Monday

Exploring Control Measures

Exploring Control Measures

Goal: Reduce morbidity and mortality due to disease.

Sometimes control measures are focused on protecting vulnerable populations (e.g. elderly people for influenza, or endangered populations of wildlife)

…***but usually the aim is to reduce disease burden in the whole population, by reducing transmission of the infection***.

From earlier lectures, we know that the effective reproductive rate for transmission within a population can be expressed:

$$R_{eff} = c\ p\ D\ (S/N)$$

where

c = contact rate
p = probability of transmission given contact
D = duration of infectiousness
S/N = proportion of the population that is susceptible

Overall disease spread can also be reduced by measures to limiting transmission among populations or among regions.

Measures to reduce the contact rate, *c*

Quarantine: reduce contacts of possible latent cases (*E*)

Case isolation: reduce contacts of known infected individuals (*I*)

ABC: ‘Abstinence’ & ‘Be faithful’

Reducing mass gatherings: school closures etc

Culling (killing of hosts): reducing population density will reduce contact rate (if it’s density dependent)

Measures to reduce the probability of transmission, *p*

Barrier precautions (masks, gloves, gowns etc.)

ABC: ‘Condomize’

Male circumcision
(now known to reduce $f \rightarrow m$ transmission of HIV)

Imperfect vaccines

Prophylactic treatment

Measures to reduce the duration of infectiousness, *D*

Treatment

Case isolation

Contact tracing

Improved diagnostics

Culling of infected hosts

Measures to reduce the proportion susceptible, *S/N*

Vaccination

'Ring' vaccination Contacts of suspect smallpox cases are traced and vaccinated when found. Can be coupled with policy of isolation of identified contacts.	Minimizes use of vaccine, and hence morbidity and mortality caused by adverse reactions to vaccination.
Targeted vaccination For example, vaccination of whole population in affected neighbourhood or city.	Highly effective during eradication campaign at containing transmission localized to a single geographic area or subpopulation. Reduced vaccine-related mortality. Not dependent on contact tracing.
Mass vaccination Vaccination of whole population of a country experiencing or threatened by an outbreak.	Effective at stopping widespread dissemination of the virus across large areas and protecting individuals from infection. Not dependent on contact tracing.
Prophylactic vaccination Vaccination before a smallpox release.	Useful for protecting essential 'first responder' personnel. If used for entire population, very effective at stopping widespread dissemination of virus. Does not have to be implemented quickly. Not dependent on contact tracing.

Measures to reduce transmission between populations

Ring vaccination

Ring culling

Movement restrictions (cordon sanitaire)

Fencing

Measures to reduce vector-borne diseases

Bednets and insect repellents

Vector population reduction
- larvicides
- removal of standing water

Biological control of vectors
- e.g. fungal pathogens of mosquitoes

Treatment of human cases

Vaccination of humans (e.g. yellow fever, malaria, Zika)

Vaccination

Vaccines, typically contain antigens, which are either the whole- or broken-cell protein envelopes from the virus or bacterium causing a specific disease.

When efficacious, the presence of such antigens illicits an immune response in the host, intended to be similar to the consequences of actual infection.

The assumption (and hope) is that the vaccine provides long-lasting immunity to the infection, preventing both transmission and disease.

Two forms of random vaccination are possible:

- pediatric vaccination to reduce the prevalence of an endemic disease;

- random vaccination of the entire population in the face of an epidemic.

The policy of mass vaccination may be applied when necessary, like in case of Ebola or smallpox outbreak.

For many potentially dangerous human infections (such as measles, mumps, rubella, whooping cough, polio, etc.), there has been much focus on vaccinating newborns or very young infants.

Pediatric Vaccination

The mathematical treatment of vaccinating newborns requires making a single addition to the S(E)IR equations.

$$\frac{dS}{dt} = \nu(1-p) - \beta IS - \mu S.$$

$$\frac{dI}{dt} = \beta SI - (\gamma + \mu)I.$$

$$\frac{dR}{dt} = \gamma I + \nu p - \mu R.$$

Conventionally, the parameter p is used to denote the fraction of newborns (or infants who have lost any maternally derived immunity) who are successfully vaccinated and are therefore "born" into the immune class.

This term, p, is the product of the actual vaccination coverage (the percentage of newborns who receive the required number of vaccine doses) and the vaccine efficacy (the probability that they successfully develop immunity).

The above equations can be dynamically explored using a simple (linear) change of variables:

$S = S'(1-p), I = I'(1-p)$, and $R = R'(1-p) + \frac{\nu}{\mu} p$.

These give rise to a new set of ODEs:

$$\frac{(1-p)dS'}{dt} = \nu(1-p) - (\beta I'(1-p) + \mu)S'(1-p),$$

$$\frac{(1-p)dI'}{dt} = \beta S' I'(1-p)^2 - (\gamma + \mu)I'(1-p),$$

$$\frac{(1-p)dR'}{dt} = \gamma I'(1-p) + \nu p - \mu R'(1-p) - \nu p.$$

Canceling out the terms $(1-p)$ on both sides of these equations simplifies to

$$\frac{dS'}{dt} = \nu - (\beta(1-p)I' + \mu)S',$$

$$\frac{dI'}{dt} = \beta(1-p)S'I' - (\gamma + \mu)I',$$

$$\frac{dR'}{dt} = \gamma I' - \mu R'.$$

These equations are identical to the basic SIR equations with a single important modification: The transmission rate β is replaced with $\beta(1-p)$.

Note that if, instead of vaccination, we were attempting to deal with dynamical consequences of a systematic change in the per capita birth rates (from ν to ν', for instance), then we would instead replace (β with $\beta\frac{\nu'}{\mu}$).

These observations simply translate into the following general conclusion:

A system either subject to constant long term vaccination of a fraction p of newsborns against an infection with a basic reproductive ratio R_0, or with a modified per capita birth rate of ν', is dynamically identical to a system with

$$R_0' = (1 - p)\frac{\nu'}{\mu} R_0$$

In order to eradicate a pathogen by long-term pediatric vaccination, we need to ensure that the fraction of susceptible individuals in the population is sufficiently small to prevent the spread of the infection (i.e., $\frac{dI}{dt} \leq 0$).

This is effectively the threshold theorem of Kermack McKendrick and means we need to ensure

$$R_0' = (1 - p) R_0 < 1,$$

which translates into vaccinating a critical proportion of the newborns

$$p_c = 1 - 1/R_0.$$

This vaccination threshold make good intuitive sense, demonstrating that greater action is required for infectious diseases with a larger basic reproductive ratio.

The relationship between p_c and R_0 is plotted in Figure 8.1. It demonstrates that for diseases with very high transmission potential, such as measles and pertussis (R_0 between 16 and 18), the vaccinated fraction of newborns needed for eradication is somewhere between 93% and 95%.

For mumps and chickenpox, on the other hand, the threshold vaccination level is lower, ranging from 87.5% to 90%. For smallpox, p_c, was below 80%.

The figure demonstrates that all incoming susceptibles need not be vaccinated to ensure the infection is not endemic. The shaded regions show the range of p_c, for the estimated R_0 of different infections.

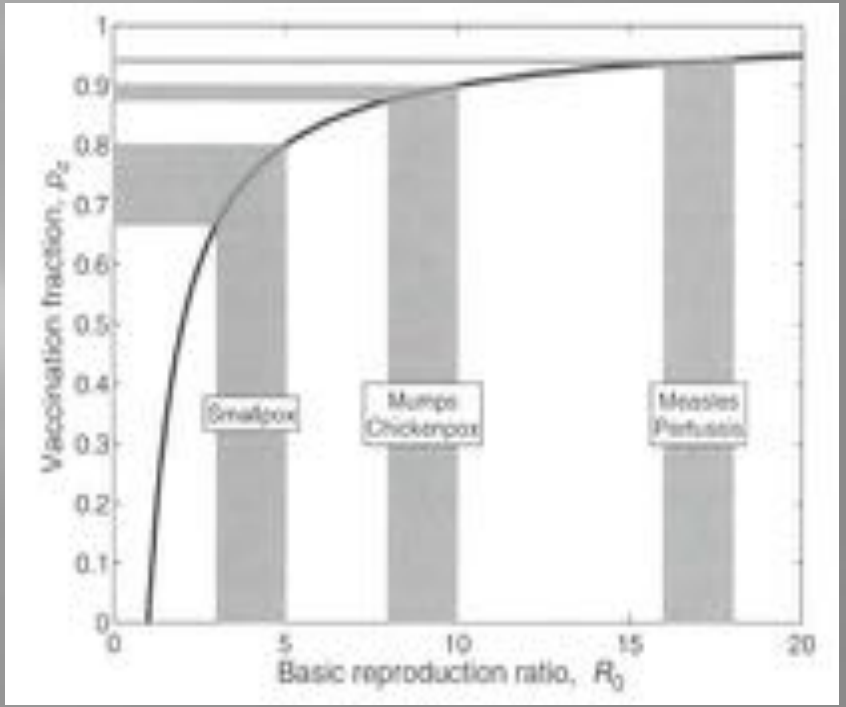


Figure 8.1. The critical fraction of newborns that must be vaccinated to eradicate an infection with a specific basic reproductive ratio (R_0).

In order to eradicate an infection, not all individuals need to be vaccinated, as long as a critical proportion (determine by the reproductive ratio of the infection) have been afforded protection.

This phenomenon is referred to as "herd immunity"

Vaccinating at the critical level p_c does not instantly lead to eradication of the disease. The level of immunity within the population requires time to build up and at the critical level it may take a few generations before the required herd immunity is achieved.

Thus, from a public health perspective, p_c, acts as a lower bound on what should be achieved, with higher levels of vaccination leading to a more rapid elimination of the disease. However, the converse is also true.

Vaccination is still a worthwhile control measure even when the critical level cannot be achieved. In such cases, vaccination reduces prevalence of infection:

$$I^* = \frac{\nu(1-p)}{(\gamma+\mu)} - \frac{\mu}{\beta}. \qquad (8.5)$$

Hence, the equilibrium fraction of infecteds decreases linearly with increasing vaccination, until eradication is achieved.

Thus, even limited vaccination provides protection at population level, as well as direct protection for those individuals vaccinated.

Comparing equation (8.5) with the unvaccinated equilibrium ($I^* = \nu / (\gamma + \mu) - \mu/\beta$), we see that $\nu p/(\gamma + \mu)$ unvaccinated individuals are saved from infection due to the ***herd immunity effects.***

Wildlife Vaccination

There are many instances, especially for wildlife diseases or perhaps when vaccine boosters are necessary, where control by vaccination means targeting the entire susceptible pool and not just the newborns.

This can occur through distributing feed containing vaccine (e.g., to control rabies in foxes), or administering vaccines (e.g., to control distemper in domestic dog populations).

In such cases, we model the random vaccination of any member of the population (irrespective of disease status), although it is only the vaccination of susceptible individuals that has any effect.

The vaccination parameter, v, now necessarily becomes a rate rather than a fraction and we are concerned with the proportion of the susceptible population immunized per unit time. The changes in the mathematical equations describing this scenario are small:

$$\frac{dS}{dt} = \mu - (\beta I + \mu - v)S,$$

$$\frac{dI}{dt} = \beta SI - (\gamma + \mu)I,$$

$$\frac{dR}{dt} = \gamma I + vS - \mu R.$$

The critical rate of vaccination, $v_c = \mu(R_0 - 1)$.

This criterion is clearly different in structure to that derived from pediatric vaccination. The two thresholds, p_c and v_c lead to the same fraction of the population needing to be vaccinated in order to eliminate the infection.

At the critical threshold, a fraction $v_c S^*$ are vaccinated daily; substituting for the values gives $\frac{1}{R_0} \times \mu\,(R_0 - 1)$, which simplifies to $\mu(1 - 1/R_0) = \mu\, p_c$ as previously derived.

Therefore, these two vaccination schemes are equivalent in terms of the numbers of susceptible hosts who need to be immunized.

The key practical difference, however, lies in the fact that wildlife vaccination assumes that the fraction of the population susceptible to infection (given by $1/R_0$) cannot be unambiguously identified and therefore vaccination effort is spread across the entire population—even though it is effective only for susceptible animals.

For this reason, regulating an infectious disease by reducing the recruitment of individuals susceptible to it (eg, pediatric vaccination) may be perhaps easier than attempting to immunize the susceptible population (eg, wildlife vaccination).

Random Mass Vaccination

For rare, non-endemic pathogens, continual vaccination at birth is not a cost-effective control measure. Instead, a mass-vaccination program may be initiated whenever there is increased risk of an epidemic.

In such situations there is a "race" between the exponential increase of the epidemic, and the logistical constraints upon mass-vaccination.

For most human diseases it is possible (and more efficient) to record who has been vaccinated, and only immunize those who have not received the vaccine—an even more refined approach would not vaccinate those individuals who have recovered from the disease because they are already protected.

We take as our most simple model:

$$\frac{dS}{dt} = -\beta SI - u,$$
$$\frac{dI}{dt} = \beta SI - \gamma I,$$
$$\frac{dR}{dt} = gI,$$
$$\frac{dV}{dt} = u,$$

Here demographics have been ignored because we are primarily interested in the short-term response to an emerging epidemic or pandemic. We obviously insist that vaccination stops once the number of susceptibles reaches zero. Two extremes of this model can be considered.

When u is small, vaccination will have little impact on the epidemic and a proportion R_∞ of the population will be infected ($R_\infty = 1 - \exp(-R_0 R_\infty)$).

At the other extreme, when u is large we can use the approximation $S(t) \approx \max(S(0) - ut, 0)$, which assumes that the level of susceptibles is decreased by vaccination although the impact of infection on the level of susceptibles is insignificant and ignored. This is a reasonable assumption if the rate of vaccination is sufficient to control the outbreak.

Under these assumptions, the number of infectious cases is given by:

$$I(t) \approx \begin{cases} I(0)\exp\left(\left[\beta S(0) - \gamma - \frac{1}{2}\beta u t\right] t\right) & t \leq \dfrac{S(0)}{u} \\ I(0)\exp\left(\frac{1}{2}\beta S(0)^2/u - \gamma t\right). & \text{otherwise} \end{cases}$$

Here, the fraction of infecteds follows a Gaussian curve, and the initial disease prevalence at the onset of immunization ($I(0)$) determines the scale of the ensuing epidemic.

This conclusion echoes a broad tenet in epidemiology, that the best way to control an epidemic is to hit it hard and hit it early—a strong response leads to the fastest reduction in the susceptible population, which in turn reduces the epidemic, and a rapid response prevents the exponential increase of cases from getting beyond logistical control.

Imperfect Vaccines and Boosting

Despite the effectiveness of vaccines in dramatically reducing the number of new infectious cases (and the severity of illness), the resurgence and epidemic outbreaks of some infectious diseases are considered to be of major public health.

Among childhood infections, measles is a well-known candidate for such outbreaks. Clinical studies have proposed several potential explanations, including decreased immunization coverage together with irregularities in the supply of vaccines, incomplete protection conferred by imperfect vaccines, and the loss of vaccine-induced immunity.

To prevent an endemic spread of measles infection, many countries, have revised their vaccination programs to include multiple schedules. The reported clinical data using the strategy of a booster MMR (measles-mumps-rubella) vaccine confirm that these countries have generally succeeded in controlling the spread of infection.

Hence, in order to achieve a global eradication, the World Health Organization recommends a booster vaccination program worldwide. The central question to ask is whether this strategy could eventually provide the conditions for global eradication.

To address this question, we can develop a framework, modified from the SEIR equations, that would predict the consequences of the introduction of a booster schedule, in terms of the known major factors associated with a vaccination.

The model present here is composed of four distinct classes: Susceptible (S) , Vaccinated (S_v), Infectious (I), and Booster vaccinated (or recovered) individuals (V) who are immune for life.

It accounts for two major aspects of an imperfect vaccine: (1) incomplete protection, and (2) waning of vaccine-induced immunity.

The first may result in the subsequent infection of the pediatric-vaccinated class, perhaps at a lower rate than that of the fully susceptible class.

The second leads to an increase in the size of the fully susceptible pool through the loss of vaccine-induced immunity.

The model also assumes that, like the natural immunity induced by the infection, the booster vaccine administered to the class of pediatric-vaccinated individuals confers complete protection against the disease.

The system can be mathematically expressed by the following system of differential equations:

$$\frac{dS}{dt} = (1-p)\mu - \beta SI - \mu S - \xi S + \delta S_v,$$

$$\frac{dS_v}{dt} = p\mu + \xi S - (1-\alpha)\beta S_v I - (\mu + \rho + \delta)S_v,$$

$$\frac{dI}{dt} = \beta SI + (1-\alpha)\beta S_v I - (\mu + \gamma)I,$$

$$\frac{dV}{dt} = \rho S_v + \gamma I - \mu V,$$

p is the fraction of newborns who receive the pediatric vaccine, α represents the efficacy of the vaccine in terms of reducing the susceptibility of (singly) vaccinated individuals,

δ is the waning rate following pediatric vaccination, $1/\gamma$ is the infectious period, μ is the natural death rate, and ρ and ξ are the rates of administration of the boaster vaccine to previously vaccinated and susceptible individuals, respectively.

The effective reproductive ratio, r_0, given by the following expression:

$$r_0 = \frac{\mu[\delta + (1-p)(\mu+\rho) + (\mu p + \xi)(1-\alpha)]\beta}{(\mu+\gamma)[(\mu+\xi)(\mu+\rho) + \mu\delta]}.$$

Naturally, there is significant public health interest to ensure control parameters that would make eradication feasible by reducing r_0 below unity.

An increase in $p, \delta,$ or β —which relate to more susceptibles entering the population, a decrease in the mean duration of vaccine-induced immunity, and a higher transmission rate respectively —can all be offset by a higher level of pediatric vaccination.

It is useful to rewrite r_0 in terms of the basic reproductive ratio for a population that is wholly susceptible, with no vaccination (R_0)

This gives:

$$r_0 = \left(1 - \frac{(\mu p + \xi)(\rho + \mu\alpha)}{(\mu + \rho)(\mu + \xi) + \mu\delta}\right) R_0,$$

where, as before, $R_0 = \beta/(\gamma + \mu)$.

Clearly, a high value of R_0 requires a high coverage level of pediatric vaccination, p, to prevent the spread of the infectious disease, regardless of the type of vaccine being administered.

However, it is practically unfeasible to vaccinate all individuals in the susceptible class (p is always significantly less than 1), particularly in countries where finances play a major role in the number of people who receive the vaccines.

Hence, the next best strategy is to determine the critical number needed to be vaccinated and try to achieve this value.

It is instructive to establish the minimum pediatric vaccination level that is required to eliminate the infectious disease in the absence of boosters ($\rho = \xi = 0$)—the equivalent to the standard vaccination model but with waning immunity and partial protection. This is given by:

$$p_c = \left(1 - \frac{1}{R_0}\right)\left(\frac{\mu + \delta}{\mu\alpha}\right), \quad (8.14)$$

such that $r_0 \leq 1$ whenever $p \geq p_c$.

Not surprisingly, this threshold reduces to $p_c = 1 - 1/R_0$ for a perfect vaccine ($\alpha = 1,\ \delta = 0$).

The most important implication of this result is that eradication may be impossible to achieve once the reproduction number, R_0, is greater than 2.

Consider the optimistic case in which the pediatric vaccine provides perfect immunity to infection ($\alpha = 1$), but where protection wanes through time ($\delta > 0$).

In this scenario, equation (8.14) means that the critical proportion of the population required to be vaccinated becomes greater than $1(p_c \geq 1)$, unless the ratio of life expectancy to the period of protection $((\mu + \delta)/\mu)$ is less than $R_0/(R_0 - 1)$.

As a result, for a pathogen with effective $R_0 = 3$, this result effectively means that eradication requires the period of protection to last for at least $2/3$ the duration of life—hence the need for booster vaccination.

Simple Isolation

Isolation or quarantining of individuals provides the most effective, means of disease control.

It is focused toward those individuals who are infected—preventing them from further contact and subsequent transmission.

Quarantine always involves isolating infected individuals as soon as they are diagnosed and can be applied in many ways.

Unlike vaccination which acts on susceptible individuals preventing from becoming infected, quarantine acts by removing infectious individuals population, dramatically reducing their risk of transmission.

This can be simply modeled by an effective decrease in the infectious period (or an increase in the recovery rate), adding a quarantine class Q to the standard SIR approach:

$$\begin{aligned}
\frac{dS}{dt} &= \mu - \beta SI - \mu S, \\
\frac{dI}{dt} &= \beta SI - \gamma I - d_I I - \mu I, \\
\frac{dR}{dt} &= \gamma I + \tau Q - \mu R, \\
\frac{dQ}{dt} &= d_I I - \tau Q.
\end{aligned} \tag{8.24}$$

where d_I is the rate at which infectious individuals are detected and "removed" to quarantine in addition to the normal recovery rate, and $1 / \tau$ is the average isolation.

We assume that individuals leave the quarantine class, Q, only recovered.

This leads to a reproductive ratio of

$$R_Q = \frac{\beta}{\gamma + d_I + \mu},$$

The critical isolation threshold that ensures $R_Q = 1$ is $d_I = \beta - \gamma - \mu$

Lets consider the effects of logistical constraints on the isolation facility. In practice, an isolation facility has a maximum capacity, Q_c, that it can accommodate.

For example, consider an isolation ward in a hospital in the context of an outbreak of Methicillin Resistant Staphylococcus Aureus (MRSA).

Similar considerations could apply locally or nationally for outbreaks of smallpox, SARS, or pandemic influenza.

When the isolation facility is operating below capacity ($Q < Q_c$ or $d_I I < \tau Q_c$), then the dynamics are governed by simple isolation model give above.

However, if infection exceeds capacity ($Q = Q_c$ and $d_I I > \tau Q_c$ isolation facility is full and there is an excess of infection), newly detected infections can enter isolation only when someone leaves, leading to:

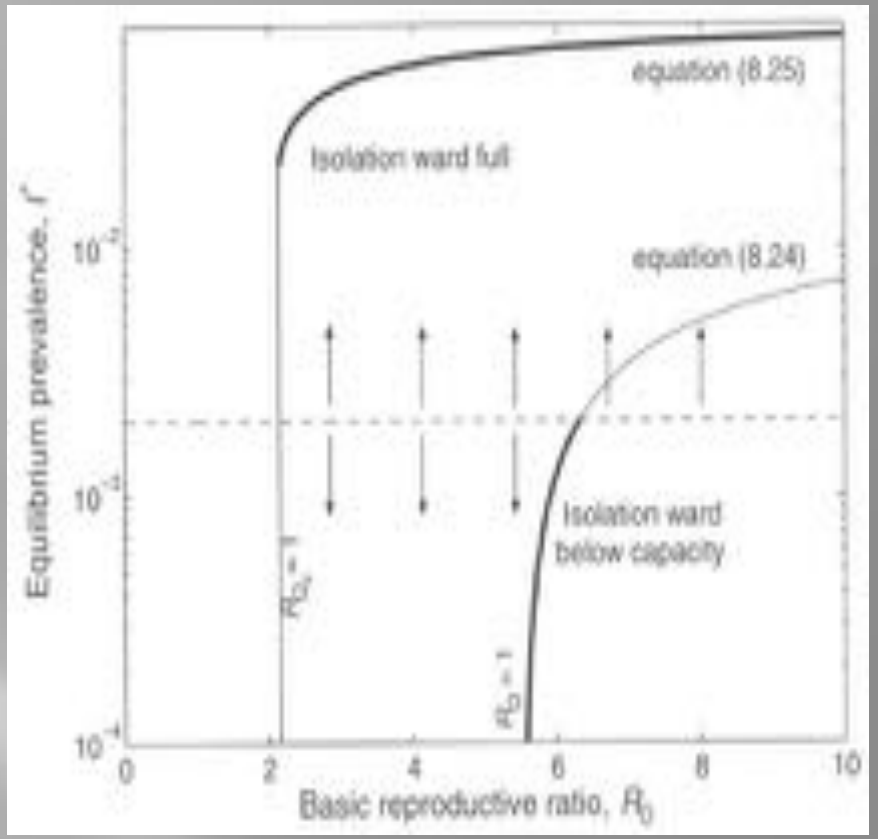
$$\begin{aligned}
\frac{dS}{dt} &= \mu - \beta SI - \mu S, \\
\frac{dI}{dt} &= \beta SI - \gamma I - \tau Q_c - \mu I, \\
\frac{dR}{dt} &= \gamma I + \tau Q_c - \mu R, \\
\frac{dQ}{dt} &= \tau Q_c - \tau Q_c,
\end{aligned} \tag{8.25}$$

in which case the reproductive ratio depends on the current number of cases and is

$$R_{Q_c} = \frac{\beta}{\gamma + \tau Q_c / I + \mu} > R_Q.$$

Given that there are now two plausible equations, (8.24) and (8.25), two equilibria possible (Figure 8.7): one when the isolation ward is below capacity

$$I_Q^* = \mu(R_Q - 1)/\beta,$$



and one when there is an excess of infection

$$\beta(\gamma+\mu){I_{Q_c}^*}^2 + (\beta\tau Q_c + \mu(\gamma+\mu-\beta))I_{Q_c}^* - \mu\tau Q_c.$$

We observe that for some intermediate values of $R_0 (= \beta/(\gamma+\mu))$ the two solutions are both stable.

This has a number of far-reaching implications.

First, once the quarantine limit is reached the prevalence of infection dramatically increases, jumping from the solution to the higher stable solution.

Second, once the quarantine capacity is reached it may be exceedingly difficult to regain spare capacity because the reproductive R_{Q_c} , may be much higher.

Finally, the size of the isolation facility needed is governed by the the prevalence predicted by equation (8.25), and not the normal lower level equation (8.24).

Thus isolation facilities may need to far exceed the usual demand if they are not to succumb to catastrophic failure.

Where isolation or quarantining capacity is limited, ***bistability*** can occur.

Home Work

List all possible control measures for the following diseases

1. Human Papillomavirus (HPV)

2. Zika

3. Typhoid Fever

4. Bovine Tuberculosis

Home Work

Currently there is no treatment nor vaccine for Ebola, incorporate into the Ebola model treatment and vaccination with imperfect vaccine.

Compute the reproduction number and write a matlab code to implement this system.

Parameter estimation and model fitting

After developing a model, one of the most important steps is to validate the model by comparing it with data.

Model validation is the process of determining the degree to which a mathematical model accurately represent the real-world data.

It is important therefore to link our model to data to help give confidence in the model and to obtain realistic estimates of the parameters.

Furthermore, it is important to determine which models are good or are bad, but this relies on statistics.

We will introduce some basic techniques here to address such question.

Data needs I. What's needed to build a model?

Individual "clinical" data

- Latent period: time from infection to transmissibility
- Infectious period: duration (and intensity) of shedding infectious stages
- Immunity: how effective, and for how long?

Population data

- Population size and structure
- Birth and death rates, survival, immigration and emigration
- Rates of contact within and between population groups

Epidemiological data

- Transmissibility (R_0)
  - density dependence, seasonality

Data needs II. What's needed to validate a model?

Time series

- Incidence: number of new cases
- Prevalence: proportion of population with disease

Seroprevalence / sero-incidence: shows individuals' history of exposure.

Age/sex/spatial structure, if present.

e.g. mean age of infection → can estimate R_0

Cross-sectional data

Seroprevalence survey (or prevalence of chronic disease)

endemic disease at steady state → insight into mixing

epidemic disease → outbreak size, attack rate, and risk groups

Where does the data come from?

First data could come from biologists or epidemiologists who collect data.

Comprehensive long-term datasets are usually collected by various health organization such as

- World Health Organization (WHO),
- Centers for Disease Control and Prevention (CDC), and
- various foundations.

These datasets can be obtained by requesting them from the health organization or online.

For instance, if you go to the WHO Data and Statistics website *http://www.who.int/research/en/*

There are a number of important diseases listed with data and statistics about them.

Where does the data come from?

Suppose we are interested in the cholera epidemic that occurred in Haiti after the devastating earthquake of January 2010.

The central WHO website gives only the number of yearly cases by country.

If we need more resolution, if, for instance, we need monthly or weekly cases we can google "cholera data monthly".

We may find the data on the Pan American Health Organization website http://new.paho.org/hq/images/Atlas_IHR/CholeraHispaniola/atlas.html.

Where does the data come from?

The third possible approach is to obtain the data from published articles.

Data in articles are often published as plots. Hence, if we want the actual coordinates, we need to extract them from the plots.

There are many routines that can be used to extract values for the points in a plot.

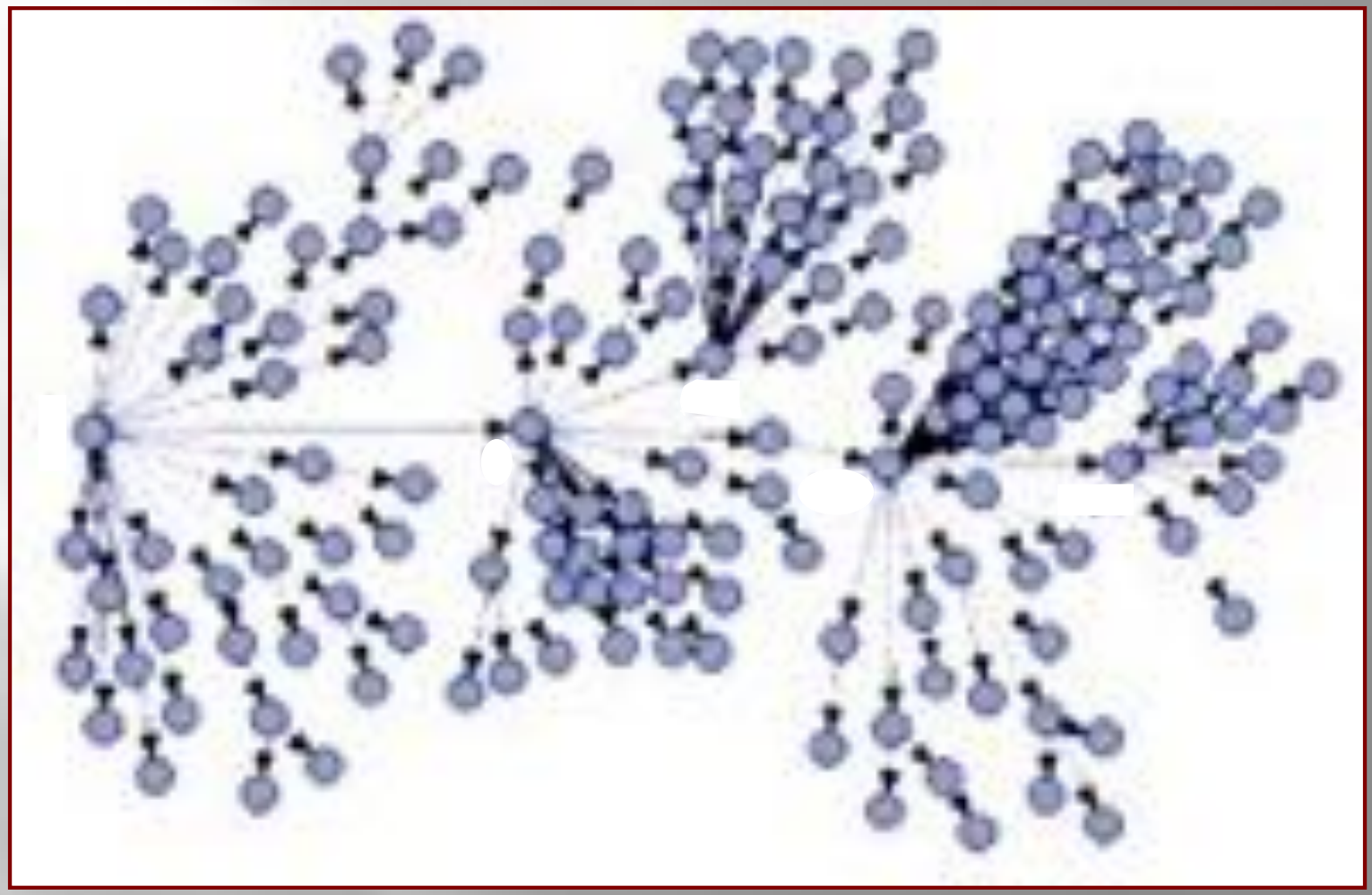
One is PlotDigitizer at http://plotdigitizer.sourceforge.net.

Another is Matlab, which has capabilities to extract data values from a plot. The matlab app (grabit) can be obtained by downloading grabit.zip.

The instructions on how to use it can be found at http://extractdata.blogsport.com/

Contact tracing

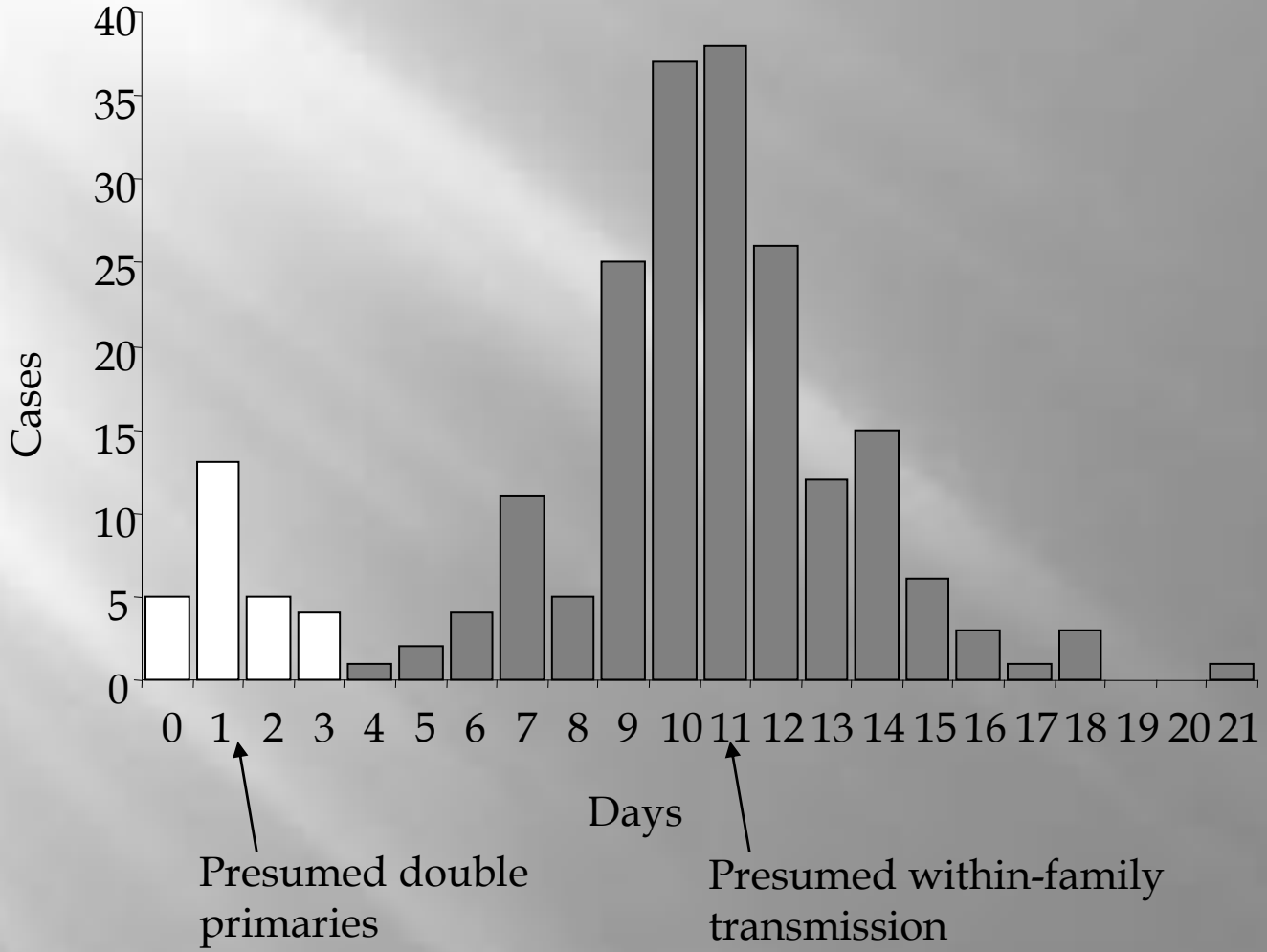
SARS transmission chain, Singapore 2003



Morbidity & Mortality Weekly Report (2003)

Household studies

Observed time intervals between two cases of measles in families of two children. Data from Cirencester, England, 1946-1952 (Hope-Simpson 1952)



Measles:

Latent period 6-9 d, Infectious period 6-7 d, Average serial interval: 10.9 d

Long-term time series

. The Diſeaſes and Caſualties this Week.

Abortive	4
Aged	45
Bleeding	1
Broken legge	1
Childbed	25
Chrisomes	9
Consumption	126
Convulsion	89
Cough	1
Dropsie	33
Feaver	348
Flox and Small-pox	11
Flux	1
Frighted	2
Gout	1
Grief	3
Griping in the Guts	72
Head-mould-shot	1
Jaundice	7
Impoſthume	1
Infants	12
Kingsevil	1
Lethargy	1
Lyce	1
Plague	43
Plurisie	1
Rising of the Lights	18
Rupture	1
Scurvy	1
Spotted Feaver	168
Stopping of the Stomach	19
Strangury	1
Suddenly	1
Teeth	111
Thrush	6
Tiſick	9
Vomiting	10
Wind	1
Worms	10

Christned: Males 96, Females 81, In all 177

Buried: Males 3777, Females 2361, In all 6138

Plague 4237

Increaſed in the Burials this Week 343

Pariſhes clear of the Plague 17 Parishes Infected 111

The Aſſize of Bread ſet forth by order of the Lord Maior and Court of Aldermen
A penny Wheaten Loaf to contain Nine Ounces and a half, and three
half penny White Loaves the like weight.

Historical mortality records provide data:
London Bills of mortality for a week of 1665

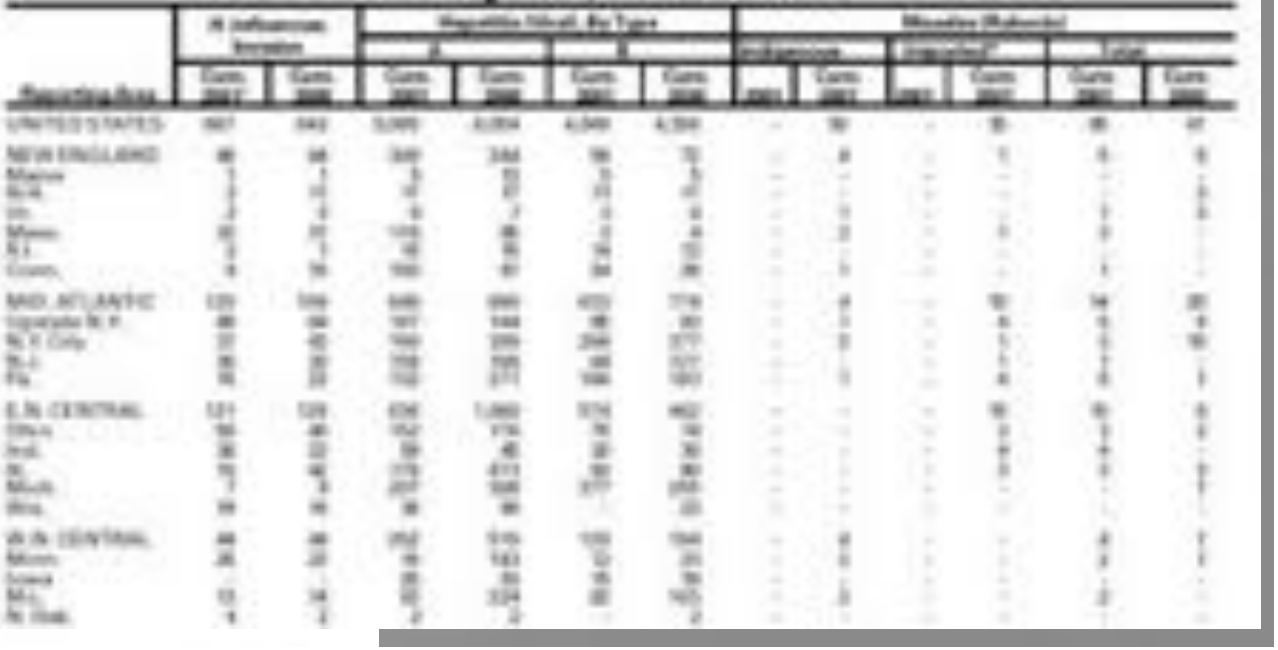
Table 3.2 Notifiable infections in the United States (1984)

Acquired Immunodeficiency Syndrome (AIDS)
Anthrax
Asthma
Aseptic meningitis
Botulism
Brucellosis
Chancroid
Cholera
Diphtheria
Encephalitis, primary
Encephalitis, post infectious
Gonorrhoea
Granuloma inguinale
Hepatitis, serum
Hepatitis, infectious
Hepatitis, unspecified
Leprosy
Leptospirosis
Lymphogranuloma venereum
Malaria
Measles
Meningococcal infections
Mumps
Pertussis
Plague
Poliomyelitis
Psittacosis
Rabies, animal
Rabies, human
Rheumatic fever, acute
Rubella
Rubella Congenital Syndrome
Salmonellosis
Shigellosis
Smallpox

Today: several infections are ‘notifiable’

CDC Morbidity and Mortality Weekly Report

TABLE III. Provisional cases of selected notifiable diseases preventable by vaccination, United States, weeks ending August 18, 2001, and August 19, 2000 (33rd Week)

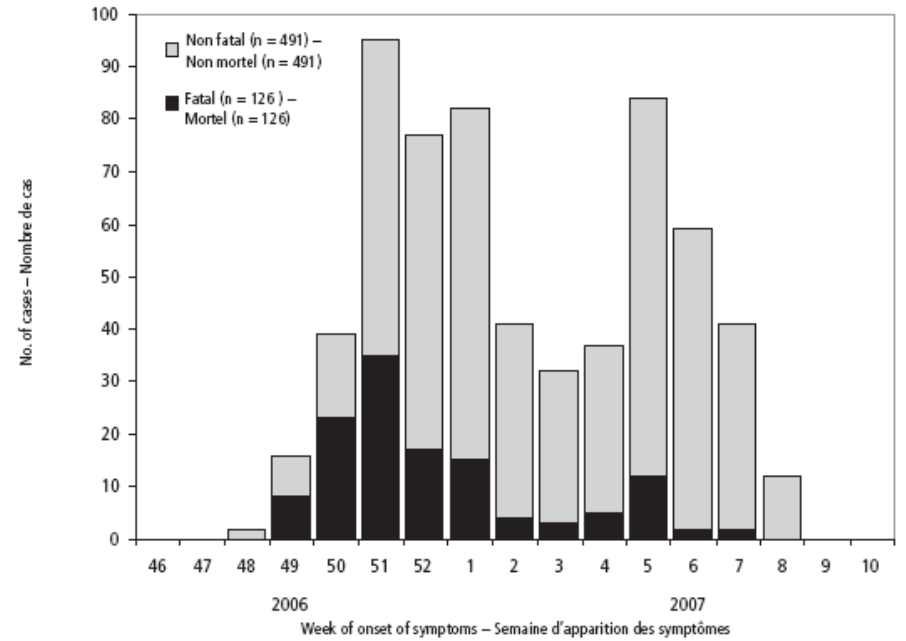


Outbreak time series

- Journal articles

Weekly epidemiological record
Relevé épidémiologique hebdomadaire

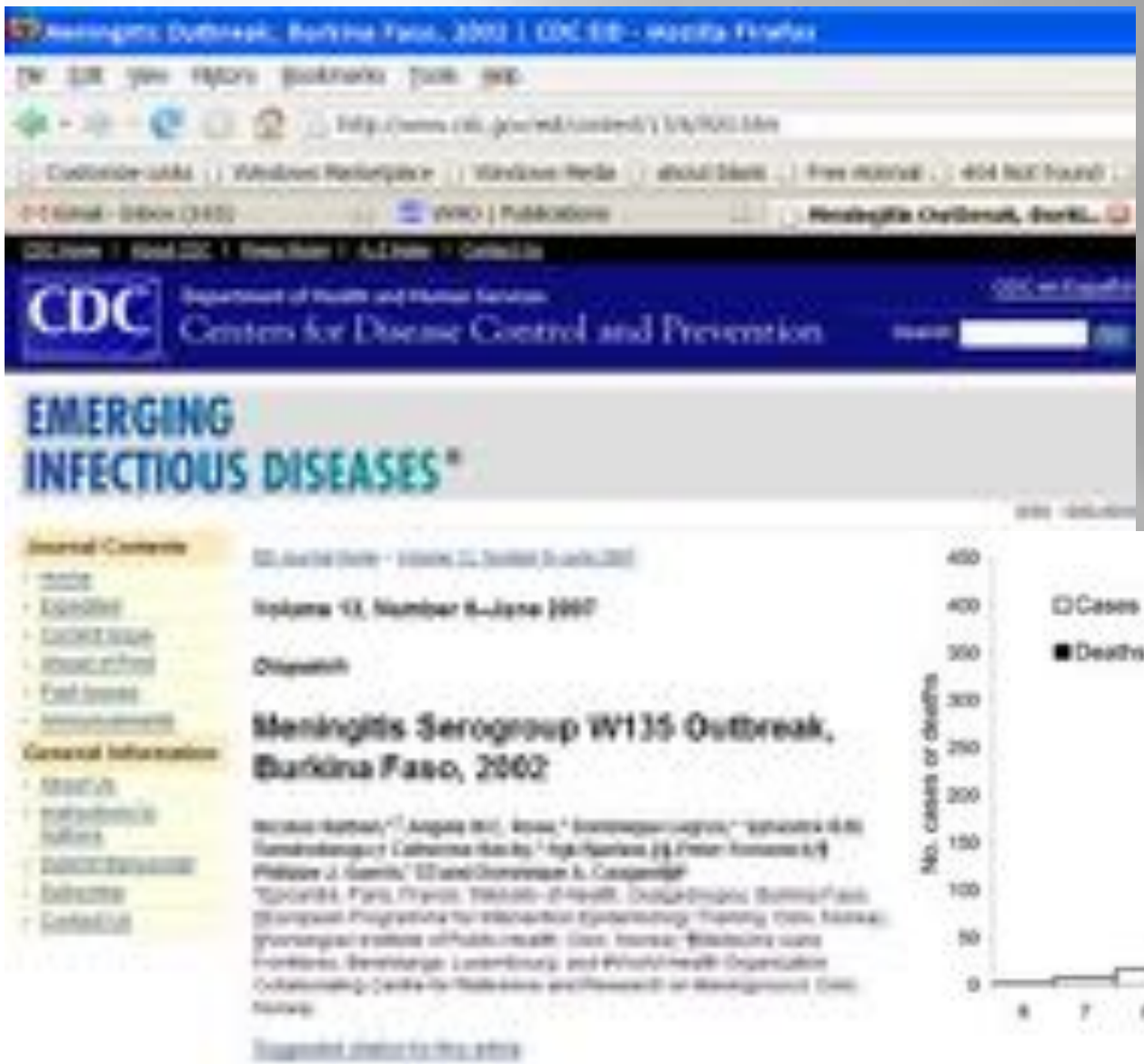
Fig. 1 **Cases of Rift Valley fever meeting inclusion criteria by date of onset of symptoms, Kenya, December 2006–February 2007 (*n* = 617)**
Fig. 1 **Cas de fièvre de la vallée du Rift satisfaisant aux critères d'inclusion par date d'apparition des symptômes, Kenya, décembre 2006-février 2007 (*n* = 617)**



http://www.who.int/wer/en/

http://www.cdc.gov/mmwr/

http://www.eurosurveillance.org



Age-incidence

Grenfell & Anderson's (1989) study of whooping cough

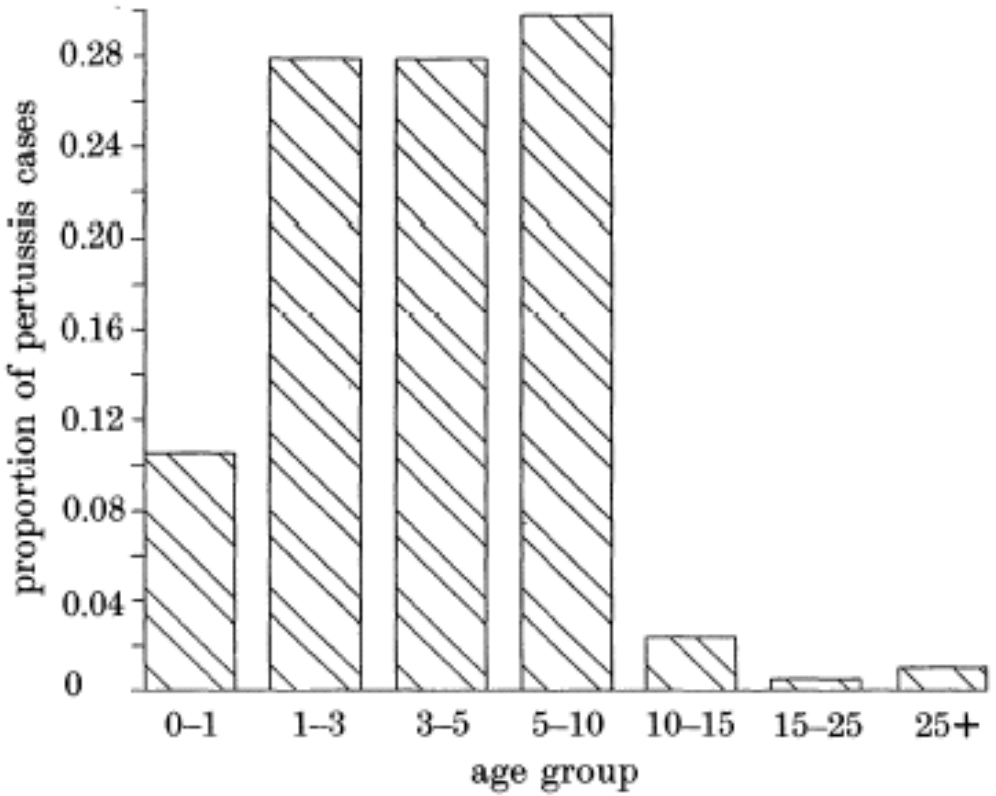
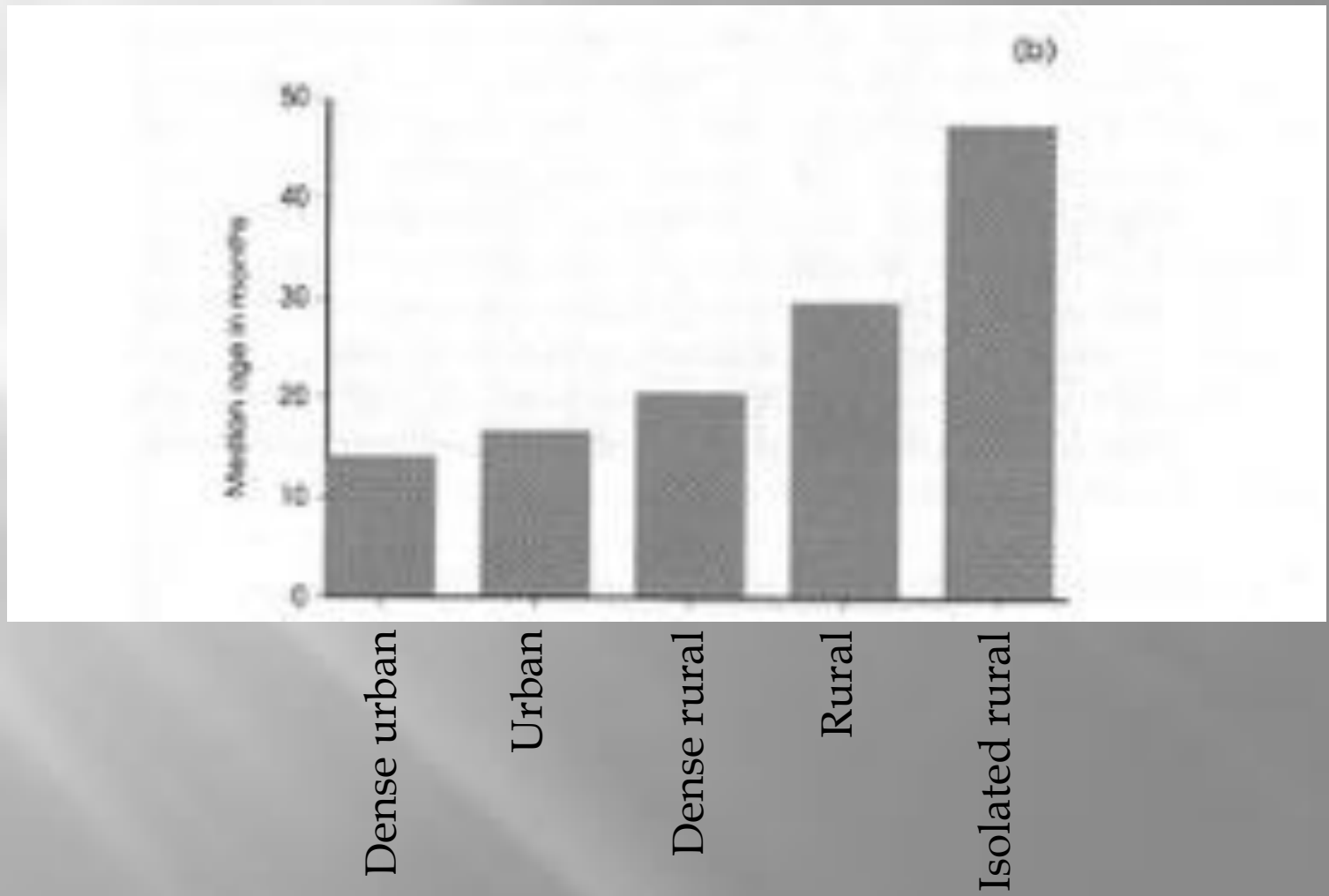


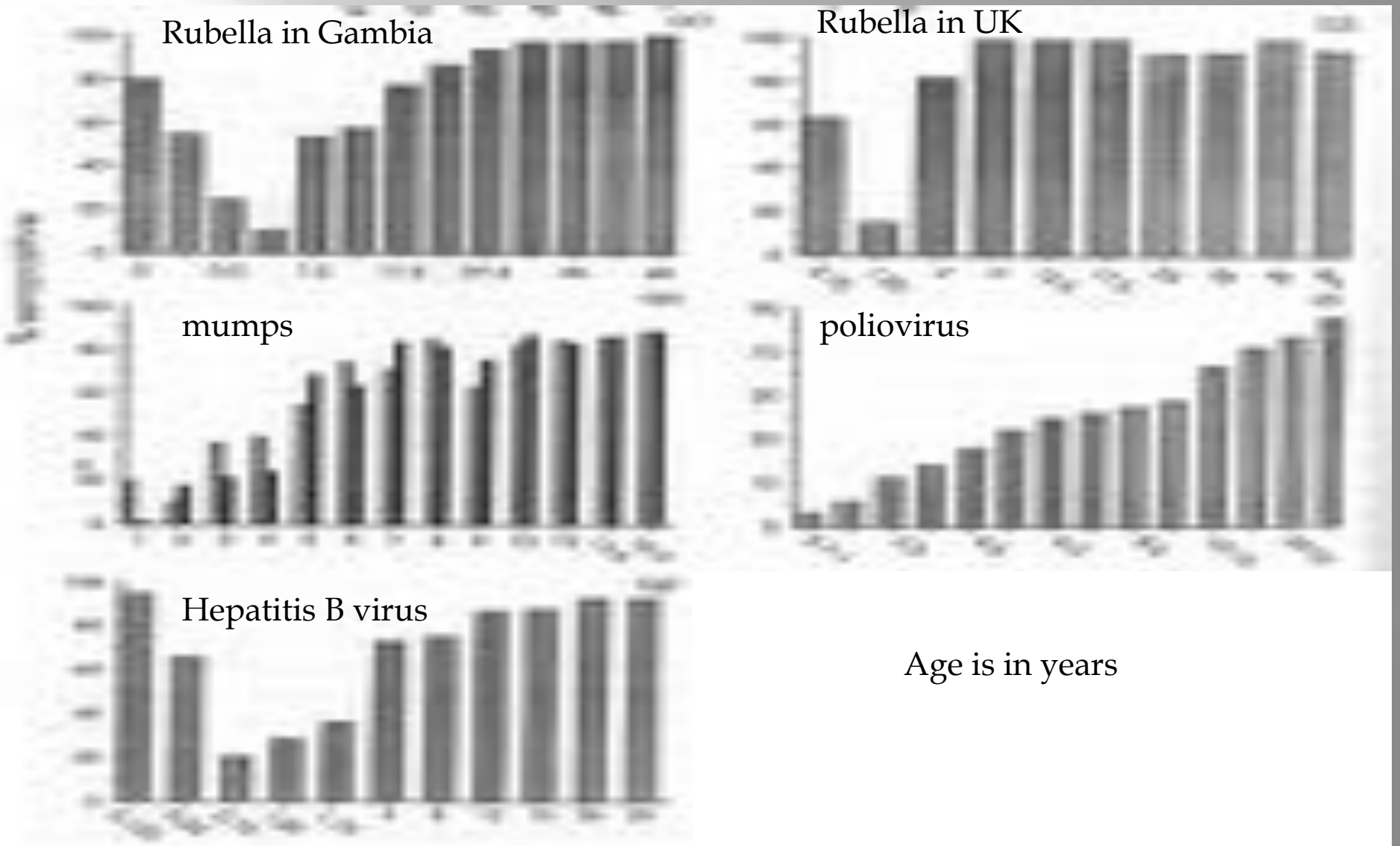
FIGURE 2. Proportional age distribution of whooping cough notifications in England and Wales for the period 1944 to 1946.

Age-incidence

e.g. Walsh (1983) of measles in urban vs rural settings in central Africa



Age-seroprevalence curves



Seroprevalence: Proportion of population carrying antibodies indicating past exposure to pathogen.

Increased transmission leaves signatures in seroprevalence profiles

e.g. measles in small (grey) and large (black) families

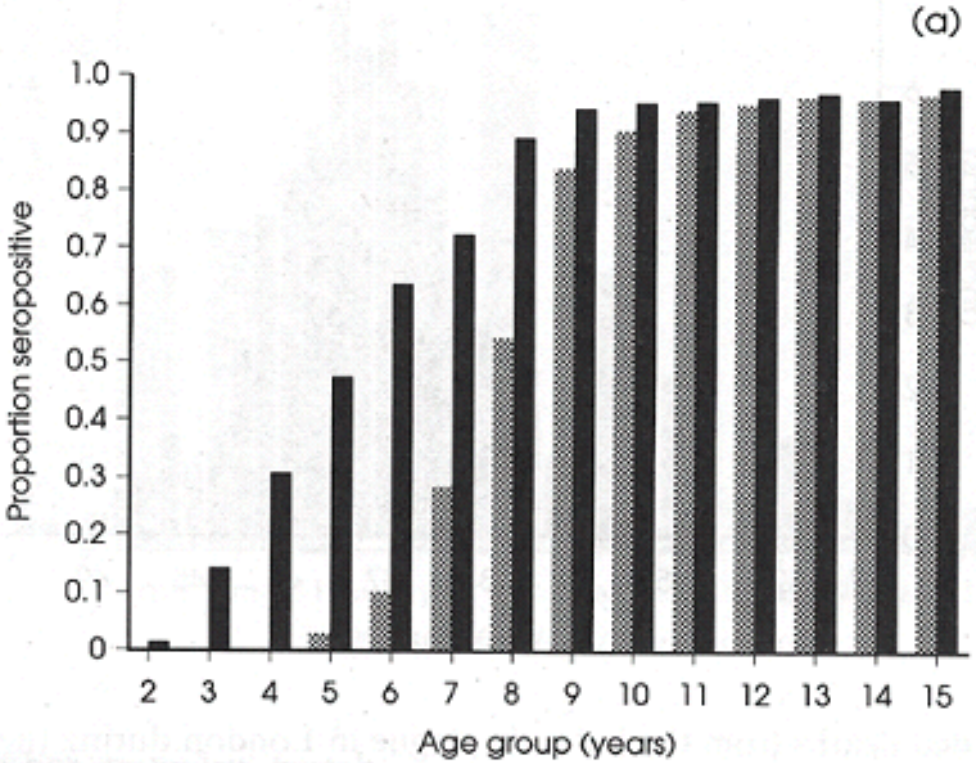
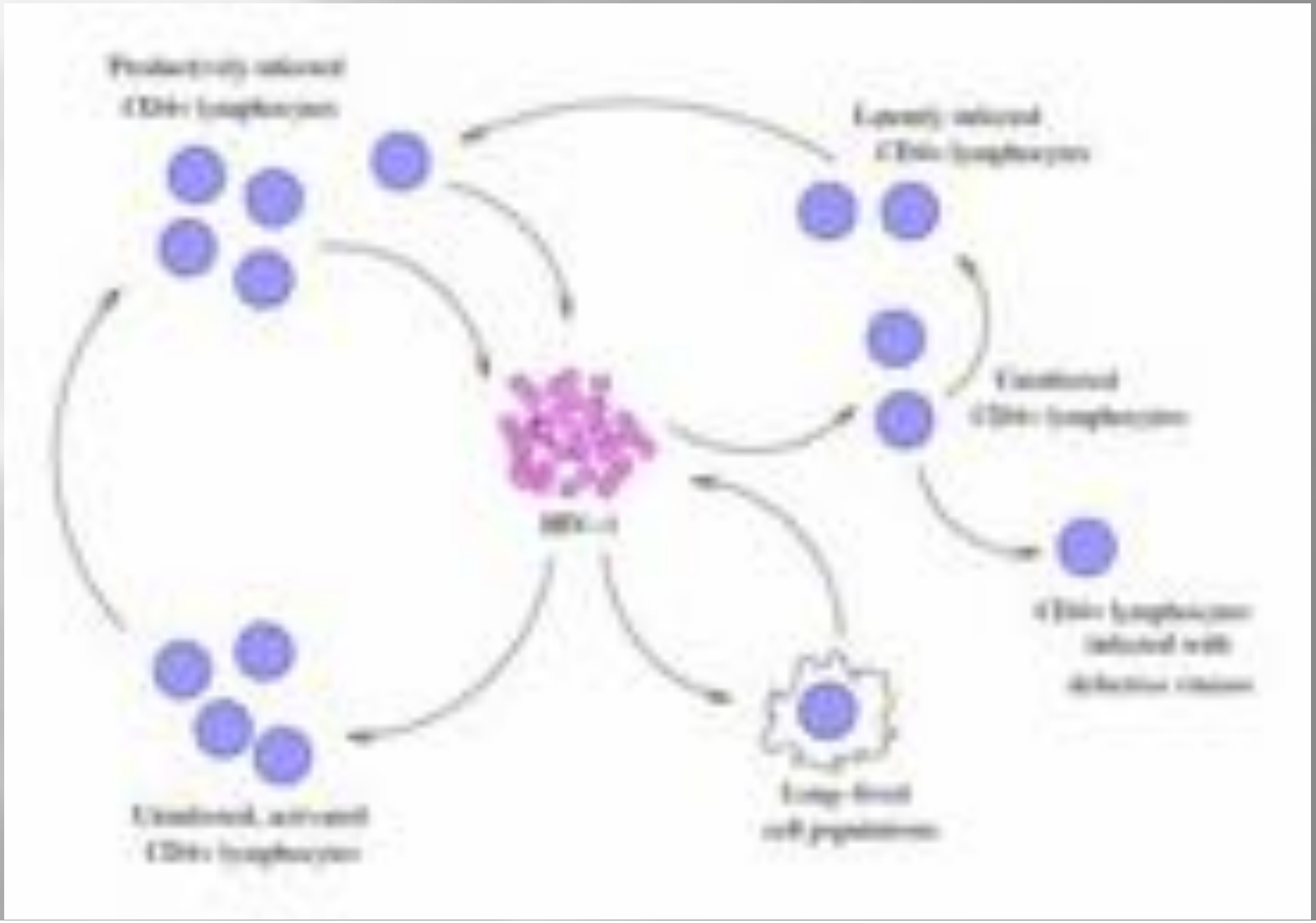


Fig. 3.10. (a) The proportion of an age group with antibodies specific to measles virus antigens in children from small and large families in the United States in 1957 prior to the introduction of mass vaccination (data from Black 1959). Family size clearly has an important influence on immunity to measles at different ages.

Other fields of disease modelling

Within-host models

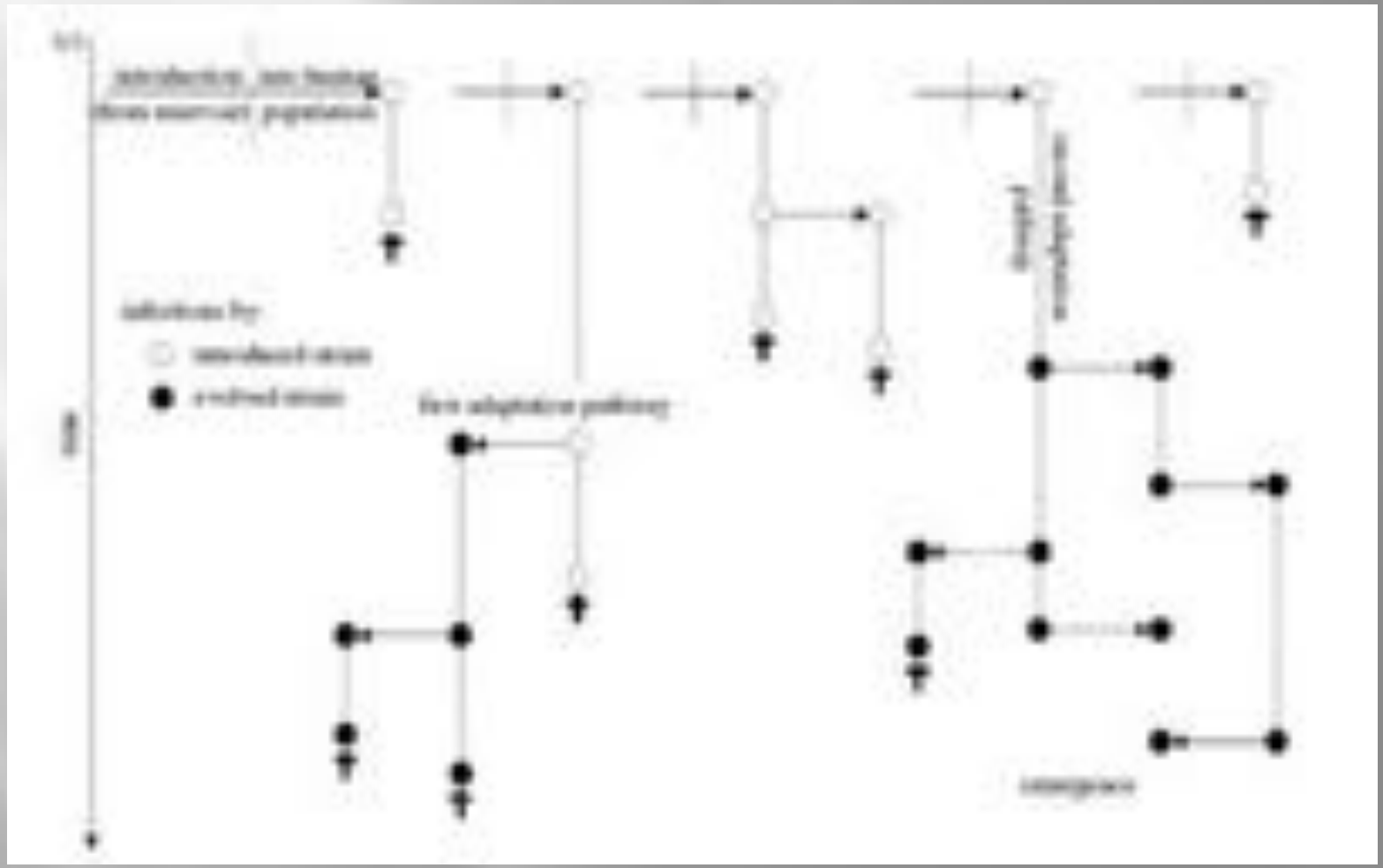
- pathogen population dynamics and immune response



Other fields of disease modelling

Pathogen evolution

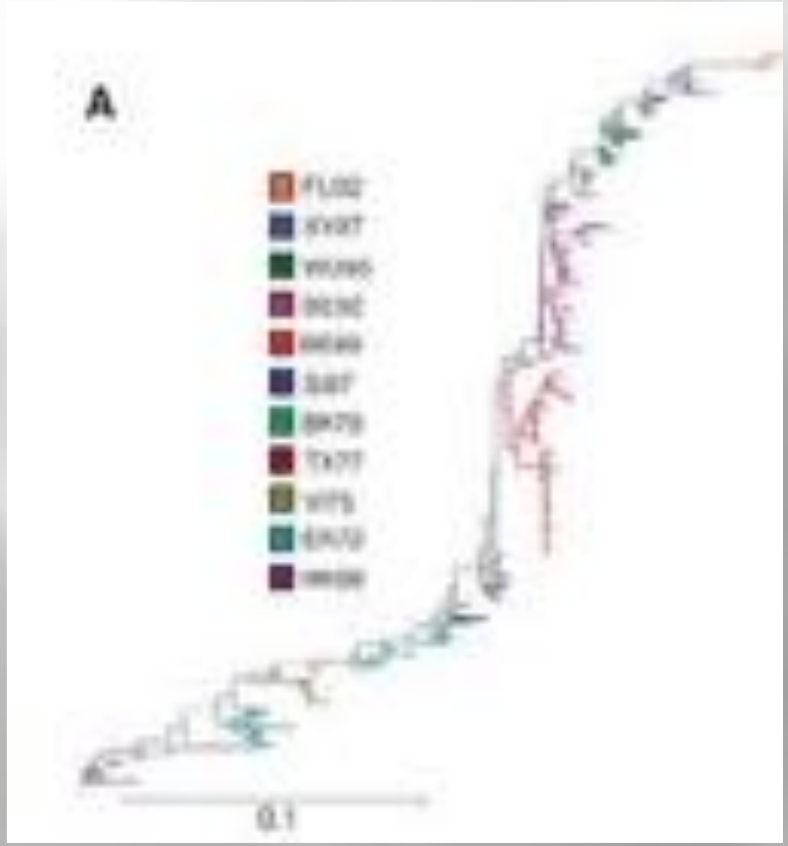
- adaptation to new host species, or evolution of drug resistance

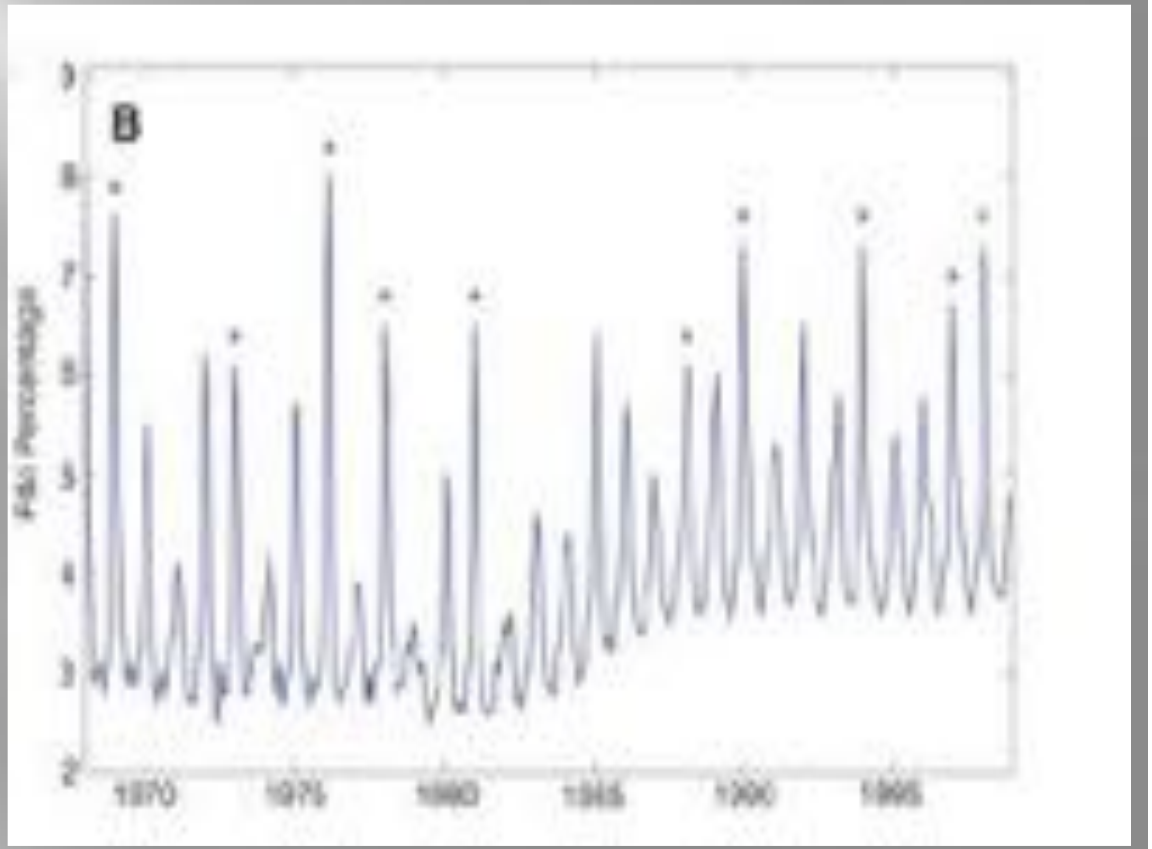


Other fields of disease modelling

Phylodynamics

- how epidemic dynamics interact with pathogen molecular evolution





Community dynamics of disease

Co-infections

What happens when multiple parasites are present in the same host?

How do they interact? Resource competition? Immune-mediated indirect competition? Facilitation via immune suppression

Multiple host species

Many pathogens infect multiple species

- when can we focus on one species?

- how can we estimate importance of multi-species effects?

Zoonotic pathogens – many infections of humans have animal reservoirs, e.g. flu, bovine TB, yellow fever, Rift valley fever

Reservoir and spillover species

Host jumps and pathogen emergence

Estimating R_0: from individual parameters

In its simplest form, $R_0 = \beta/\gamma = c\,p\,D$ where

c = contact rate

p = probability of transmission given contact

D = duration of infectiousness

So why can't we just estimate it from individual-level parameters?

Problems:

- for many diseases we can't estimate the contact rate, since "contact" is not precisely defined. The exceptions are STDs and vector-borne diseases, where contacts are (in principle) countable, though heterogeneity complicates this.

- Estimates based on R_0 expressions are highly model-dependent.

- $E(c\,p\,D) \neq E(c)\,E(p)\,E(D)$ in general.

Estimating R_0: from epidemic data

Epidemic time series data are very useful in estimating R_0. Simple analysis of the SIR model yields two useful approaches:

1) If the exponential growth rate of the initial phase of the epidemic is r, then $\boldsymbol{R_0 = 1 + rD}$

2) Equivalently, if t_d is the doubling time of the number infected, then

$$R_0 = 1 + \frac{\boldsymbol{D} \ln 2}{t_d}$$

3) If s_0 and s_∞ are the susceptible proportions before the epidemic and after it runs to completion, then

$$R_0 = \frac{\ln(s_0) - \ln(s_\infty)}{s_0 - s_\infty}$$

Estimating R_0: from epidemic data

All of those estimates are based on simple ODE models, and hence assume exponentially distributed infectious periods.

Wallinga and Lipsitch (2007, Proc Roy Soc B 274: 599-604) analyze how the distribution of the serial interval influences the relationship between r and R_0.

They find $R_0 = \frac{1}{\boldsymbol{M}(-\boldsymbol{r})}$

where $\boldsymbol{M}(\boldsymbol{z})$ is the moment generating function for the distribution of the serial interval.

→ 1. Can calculate R_0 from r for any distribution of serial interval.

→ 2. Prove that the upper bound on R_0 is $R_0 = e^{rT}$ where T is the mean serial interval.

Fitting Models to Data

Lets suppose we have data in the form of a time series for one or more of the classes in the model.

The data could be the disease prevalence or; the incidence; and sometimes, it may be the number of recovered individuals.

Curve-fitting or calibration is the process of identifying the parameters of the model so that the solution best fits the data.

What does it mean for a solution to best fit the data? Ideally, this is when the solution passes through all the data points. This type of fit is called ***interpolation.***

However, interpolation is not always the best approach to fit real data, since the data may contain errors, and capturing every tiny change in them may be impractical.

A better way to fit the solution to the data is the least-squares approach.

In the least-squares approach, we assume that the time coordinates of the data are exact, but their y-coordinates may be noisy or distorted.

We fit the solution curve through the data (see Fig. 6.1) so that the sum of the squares of the vertical distances from the data points to the point on the curve is as small as possible.

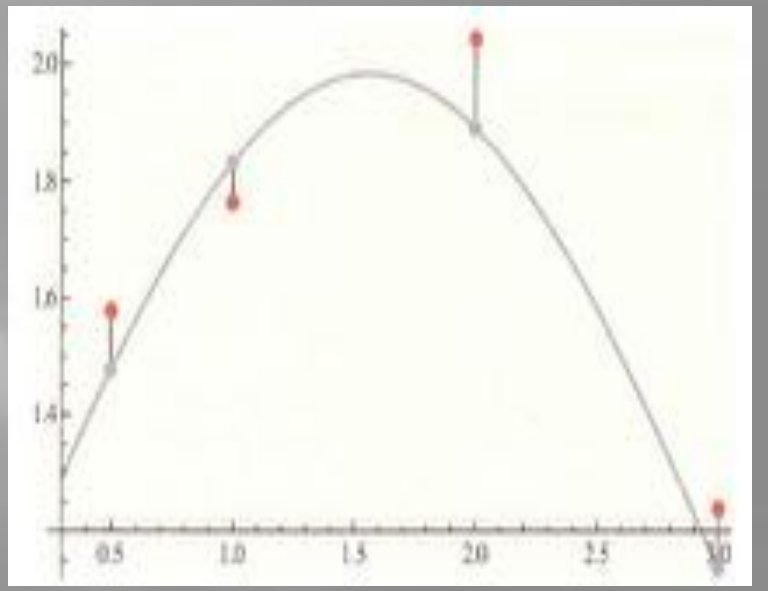


Fig. 6.1

In particular, suppose we are fitting the prevalence $I(t)$, and we are given the data $\{(t_1, Y_1), \ldots, (t_r, Y_r)\}$.

Then we consider the ***sum- of- squares error***

$$\text{SSE} = \sum_{j=1}^{n} (Y_j - I(t_j))^2.$$

The sum-of-squares error SSE is a function of the parameters of the model.

So the basic problem is to identify the parameters such that the SSE is as small as possible:

$$\text{SSE} \longrightarrow \min .$$

Minimizing the SSE is an optimization problem with its own difficulties.

Differential equation epidemic models are typically nonlinear and cannot be solved explicitly.

Hence, the resulting minimization problem is also highly nonlinear. As a result, in the general case, this problem is solved numerically with the use of computer apps like Matlab.

The code requires two basic components:

- a differential equation solver and
- a minimization routine.

The minimization is performed iteratively. The user specifies parameter values, and the computer solves the differential equations with those parameter values, evaluates the SSE, and improves the parameter values so that the SSE is reduced.

The process is repeated until the SSE no longer becomes smaller.

The minimization process is local, so depending on the initially specified parameter values, a minimization may occur for different sets of parameter values, and the SSE may be different.

Thus, it is advisable to check several sets of initial parameter values and use the smallest SSE obtained.

The English Boarding School Influenza Epidemic

In January-February 1978, an influenza epidemic occurred in a boarding school in the north of England. The boarding school housed a total of 763 boys, who were at risk during the epidemic. On January 22, three boys were sick. The table below gives the number of boys ill on the nth day after January 22 (n - 1).

Table 6.1 Daily number influenza infected boys

Day	No. infected[a]	Day	No. infected
3	25	9	192
4	75	10	126
5	227	11	71
6	296	12	28
7	258	13	11
8	236	14	7

[a] Data taken from "Influenza in a Boarding School," *British Medical Journal*, 4 March 1978

To fit with Matlab, we do not need to know the final size of the epidemic. Once we have the data, the first question that we have to answer is, what model we should fit to the data? Since these are outbreak data, we need an epidemic model without demography.

The SIR model without demography is appropriate for this case:

$$S'(t) = -\beta S(t)I(t),$$
$$I'(t) = \beta S(t)I(t) - \alpha I(t),$$

Next we determine which model parameters we should fit and which parameter we should pre-estimate and fix.

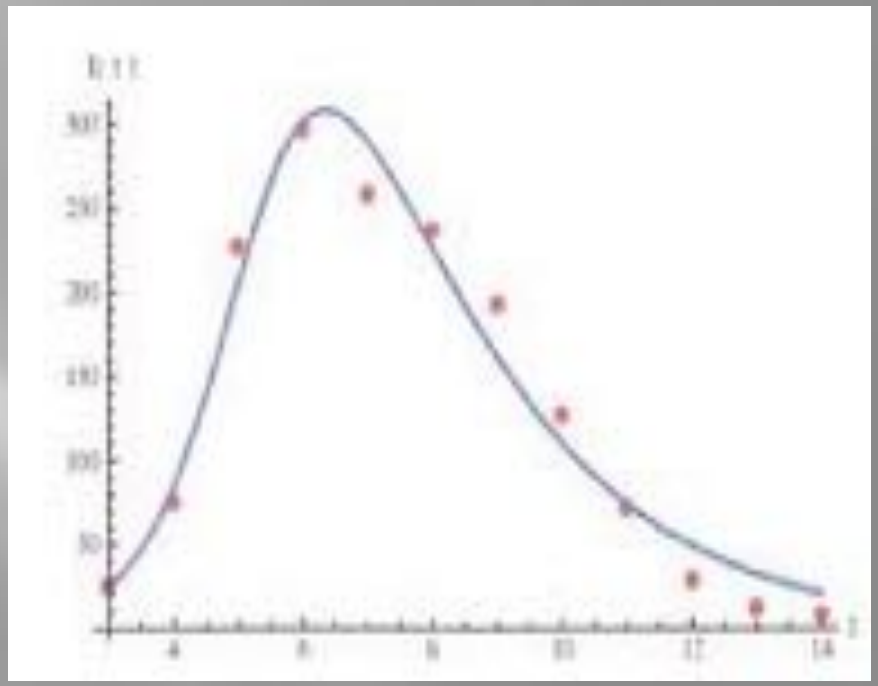
We can fit α, β, and the initial conditions-four parameters altogether. We can pre-estimate α from the duration of infectiousness, and the two initial conditions from the given data.

For instance, we know from the data that $I(3) = 25$, and therefore $S(3) = 763 - 25 = 738$.

The duration of infectiousness is 2-4 days, so we may take $\alpha = 0.30$. Even if we plan to fit all these parameters, pre-estimating what we can is useful with the initial guess of the parameters.

Before fitting the model to the data using Matlab, the original SSE $= 7.2 \times 10^4$. After the optimization. the newly computed $\alpha = 0.465$ and $\beta = 0.00237$ and SSE $= 4 \times 10^3$.

The newly computed value of α gives the duration of the infectious period as $1/\alpha = 2.15$ days. This infectious period is meaningful, since infected students showing symptoms were quarantined.



Always ask whether the computed parameters have a sensible biological interpretation. If that is not the case we should refit, using upper and lower bounds for the parameters.

Summary of Basic Steps

When you prepare to fit a mathematical model to data, think about the following basic steps in the fitting process:

1. Examine your data. Are the values involved too large or too small? If yes, determine units that allow you to work with average-size numbers.

2. Choose your model. Is your model sensible for the disease you are modeling? Should your model include demography? Decide whether your data are epidemic or endemic. What is the time span modeled?

3. Decide which model parameters to fit and which to pre-estimate and fix. Don't forget that the initial conditions for the differential equations are also in the parameter Set. *Never fit more parameters than the number of data points.*

4. Choose initial guesses for the parameters that will be fitted. Use biological sense or prefit.

5. Perform the fit. Plot the solution alongside the data and examine the fit. Does the solution agree with the data?

Plot the residuals. Are the residuals small and random? If they are not random, you may need a better model.

6. Determine the best-fitted parameters. Interpret them biologically. Do they make sense? If not, refit specifying upper and lower bounds for those parameters.

7. Determine the standard errors and 95% Cl. Are they small? If they are not small, that may mean that some of the parameters are unidentifiable. Refit, fixing some more parameters.

References

1. Matt J. Keeling and Pejman Rohani.
Modeling Infectious Disease in Humans and Animals. Princeton University Press 2008.

2. Jamie Lloyd-Smith
Introduction to infectious diseases
Center for Infectious Disease Dynamics, Pennsylvania State University

3. Steve Bellan Introduction to Infectious Disease Modelling Clinic on the Meaningful Modeling of Epidemiological Data, (2015)
African Institute for Mathematical Sciences Muizenberg, South Africa.

4. Jamie Lloyd-Smith
Incidence functions and population thresholds
Center for Infectious Disease Dynamics, Pennsylvania State University

5. Nakul Chitnis, James M. Hyman, Jim M. Cushing,
Determining Important Parameters in the Spread of Malaria Through the Sensitivity Analysis of a Mathematical Model. Bulletin of Mathematical Biology (2008)
DOI 10.1007/s11538-008-9299-0

Fitting Models to Data

- Develop a model
- Validate the model by comparing it with data.

Model validation is the process of determining the degree to which a mathematical model accurately represent the real-world data.

Where does the data come from?

- Biologists or epidemiologists
- Health organization such as WHO, CDC, and
- Published articles.

The article data can be extracted using

- PlotDigitizer
- Matlab app (grabit)

Fitting Models to Data

Suppose we have time series data for one or more of the classes in the model.

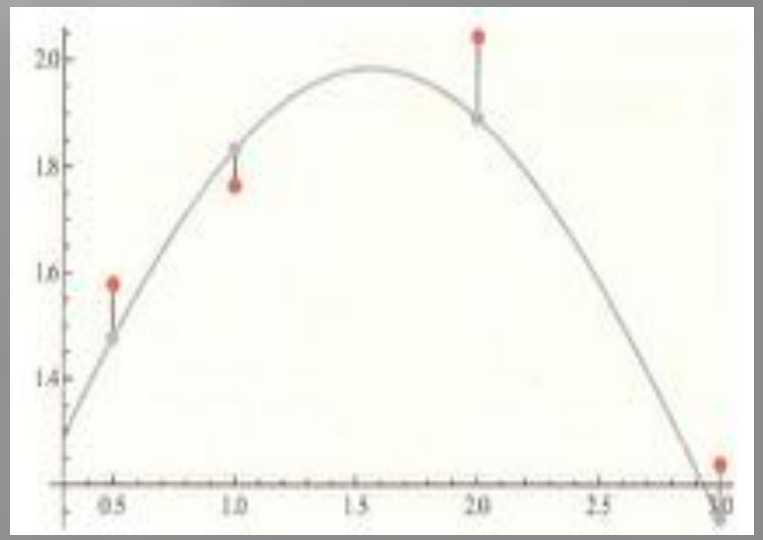
The data could be the disease prevalence or; the incidence; and sometimes, it may be the number of recovered individuals.

We can fit the solution of model to data using lease-squares approach.

This approach require us to minimize the ***sum- of- squares error***

$$\text{SSE} \longrightarrow \min.$$

$$\text{SSE} = \sum_{j=1}^{n} (Y_j - I(t_j))^2.$$



Model Fitting

So the basic problem is to identify the parameters such that the SSE is as small as possible:

The code requires two basic components:

- a differential equation solver and
- a minimization routine.

The minimization is performed iteratively.

- Specify the parameter values, and
- Solves the model with those parameter values,
- Evaluates the SSE, and improves the parameter values so that the SSE is reduced.
- Repeated until the SSE no longer becomes smaller.
- Check if computed parameters have a sensible biological interpretation.
- If not repeat steps above.

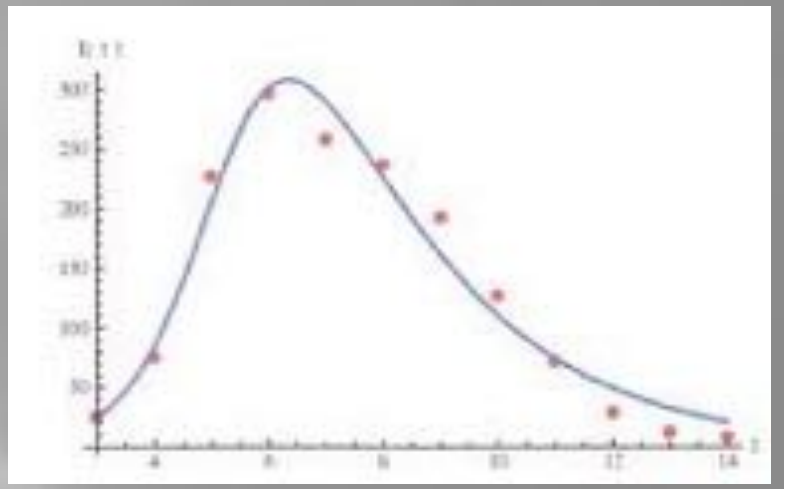
Fitting Models to Data

Table 6.1 Daily number influenza infected boys

Day	No. infected[a]	Day	No. infected
3	25	9	192
4	75	10	126
5	227	11	71
6	296	12	28
7	258	13	11
8	236	14	7

[a] Data taken from "Influenza in a Boarding School," *British Medical Journal*, 4 March 1978

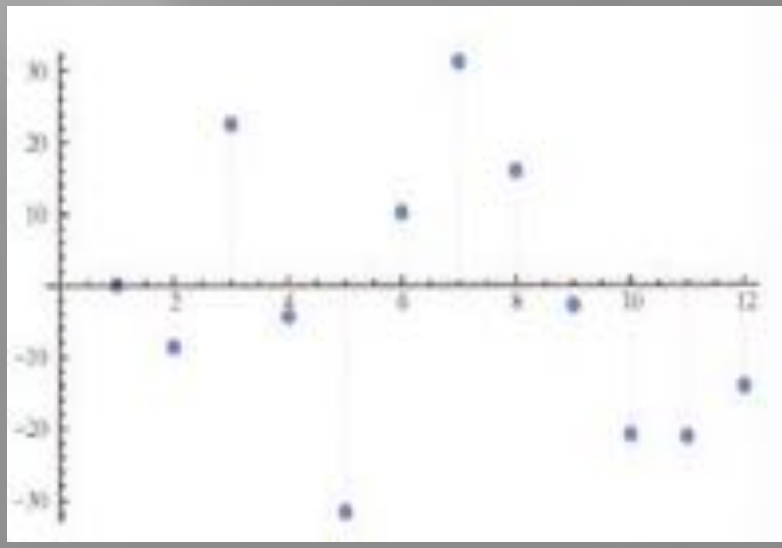
$$S'(t) = -\beta S(t)I(t),$$
$$I'(t) = \beta S(t)I(t) - \alpha I(t),$$



Fit $\alpha, \beta,$ and the initial conditions.

Pre-estimate α from the duration of infectiousness, and the two initial conditions from the given data.

From the data $I(3) = 25$, and $S(3) = 738$. $\alpha = 0.30$ from the 2-4 days duration of infectiousness.
After the optimization. $\alpha = 0.465$ and $\beta = 0.00237$ and SSE $= 4 \times 10^3$.



Fitting World HIV/AIDS Prevalence

HIV infection is a disease of the immune system caused by the HIV virus.

It is transmitted primarily via unprotected sexual intercourse, contaminated blood transfusions, and vertically from mother to child during pregnancy, delivery, or breastfeeding.

The virus causes acute infection upon entering the body, with flulike symptoms. The acute infection is followed by a long asymptomatic period.

As the illness weakens the immune system as it progresses, making the infected individual much more likely to get other opportunistic infections.

There is no cure or vaccine against HIV; however, antiretroviral treatment can slow the course of the disease and may lead to a near-normal life expectancy.

People with HIV now live longer, and the incidence of HIV is declining, although the number of individuals infected with HIV or having advanced-stage AIDS is still slowly increasing worldwide.

The number of people living with HIV as well as the incidence and the number of deaths from HIV worldwide can be found United Nations Millennium Development Goals Report 2010.

The prevalence data is given below

Table 6.2 Prevalence (in millions) of HIV worldwide 1990-2011, 1990 gives $t = 0$

Year	Time (in years)	Prevalence	Year	Time (in years)	Prevalence
1990	0	7.3	2001	11	29.0
1991	1	9.2	2002	12	30.0
1992	2	11.3	2003	13	30.8
1993	3	13.5	2004	14	31.4
1994	4	15.9	2005	15	31.9
1995	5	18.3	2006	16	32.4
1996	6	20.6	2007	17	32.8
1997	7	22.7	2008	18	33.4
1998	8	24.6	2009	19	33.3
1999	9	26.3	2010	20	34.0[a]
2000	10	27.8	2011	21	34.0[a]

[a]Other sources

Our main objective is to determine a model that can be fitted to the data. The simplest HIV model is the SI epidemic model with disease-induced mortality.

However, this model does not fit the data well.

The main reason for that, is the fact that a simple SI model has an exponential distributed time spent in the infectious stage, that is, the probability of surviving in the stage declines exponentially.

That is not very realistic for HIV where the infectious stage is long and the duration is subject to significant variation. That requires that the distribution of the waiting time in the infectious class have a nonzero mode.

To incorporate this effect, a typical approach is to use ***Erlang's "method of stages".***

This approach is primarily applied with stochastic HIV models, but its deterministic variant requires the infectious period to be represented as a series of k stages such that the duration of stay in each stage are independent identically distributed exponential variables.

To this end, we divide the infectious class $I(t)$ into to four classes: $I_1(t), I_2(t), I_3(t), I_4(t)$ with an exit rate γ.

Individuals in all four stages are infectious and can infect susceptibles individuals $S(t)$.

Denote by $I(t)$ the sum of all infectious classes:

$$I(t) = I_1(t) + I_2(t) + I_3(t) + I_4(t).$$

Assume the force of infection $\lambda(t)$ is nonmonotone and is given by

$$\lambda(t) = \beta e^{-\alpha I(t)/N(t)} I(t)/N(t),$$

where $N(t)$ denotes the total population size:

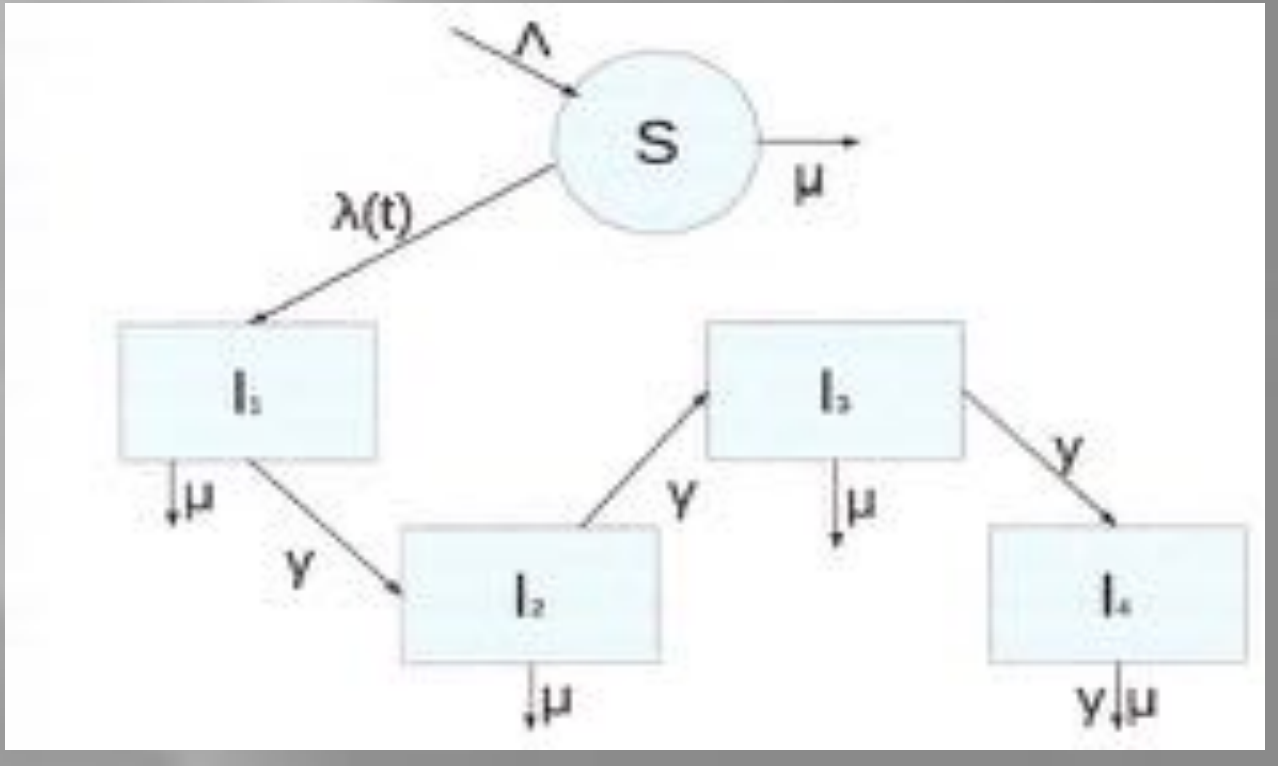
$$N(t) = S(t) + I_1(t) + I_2(t) + I_3(t) + I_4(t).$$

$$
\begin{aligned}
S'(t) &= \Lambda - \lambda(t)S(t) - \mu S(t),\\
I_1'(t) &= \lambda(t)S(t) - (\gamma+\mu)I_1(t),\\
I_2'(t) &= \gamma I_1(t) - (\gamma+\mu)I_2(t),\\
I_3'(t) &= \gamma I_2(t) - (\gamma+\mu)I_3(t),\\
I_4'(t) &= \gamma I_3(t) - (\gamma+\mu)I_4(t).
\end{aligned}
$$

This force of infection is sensible for HIV since as the infection spreads, it is likely that the remaining susceptible individuals become more cautious about their contacts and potential exposure to HIV and the force of infection begins to decline.

The last exit rate γ from the class I_4 is considered to be disease-induced mortality.

Fit the model to the data given above.



To fit the model to the data, first decide on the units to fit. The data are given in millions, and as such, they are neither too large nor too small as numbers.

The round-off errors may be large, and the fit may be bad, if the numbers we fit are too large or too small. Therefore, use units that make our numbers reasonable.

Furthermore, fit in years. Next decided on the units, the parameters to fit, and which to pre-estimate.

We decide to fix Λ and μ as well as the initial values.

The natural mortality rate of humans can be taken to be $1/70$.

Because the current world population size is 7 billion, that is, 7000 million, then if we take that to be the equilibrium population, we have $7000 = \Lambda/\mu$.

We estimate that $\Lambda = 100$ million people per year. We further assume that in 1990, all individuals infected with HIV were actually in class I_1. Hence, S(0) = 6992.7 and $I_1(0) = 7.3$ million people.

We set the remaining initial conditions to zero. In this fitting, we do not fit the initial conditions.

We fit α, β, and γ and obtain
$\alpha = 260.4972$, $\beta = 0.334547$, and $\gamma = 0.339958755$.
The SSE = 0.47 with these parameters.

Home Work

Fitting the Incidence of TB in the United States

After a mild increase in tuberculosis (TB) cases in the United States in the late 1980s and the beginning of the 1990s, the incidence (number of new cases per year) of TB has been steadily declining. Table 6.8 gives the TB incidence starting in 1990.

Table 6.8 Number of new cases of TB in USA[a]

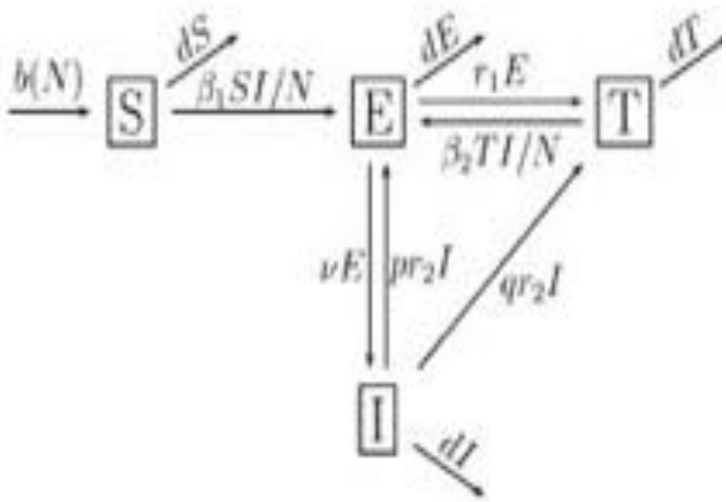
Year	Number cases	Year	Number cases	Year	Number cases
1990	25,701	1998	18,287	2006	13,732
1991	26,283	1999	17,500	2007	13,286
1992	26,673	2000	16,309	2008	12,905
1993	25,107	2001	15,945	2009	11,537
1994	24,205	2002	15,055	2010	11,182
1995	22,727	2003	14,835	2011	10,528
1996	21,210	2004	14,499	2012	–
1997	19,751	2005	14,068	2013	–

[a] http://www.cdc.gov/tb/statistics/reports/2010/default.htm

Fit the following TB model to the data above

$$
\begin{aligned}
S'(t) &= \Lambda - \beta_1 SI/N - \mu S, \\
E'(t) &= \beta_1 SI/N + \beta_2 TI/N - (\mu + \kappa + r_1)E + pr_2 I, \\
I'(t) &= \kappa E - (r_2 + \mu)I, \\
T'(t) &= r_1 E + qr_2 I - \beta_2 TI/N - \mu T,
\end{aligned}
$$

where $T(t)$ is the number of treated individuals,
$I(t)$ is the number of individuals with active TB,
$E(t)$ is the number of exposed,
r_1 is the treatment rate of exposed individuals,
r_2 is the treatment rate of infectious individuals,
k is the progression to the infectious state.
We assume that $p + q = 1$.



Hint: (l) You should be fitting the incidence $\beta_1 SI/N + \beta_2 TI/N$ to the data
(2) Set the incidence as a new equation in the code
(3) You may fit the initial conditions or take them as follows:
$S(0) = 290{,}000{,}000, E(0) = 25{,}000, I(0) = 25{,}000, T(0) = 22{,}000$