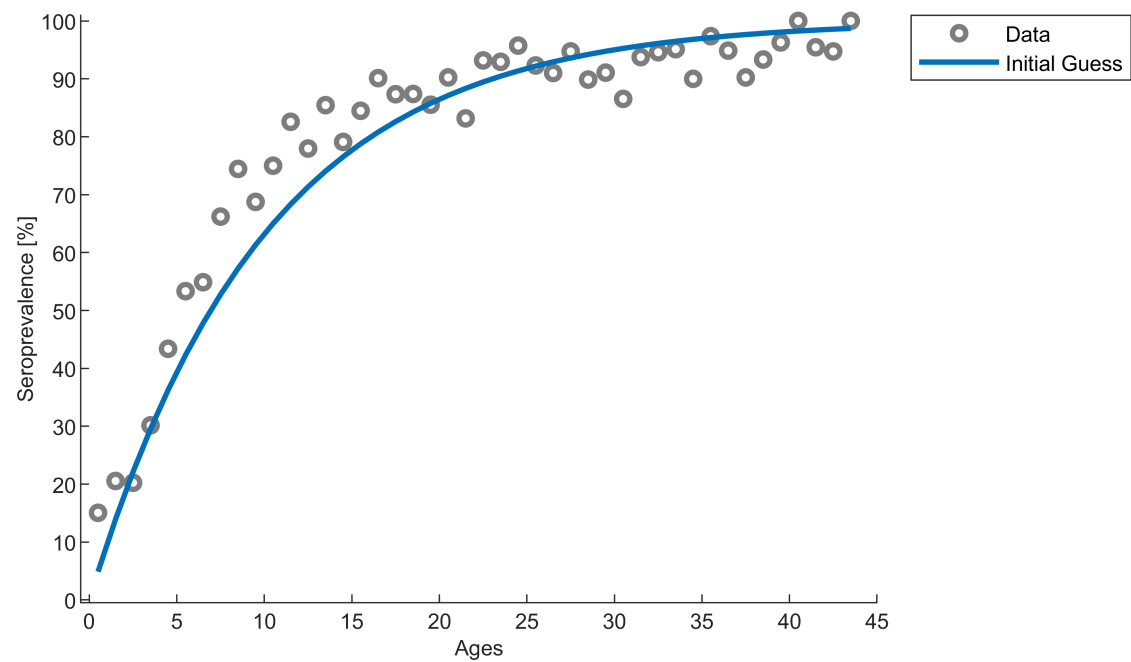
UK

We will first analyze the data from the UK ("seroprevalence_uk"):

Ignoring the maternal antibodies, write the formula for the proportion of age "a" who have ever been infected in terms of force of infection " λ ".

1 Assume the initial value for the force infection in the UK to be 0.1. Do you think the true value for the force infection in the UK was greater or smaller than that currently assumed? What is the current value for the L^2 -error?

The L2 error before estimation is 0.4355

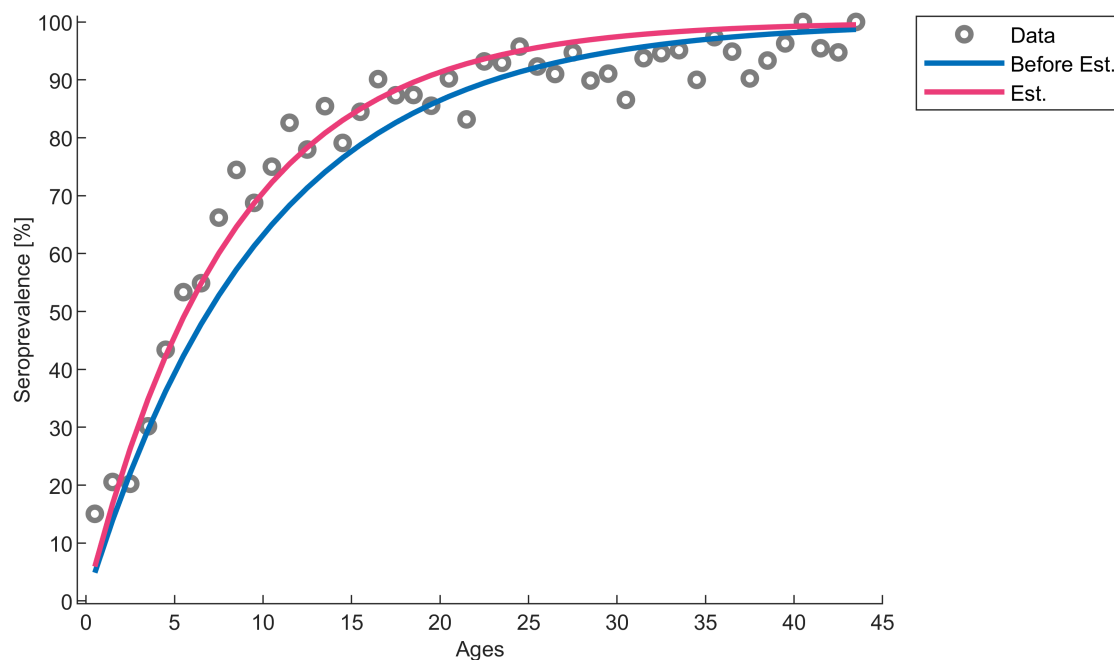


- We can see the true Fol is greater than now.

2 What is the best-fitting value for the force of infection and the current value for the L^2 -error? Plot a graph of model predictions and observed data.

The estimated FoI value is 0.1223

The L2 error after estimation is 0.3295



3 For which age groups does the model underestimate the proportion of individuals who are seropositive? For which age groups does it overestimate it?

- The sero-prevalences of the young who are less than about 15 are underestimated. At the same time, the ones from the others are overestimated.

4 According to the formula, what is the average age at infection in the UK assuming that the force of infection is independent of age?

- Assuming rectangular population structure by age, average age at infection, A , can be written as $A = \frac{1}{\lambda}$.

In our case, A is

$$A = 8.1771$$

5 Assuming that the average life expectancy (L) is 60 years, what is the R_0 for this population according to the expression $R_0 = \frac{L}{A}$? What is the herd immunity threshold?

- R_0 is

$$R_0 = 7.3375$$

- HIT, which can be get as $1 - \frac{1}{R_0}$ from it is

$$\text{HIT} = 86.3715$$

which means the minimal proportion of the protected (by anything) for curbing transmission of disease.

China

Fit the catalytic model to China data ("seroprevalence_china") to estimate the force of infection:

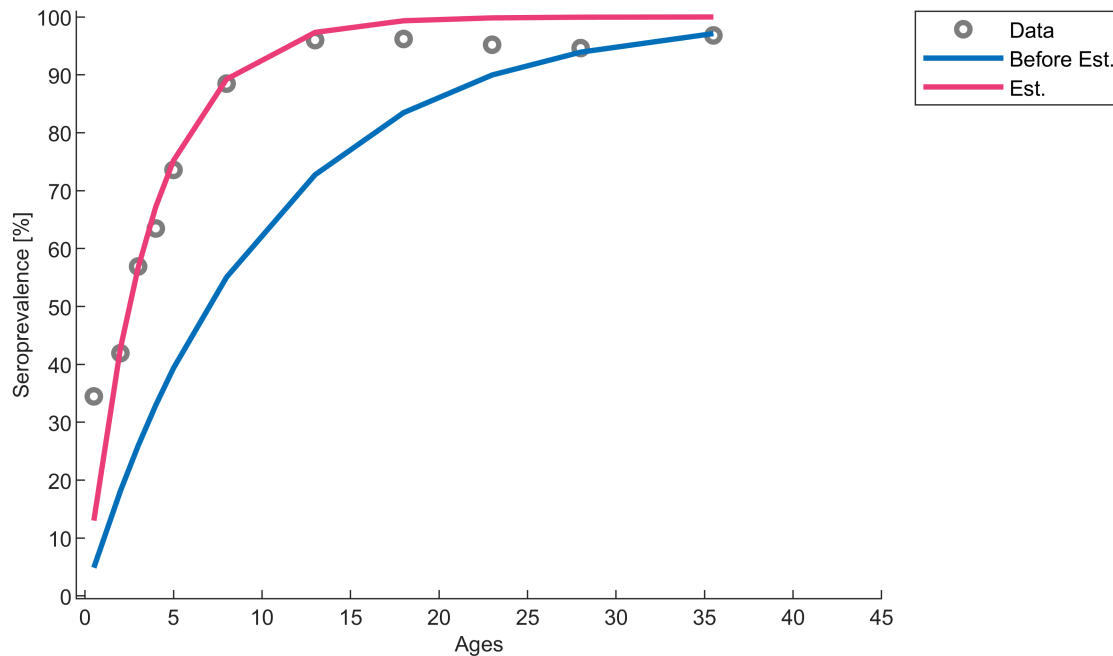
6 Determine the best-fitting force of infection and plot a graph of model predictions and observed data.

Calculate the average age at infection, the R_0 (assuming that the life expectancy is the same as that in the UK) and herd immunity.

The L2 error before estimation is 0.7969

The estimated FoI value is 0.278864

The L2 error after estimation is 0.2345



- Average age at infection, A

$$A = 3.5860$$

- The basic reproduction number, R_0 , with assumption of $L = 60$

$$R_0 = 16.7318$$

- The HIT is

$$\text{HIT} = 94.0234$$

7 How do the values for the force of infection, average age at infection, R_0 and herd immunity threshold in China compare against those for the UK? Suggest possible reasons for these differences.

- The risk of infection of China is larger, which causes early infection, high R_0 and HIT. The main difference is the positive ratio of the young. This is for hygiene problems or if the seroprevalence data of China is collected after vaccination program begins, it can increase the number of sero-positive ratio of the young.

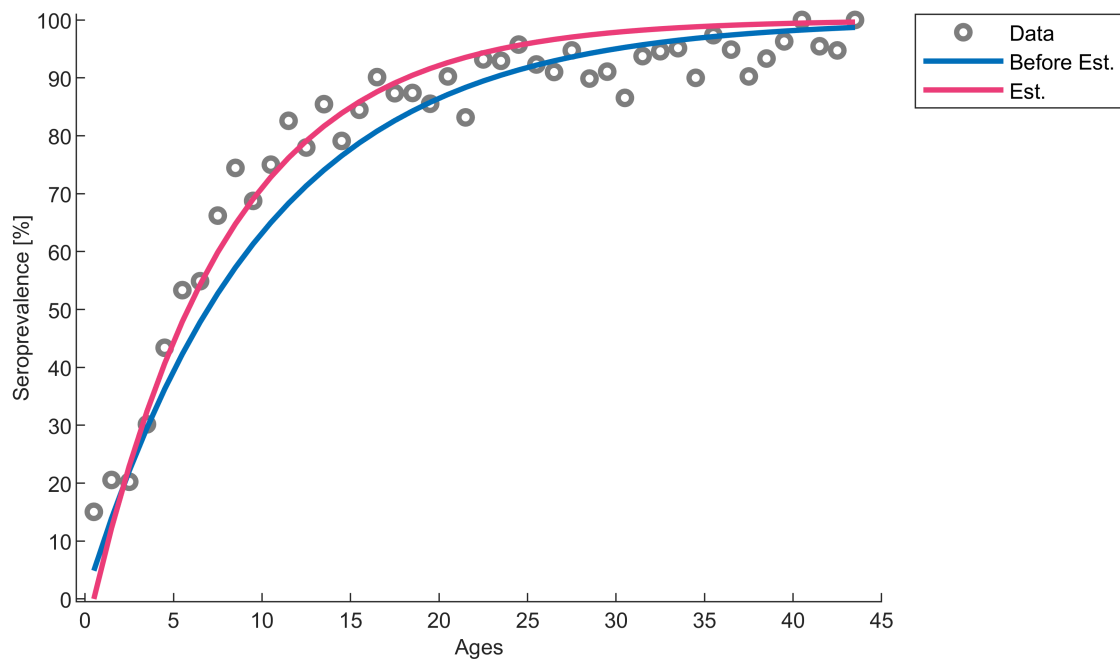
8 Modify the expression for the prevalence of previous infection at each age assuming that individuals are immune for the first 6 months of life as a result of maternal antibodies.

- The original catalytic analysis says proportions of seropositive people aged a is $1 - e^{-\lambda a}$. When we applied the maternal immunity into it, which gives perfect protections to the less-than-6-months-aged children, we should regard the seroprevalence ratio of the aged- a of data as the one of the aged- $(a - 0.5)$. Therefore, the fitting function has the form of $1 - e^{-\lambda(a-0.5)}$.

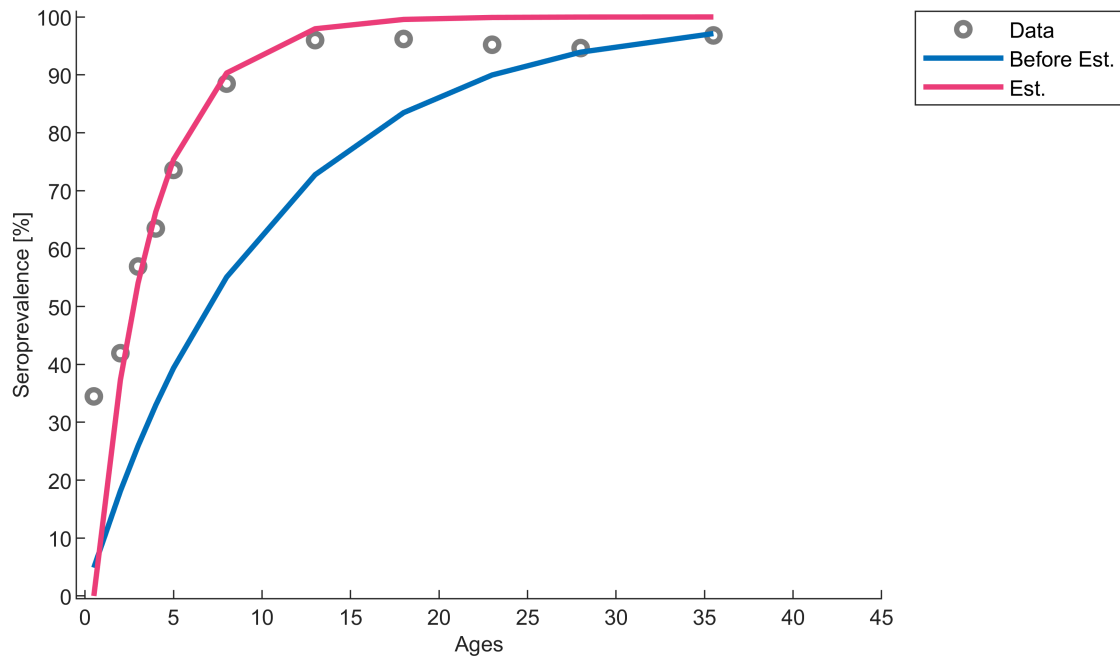
Maternity

9 Assuming that individuals are immune for the first 6 months of life and are then susceptible, refit the model to estimate the force of infection in the UK and China. Plot a graph of model predictions and observed data.

Estimated FoI of UK with considering maternity is 0.1305



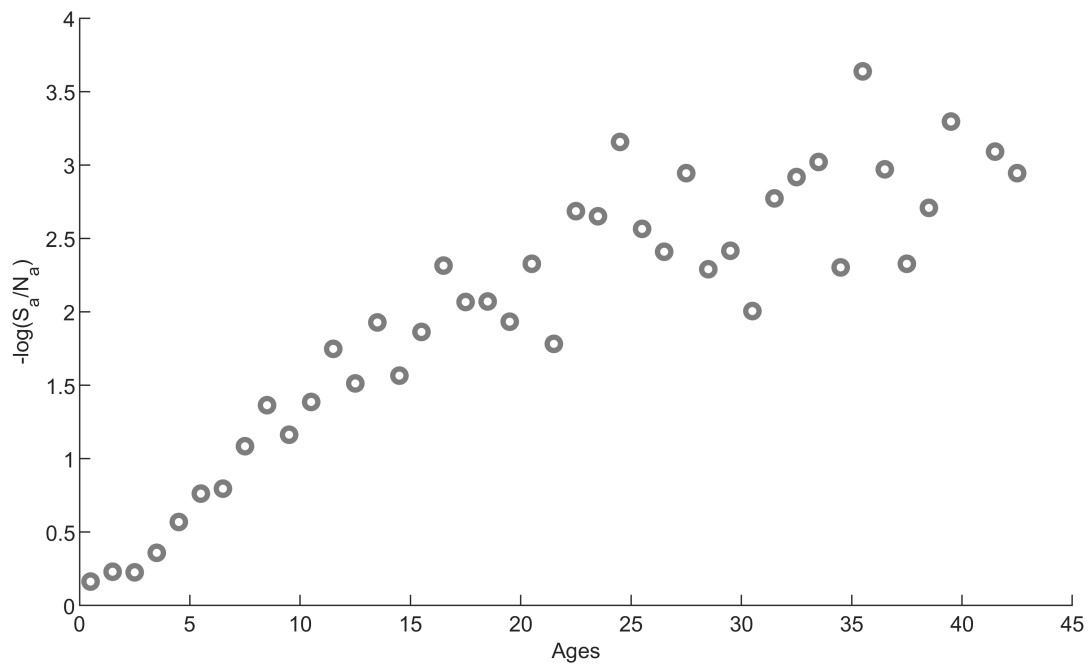
Estimated FoI of China with considering maternity is 0.3113



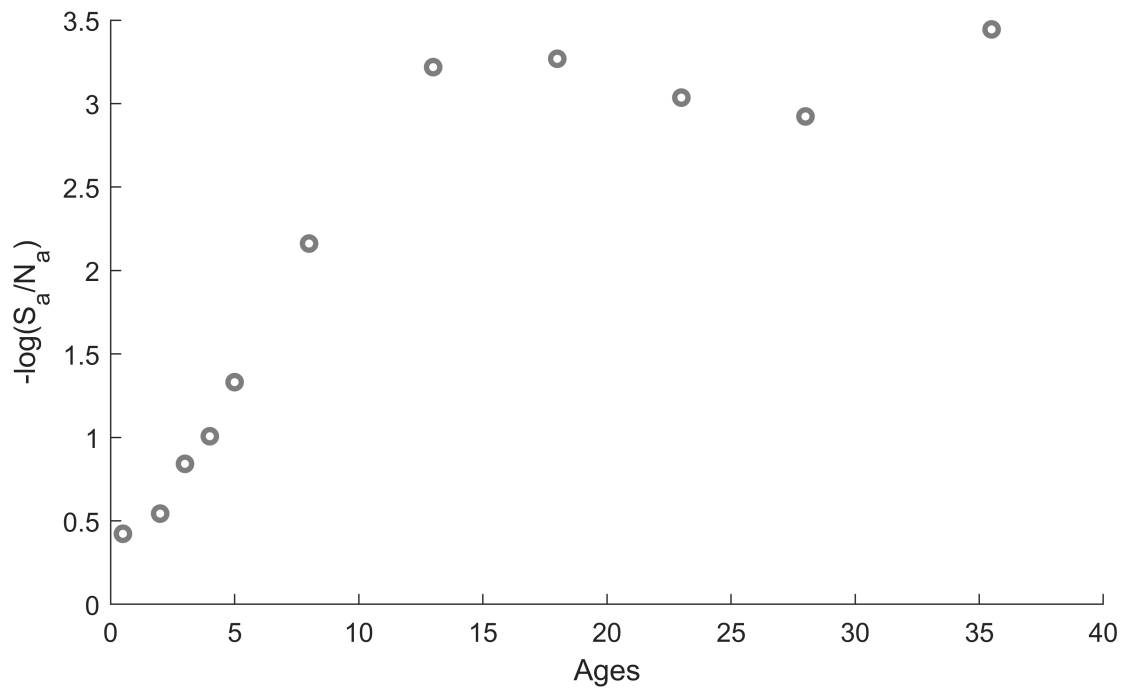
Age-specific FoI

Plot the graphs of $-\ln\left(\frac{S_a}{N_a}\right)$ for the UK and China, where S_a is the number of susceptible at age “ a ” and N_a is the number of population at age “ a ”.

- Draw a plot for UK



- Draw a plot for China



10 According to the plots, is the assumption that the force of infection is independent of age in these populations justified? At what age does it look as though the force of infection changes in these populations?

- A single age-independent FOI can't be justified with two graphs. People have different values of FOI before-and-after around 15 years old.

11 Estimate the age-specific forces of infection using 2 age groups for the UK and China. Suggest reasons for the differences in the force of infection between China and the UK.

- Set the age-specific force of infection and make a function to produce seroprevalence from the catalytic model with age-specific FOI, named `get_seroprev_from_foi_with_age_specificity`.

With the example of 2 age groups differentiated at 15-years-old, the analytic solution of catalytic model is

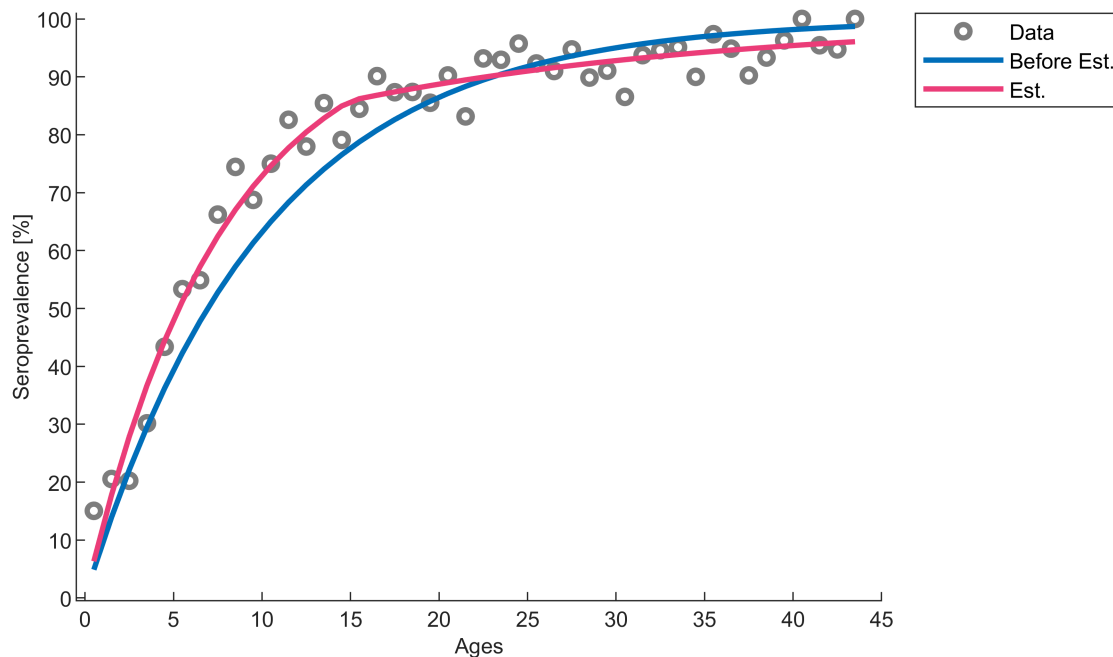
$$s(a) = \begin{cases} e^{-\lambda_1 a} & \text{for } a < 15 \\ e^{-15(\lambda_1 - \lambda_2)} e^{-\lambda_2 a} & \text{for } a \geq 15 \end{cases}$$

Therefore, the sero-prevalence for each age is

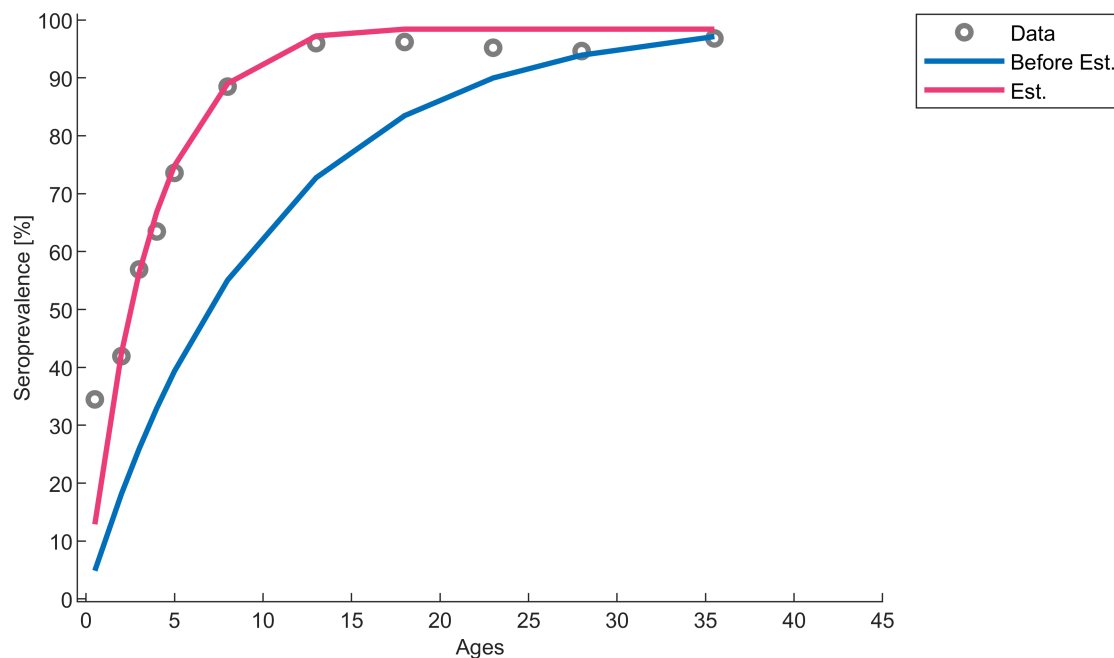
$$z(a) = \begin{cases} 1 - e^{-\lambda_1 a} & \text{for } a < 15 \\ 1 - e^{-15(\lambda_1 - \lambda_2)} e^{-\lambda_2 a} & \text{for } a \geq 15 \end{cases}$$

```
initial_guess = 1x2
0.1000      0.1000
```

Estimated age-specific FoI of UK is 0.1306 (less than 15) and 0.0448 (greater than or equal to 15)



Estimated age-specific FoI of China is 0.2761 (less than 15) and 0.0000 (greater than or equal to 15)



The seropositivity of the young is different. Reasons for this difference make the other differences. The difference of household hygiene or contacts of the young, which is a possible route of transmission, be the reason.

Part II: Fitting transmission model to prevalence data to estimate the transmission rate

Set up the SEIR model of the transmission dynamics of measles in a closed population using differential equations.



We assume that individuals mix randomly and parameter and initial values are given as follows:

1. Population: 100000 people
2. Pre-infectious period: 8 days
3. Infectious period: 7 days
4. Initial values: $(S, E, I, R) = (99999, 0, 1, 0)$

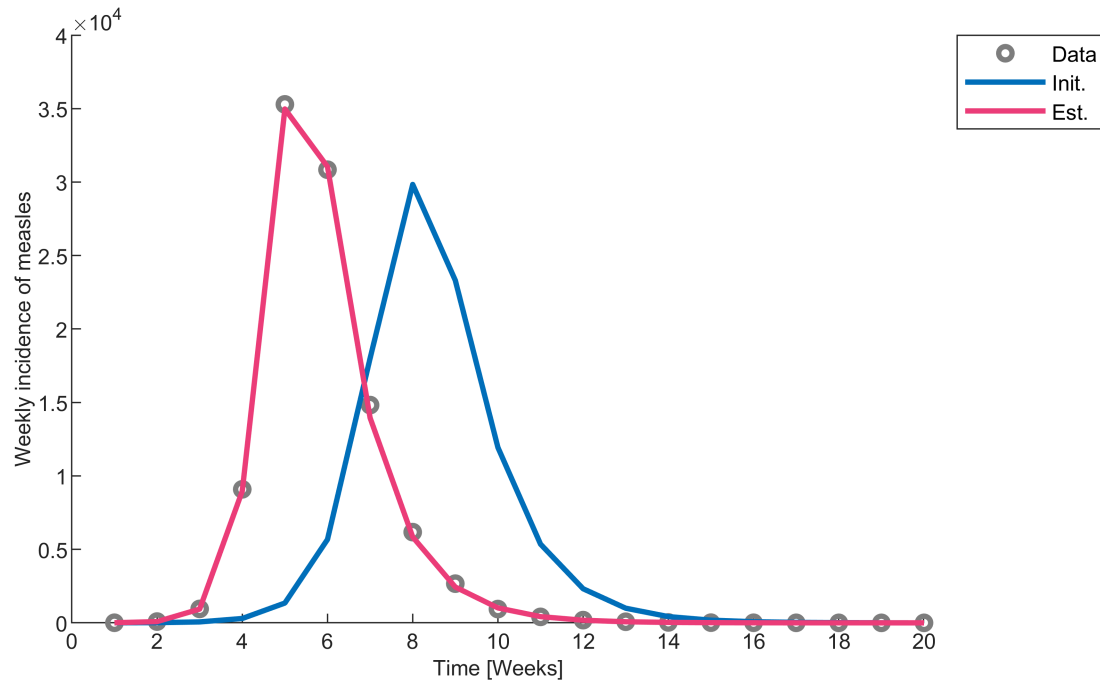
Fit the model to prevalence data to estimate the transmission rate ("incidence_measles"):

12 What is the best-fitting value for the transmission rate? Plot a graph of model predictions and observed data.

- Find the best-fitting transmission rate β with the above parameter sets and initial guess.

The estimated value of beta is 1.858e-05

- Draw a plot of solution with the parameters including estimated_beta value and initial_guess. Plot should include the data points also.



13 Calculate R_0 and herd immunity threshold.

- R_0 is $\frac{\beta N}{\gamma}$ which is

$$R_0 = 13.0084$$

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

$$HIT = 0.9231$$