Estimation of R_0 for endemic diseases

[Practical] (August 5, 2015)

Hiroshi Nishiura
E-mail: nishiurah@m.u-tokyo.ac.jp

Estimation of R_0 from endemic (sero-prevalence) data

Open the spreadsheet "5Aug2015_4.xls". You will see

a) Blue cells: Data on the age-specific seroprevalence of rubella in the UK.

b) Pink cells: in which you will set up equations for the age-specific seroprevalence

c) Yellow cells: containing a value for the constant force of infection. This value has been given the name "foi_uk"

d) A graph plotting the observed proportion of seropositive

(Exercise B-1)

For now, let's ignore contribution of maternal antibodies. What would be the Excel formula for the prevalence of infection among 0.5 year-old in terms of foi_uk?

[Hint] The Excel expression for a function of the form exp(-c) is "=exp(-c)".

Please then copy it down for all the age groups in the data set (in the column labelled "Expected prevalence of infection")

(Exercise B-2)

Currently, foi_uk is assumed as 0.1 per year. Is the true force of infection greater or smaller than that currently assumed?

We will move on to find the best fit value of the force of infection for the UK data. Select rows 5 and 10. Right-click the mouse button and select "unhide" option.

Now you should see green cells containing an expression for the goodness of fit of the model to the data. This is what we call "loglikelihood" which is the negative logarithm of the likelihood function (that we have just learned). To find the maximum likelihood estimate of λ, we will have to minimize this value. If you're interested in seeing how the deviance is calculated, select columns F and I, then click with the right

mouse button and select the unhide option (we ignore the details for now, but please ask questions if you want to know more).

(Exercise B-3)
What is the current value for the deviance?

We will now find the force of infection which yields to the smallest deviance.
Select the "solver" option from the "Tools" menu if it's available. If it's not available, select the "add-ins" option in the Tools menu and click on the "Solver add-in" option, click on OK and then select the "solver" option.

The "Solver" option enables you to minimize a function you specify and find some estimate of the force of infection. It requires you to specify

a) the location of the cell to minimize (this should be cell D6)
b) the cells which you want to change (i.e. the force of infection in cell D4)

In this way, please set up the "Target cells" and the "By changing cells" options.
Select the "Min" option under the "Equal to" option and click on the "Solve" button.

(Exercise B-4)
What is the best-fitting value for the force of infection? Is that consistent with your answer to B-2?

(Exercise B-5)
For which age groups does the model underestimate the proportion seropositive?

(Exercise B-6)
Using the estimated age-independent force of infection, what is the average age at infection?

(Exercise B-7)
Assuming that the average life expectancy at birth, L, is 60 years, and assuming that the demographic pattern follows Type I distribution, what is the estimate of R0 of rubella for this population? Moreover, what is the critical coverage of immunization to eradicate rubella?

Open the spreadsheet "5Aug2015_5.xls". You will see the same information (but, in this time, the data is rubella in China) as it was contained in the UK data.

(Exercise C-1)
Please follow the steps that you carried out using the UK data to calculate
(1) the best-fitting force of infection
(2) the average age at infection
(3) R0
(4) herd immunity threshold (i.e. critical coverage of immunization)

(Exercise C-2)
How do the values for the force of infection, average age at infection, R0 and herd immunity threshold in China compare against those for the UK? Please suggest possible reasons for these differences.

(Exercise C-3)
Although we don't explicitly address maternal antibody, how should your estimate of the force of infection change if you were to assume that individuals are immune for the first 6 months of life as a result of maternal antibodies?

We will now assess whether our assumption of age-independency of the force of infection is justified in these settings. If you have not already done so, unhide column I for "5Aug2015_5.xls". Enter the appropriate formula into the purple cells.

(Exercise D-1)
According to Figure 2, is the assumption that the force of infection is independent of age in this population justified? At what age does it look as though the force of infection changes in this population?

The following contrasts the age-specific forces of infection calculated for the UK against those for China.

Population	Force of infection (per year)	
	< 15 y/o	>= 15 yo
UK	13.3	4.2
China	24.7	"0"

(Exercise D-2)
What do you notice about these values? Please suggest reasons for the differences in the force of infection between China and the UK.

Estimation of R_0 for endemic diseases

[Practical] (August 5, 2015)

(B-1)

The expression should be

= 1 - exp(-A25*foi_uk)

(B-2)

The model underestimates the proportion seropositive for 5-20 year olds and overestimates that for older individuals, suggesting that the true force of infection was higher and possibly lower among younger and older individuals respectively than that assumed.

(B-3)

The deviance should be 176.46.

(B-4)

The best-fitting force of infection should be 0.1159 (i.e. 11.59 % per year) which is slightly higher than that assumed initially. This should be consistent with your answer to B-2.

(B-5)

The model underestimates the observed proportion seropositive for 5-14 year olds and overestimates that for individuals aged > 30 years. This suggests that the force of infection in this population is age-dependent.

(B-6 & 7)

Using A = 1/λ, R0 =L/A and herd immunity = 1-1/R0, you should get

A = 8.6 y/o, R0 = 6.95, critical immunization level = 85.6 %

(C-1)

foi = 20.3 % pear year, A = 12.2 y/o, R0 = 4.9, herd immunity = 91.8 %

(C-2)

FOI, R0 and herd immunity threshold are higher in China than in the UK. This could be attributable to differences in contact patterns between the two populations. For example, individuals in China may contact more individuals each day than do those in the UK possibly as a result of more crowded living conditions (?)

(C-3)

After incorporating maternal immunity, the force of infection should be

estimated higher than that estimated assuming that all individuals are susceptible at birth. As a quick and dirty solution to confirm this point, you can see this by using

$$1 – \exp(-foi\_c * (age-0.5))$$

in the Excel formula (assuming that the duration of maternal antibody is exactly 6 months for all individuals).

(D-1)

The points in the graph do not lie on a straight line, so the assumption that the force of infection is independent of age is not justified. Intuitively, it looks the force of infection changes at about age 15 years.

(D-2)

You should notice that for each population the force of infection is much higher for individuals aged less than 15 years than for those aged over 15 years. Such differences may be attributable to factors such as

1. Differences in exposure to infection
2. Differences in susceptibility

The “0” force of infection estimated for older individuals in China is unrealistic. It follows from the fact that the majority of adults are found to be seropositive: since the percentage of individuals who are seropositive cannot go above 100%, the range of values in which the force of infection can lie is restricted.

Estimation of R_0 for endemic diseases (August 5, 2015)

Hiroshi Nishiura

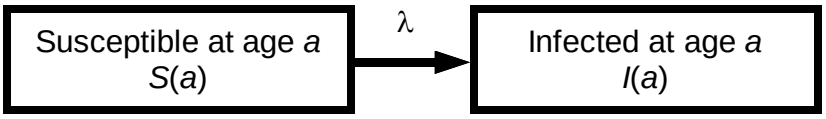
Appendix: Basics of differential equations and estimation of R_0
(Please read Section 1 and work on the exercises before the lecture)

Section 1. How to solve differential equations analytically

Although SIR model with non-linear term (i.e. $\beta S(t)I(t)$) can be solved only numerically, we can analytically solve models with linear terms alone. Since analytical solution of infected individuals is used as the "expected value" of our estimator, here we briefly sketch the basic analytical methods.

[1-1] SI model as a function of age

Consider the following compartments which are intended to model age-specific sero-prevalence data (i.e. proportions of susceptible and infected individuals as a function of age, a, $S(a)$ and $I(a)$):



Here we assume the transition rate, λ (i.e. force of infection), is constant and age-independent. We also assume everyone is born susceptible ($S(0) = 1$ and $I(0) = 0$). We have equations;

$$\begin{aligned} \frac{dS(a)}{da} &= -\lambda S(a) \\ \frac{dI(a)}{da} &= \lambda S(a) \end{aligned} \tag{1}$$

To solve these equations analytically, we first look at $S(a)$. The first subequation can be rewritten as

$$\frac{1}{S(a)}\frac{dS(a)}{da} = -\lambda \tag{2}$$

Now, left-hand side of equation (2) consists of $S(a)$ alone. Moreover, right-hand side consists of constant parameter λ alone. Then, we can integrate both sides from 0 to a.

$$\ln S(a) - \ln S(0) = -\lambda(a - 0) \tag{3}$$

Equation (3) can be rearranged as

$$\ln \frac{S(a)}{S(0)} = -\lambda a \tag{4}$$

Equation (4) is further rearranged as

$$\frac{S(a)}{S(0)} = \exp(-\lambda a) \tag{5}$$

Since $S(a) = 1$, the proportion susceptible at age a is

$$S(a) = \exp(-\lambda a) \tag{6}$$

Let's replace $S(a)$ in the second subequation of (1), i.e.,

$$\frac{dI(a)}{da} = \lambda \exp(-\lambda a) \tag{7}$$

The left-hand side consists of $I(a)$ alone. The right-hand side consists of parameter λ alone. Thus, again we can integrate (7):

$$I(a) = -\exp(-\lambda a) + C \tag{8}$$

where C is constant. Since $I(0) = 0$,

$$I(0) = -1 + C = 0 \tag{9}$$

Thus, $C = 1$. Consequently, we get

$$I(a) = 1 - \exp(-\lambda a) \tag{10}$$

Understanding the above written process is important, because equation (10) is indeed the "expected prevalence" at age a which will be used as an estimator to infer λ from the observed data. When we have to develop a different model with different assumptions, the similar analytical solution has to be obtained using the similar system of equations to write down the expected value.

(Exercise 1-1)

Then, let's think about the age-specific force of infection $\lambda(a)$ at age a (and assume that $S(0) = 1$ and $I(0) = 0$). The similar SI model is governed by

$$\begin{aligned} \frac{dS(a)}{da} &= -\lambda(a)S(a) \\ \frac{dI(a)}{da} &= \lambda(a)S(a) \end{aligned} \tag{11}$$

In this case, equation (3) is replaced by

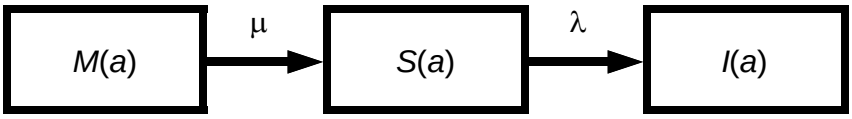
$$\ln \frac{S(a)}{S(0)} = -\int_0^a \lambda(\sigma) d\sigma \tag{12}$$

Using $S(a)+I(a)=1$, what is the expected proportion seropositive ($I(a)$) at age a?

[1-2] Model with maternal antibody

In reality, newborns receive immunity from their mothers (through placenta or through milk). Thus, in addition to the SI model written above, MSI model with $M(a)$

individuals, who are protected by maternal immunity at age *a*, may be more realistic.



Here, μ is the rate at which newborns loose maternal immunity and become susceptible. We assume that the proportion of maternally protected individuals at birth, $M(0) = m$ (where $0 < m < 1$). $S(0) = 1\text{-}m$ and $I(0) = 0$. The equations are

$$\begin{aligned} \frac{dM(a)}{da} &= -\mu M(a) \\ \frac{dS(a)}{da} &= \mu M(a) - \lambda S(a) \\ \frac{dI(a)}{da} &= \lambda S(a) \end{aligned} \tag{13}$$

As we did in the above, the first subequation of (13) is solved as

$$\frac{M(a)}{M(0)} = \exp(-\mu a) \tag{14}$$

Thus,

$$M(a) = m\exp(-\mu a) \tag{15}$$

Now we replace the second subequation of (13) by (15), i.e.,

$$\frac{dS(a)}{da} = \mu m\exp(-\mu a) - \lambda S(a) \tag{16}$$

We rearrange (16) as

$$\frac{dS(a)}{da} + \lambda S(a) = \mu m\exp(-\mu a) \tag{17}$$

so that $S(a)$ is included only in the left-hand side. We multiply $\exp(\lambda a)$ to the both sides of (17), i.e.,

$$\exp(\lambda a)\frac{dS(a)}{da} + \lambda\exp(\lambda a)S(a) = \mu m\exp\left[(\lambda-\mu)a\right] \tag{18}$$

Then, it should be noted that the left-hand side of (18) can be rewritten as

$$\frac{d\exp(\lambda a)S(a)}{da} = \mu m\exp\left[(\lambda-\mu)a\right] \tag{19}$$

Then, it's possible to integrate both sides of (19):

$$\exp(\lambda a)S(a) = \frac{\mu m}{\lambda-\mu}\exp\left[(\lambda-\mu)a\right] + C \tag{20}$$

where C is constant. Since $S(0) = 1\text{-}m$, we get

$$1-m = \frac{\mu m}{\lambda-\mu} + C \tag{21}$$

Thus, C is $1\text{-}m\text{-}\mu m/(\lambda\text{-}\mu)$.

$$\exp(\lambda a)S(a) = \frac{\mu m}{\lambda - \mu}\exp\left[(\lambda - \mu)a\right] + 1 - m - \frac{\mu m}{\lambda - \mu} \tag{22}$$

Multiplying exp(-λa) to the both sides of (22), we get

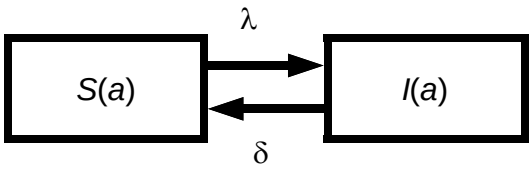
$$S(a) = \frac{\mu m}{\lambda - \mu}\left[\exp(-\mu a) - \exp(-\lambda a)\right] + (1 - m)\exp(-\lambda a) \tag{23}$$

(Exercise 1-2)

Please solve I(a) for the above written MSI model.

[1-3] SIS model with antibody decay

In reality, there are many diseases which do not confer life-long immunity. In the absence of exposure, acquired immunity frequently decays with age. Thus, we might better to consider the following model



where δ is the rate at which infected individuals loose antibody (or loose immunity). We assume $S(0) = 1$ and $I(0) = 0$. The equations are

$$\begin{aligned} \frac{dS(a)}{da} &= -\lambda S(a) + \delta I(a) \\ \frac{dI(a)}{da} &= \lambda S(a) - \delta I(a) \end{aligned} \tag{24}$$

To solve these equations, we use $S(a)+I(a)=1$ for any a. That is, the first subequation of (24) is rewritten as

$$\frac{dS(a)}{da} = -\lambda S(a) + \delta(1 - S(a)) \tag{25}$$

The equation (25) is rearranged as

$$\frac{dS(a)}{da} + (\lambda + \delta)S(a) = \delta \tag{26}$$

As we did in equation (19), we multiply exp[$(\lambda+\delta)a$] to the both sides of (26), i.e.,

$$\frac{d\exp\left[(\lambda + \delta)a\right]S(a)}{da} = \delta\exp\left[(\lambda + \delta)a\right] \tag{27}$$

Integrating (27) with respect to a, we get

$$\exp\left[(\lambda + \delta)a\right]S(a) = \frac{\delta}{\lambda + \delta}\exp\left[(\lambda + \delta)a\right] + C \tag{28}$$

where C is constant. Since $S(0) = 1$, we find that $C = 1-\delta/(\lambda+\delta)$. We get

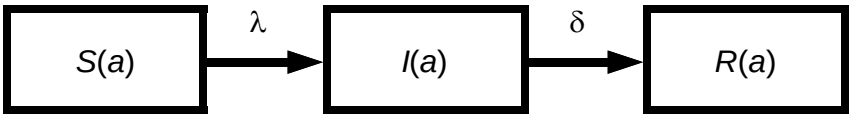
$$S(a) = \exp\left[-(\lambda+\delta)a\right] + \frac{\delta}{\lambda+\delta}\left\{1-\exp\left[-(\lambda+\delta)a\right]\right\} \tag{29}$$

(Exercise 1-3)

Please solve I(a) for the above written SIS model.

(Exercise 1-4)

Although antibody frequently decays, that may not be associated with actual loss of immunity. In that case, it's more appropriate to consider the following SIR model (where $R(a)$ denotes the number of immune but antibody-negative individuals). Here, δ is the rate at which antibody decays. $S(0) = 1$ and $I(0) = R(0) = 0$. Please solve $I(a)$ for the following SIR model.



Section 2. Deriving R_0 using force of infection

In an endemic steady state, we assume that the prevalence is constant over time. Therefore, we assume that the transmission potential can be inferred from age-specific profile of infection alone (i.e. age-specific seroprevalence directly informs R_0). Here we show how to derive R_0 using age-independent force of infection λ.

Given the stationary assumption, the effective reproduction number, R, as a function of time is 1 which is expressed as

$$R = R_0 S^* = 1 \tag{30}$$

where S^* is the proportion susceptible in a population. That is, supposing that $\bar{S}$ and $\bar{N}$ are the total numbers of susceptibles and hosts, S^* is given by

$$S^* = \frac{\bar{S}}{\bar{N}} \tag{31}$$

$\bar{S}$ and $\bar{N}$ are given by

$$\bar{S} = \int_0^\infty S(a)da \tag{32}$$

and

$$\bar{N} = \int_0^\infty N(a)da \tag{33}$$

where $S(a)$ and $N(a)$ denote the age-specific numbers of susceptible individuals and

hosts.

For Type I mortality, the age-specific survivorship of which, $l(a)$, is described by

$$l(a) = \begin{cases} 1 & \text{for } a < \mathrm{L} \\ 0 & \text{for } a > \mathrm{L} \end{cases} \tag{34}$$

where L is the average life expectancy at birth, $\bar{S}$ and $\bar{N}$ are

$$\bar{S} = N(0)\frac{1-\exp(-\lambda L)}{\lambda} \tag{35}$$

and

$$\bar{N} = N(0)L \tag{36}$$

Therefore, R_0 is

$$R_0 = \frac{1}{S^*} = \frac{\bar{N}}{\bar{S}} = \frac{\lambda L}{1-\exp(-\lambda L)} \tag{37}$$

If $\lambda L >> 1$, the equation (37) is approximated by λL.

For Type II mortality with $l(a) = \exp(-\mu a)$ where μ is the death rate, $\bar{S}$ and $\bar{N}$ are

$$\bar{S} = \frac{N(0)}{\lambda + \mu} \tag{38}$$

and

$$\bar{N} = \frac{N(0)}{\mu} \tag{39}$$

Thus, R_0 is

$$R_0 = \frac{1}{S^*} = \frac{\lambda + \mu}{\mu} = 1 + \lambda L \tag{40}$$

(Please note $L = 1/\mu$ for Type II mortality).

References

Anderson RM. May RM. Infectious Diseases of Humans. Dynamics and Control. Oxford University Press, 1991.

Dietz K. Transmission and control of arbovirus diseases. In: Epidemiology, Ludwig D and Cooke KL Eds. Philadelphia, 1975, pp. 104-121.

Dietz K, Schenzle D. Proportionate mixing models for age-dependent infection transmission. J Math Biol 1985; 22: 117-120.

Farrington CP, Kanaan MN, Day NJ. Estimation of the basic reproduction number for infectious diseases from age-stratified serological survey data. Appl Statist 2001; 50: 251-92.

[Solutions for written exercises]

(1-1)

$$I(a) = 1 - \exp\left(-\int_0^a \lambda(\sigma) d\sigma\right)$$

(1-2)

$$I(a) = \frac{m\mu}{\lambda - \mu}(\exp(-\lambda a) - 1) - \frac{m\lambda}{\lambda - \mu}(\exp(-\mu a) - 1) - (1 - m)(\exp(-\lambda a) - 1)$$

(1-3)

$$I(a) = \frac{\lambda}{\lambda + \delta}(1 - \exp(-(\lambda + \delta)a))$$

(1-4)

$$I(a) = \frac{\lambda}{\delta - \lambda}\left[\exp(-\lambda a) - \exp(-\delta a)\right]$$