# Parameter Estimation

## Parameter Estimation

- We've seen that basic reproductive ratio, $R_0$, is a central and useful quantity
- How do we calculate it from data?
- Review some simple methods
- In general, how do we achieve parameter estimation from epidemiological data?

# 1a. Final outbreak size

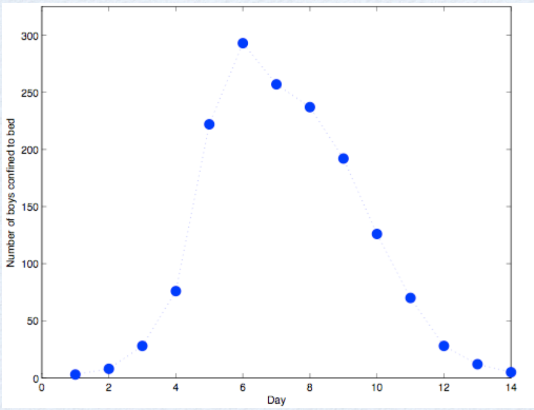
- From lecture 1: for an SIR system at end of epidemic:
  - $S(\infty) = 1 - R(\infty) = S(0)\, e^{-R(\infty) R_0}$

- So, if we know population size (N) , initial susceptibles (to get S(0)), and total number infected (to get R(∞)), we can numerically calculate $R_0$

Note: Ma & Earn (2006) showed this formula is valid even when numerous assumptions underlying simple SIR are relaxed

# 1. Final outbreak size

- Worked example:




Influenza epidemic in a British boarding school in 1978

$N = 764$
$X(0) = 763$
$Z(\infty) \sim 512$

Gives

$R_0 \sim 1.65$

# 1B. Final outbreak size

- Becker (1989) showed that with more information, we can also estimate $R_0$ from

$$R_0 = \frac{(N-1)}{C}\ln\left\{\frac{X_0+\frac{1}{2}}{X_f-\frac{1}{2}}\right\} \qquad (\sim 1.66)$$

- Again, we need to know population size (N) , initial susceptibles ($X_0$), total number infected (C)
- Usefully, standard error for this formula has also been derived

$$SE(R_0) = \frac{(N-1)}{C}\sqrt{\sum_{j=X_f+1}^{X_0}\frac{1}{j^2}+\frac{CR_0^{\ 2}}{(N-1)^2}} \qquad (\sim 1.85)$$

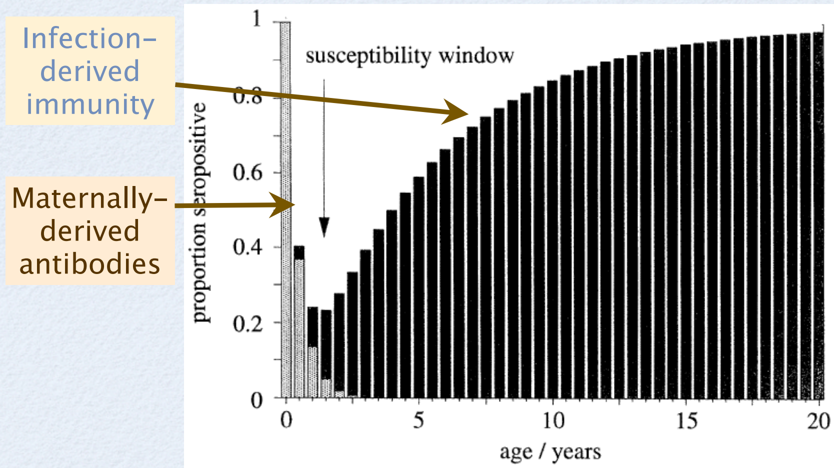
# 2. Independent data: mean age at infection

- An epidemiologically interesting quantity is mean age at infection – how do we calculate it in simple models?
- From first principles, it's mean time spent in susceptible class
- At equilibrium, this is given by $1/(\beta I^*)$, which (we'll see in next lecture) leads to

$$A \approx \left(\frac{1}{\mu(R_0-1)}\right)$$

- This can be written as $R_0$-1 ≈ L/A (L= life expectancy)
- Historically, this equation's been an important link between epidemiological estimates of A and deriving estimates of $R_0$

# Measles age-stratified seroprevalence




Mean age at infection is ~4.5 yrs

Assume L~75, ⇒ $R_0$ ~ 16.6

# Historical Significance

Anderson & May (1982; *Science*)

Table 2. The intrinsic reproductive rate, $R_0$, and average age of acquisition, $A$, for various infections [condensed from (25); see also (36)]. Abbreviations: r, rural; u, conurbation.

| Disease | Average age at infection, $A$ (years) | Geographical location | Type of community | Time period | Assumed life expectancy (years) | $R_0$ |
|---|---|---|---|---|---|---|
| Measles | 4.4 to 5.6 | England and Wales | r and u | 1944 to 1979 | 70 | 13.7 to 18.0 |
| | 5.3 | Various localities in North America | r and u | 1912 to 1928 | 60 | 12.5 |
| Whooping cough | 4.1 to 4.9 | England and Wales | r and u | 1944 to 1978 | 70 | 14.3 to 17.1 |
| | 4.9 | Maryland | u | 1908 to 1917 | 60 | 12.2 |
| Chicken pox | 6.7 | Maryland | u | 1913 to 1917 | 60 | 9.0 |
| | 7.1 | Massachusetts | r and u | 1918 to 1921 | 60 | 8.5 |
| Diphtheria | 9.1 | Pennsylvania | u | 1910 to 1916 | 60 | 6.6 |
| | 11.0 | Virginia and New York | r and u | 1934 to 1947 | 70 | 6.4 |
| Scarlet fever | 8.0 | Maryland | u | 1908 to 1917 | 60 | 7.5 |
| | 10.8 | Kansas | r | 1918 to 1921 | 60 | 5.5 |
| Mumps | 9.9 | Baltimore, Maryland | u | 1943 | 70 | 7.1 |
| | 13.9 | Various localities in North America | r and u | 1912 to 1916 | 60 | 4.3 |
| Rubella | 10.5 | West Germany | r and u | 1972 | 70 | 6.7 |
| | 11.6 | England and Wales | r and u | 1979 | 70 | 6.0 |
| Poliomyelitis | 11.2 | Netherlands | r and u | 1960 | 70 | 6.2 |
| | 11.9 | United States | r and u | 1955 | 70 | 5.9 |

# 3. Epidemic take-off

More common approach is to study epidemic take off

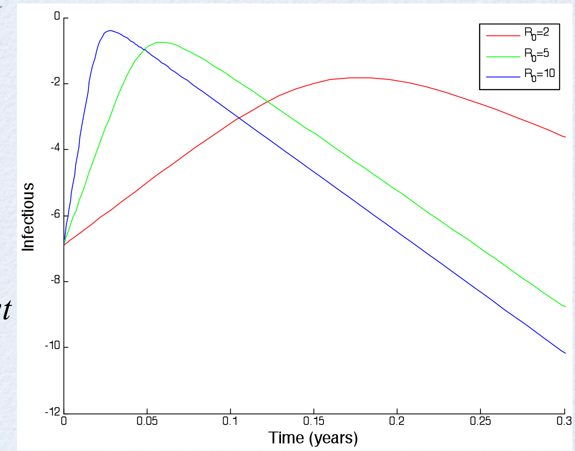
Recall from linear stability analysis that

$$I_{SIR} \approx I(0) \times e^{(R_0 - 1)\gamma t}$$

Take logarithms

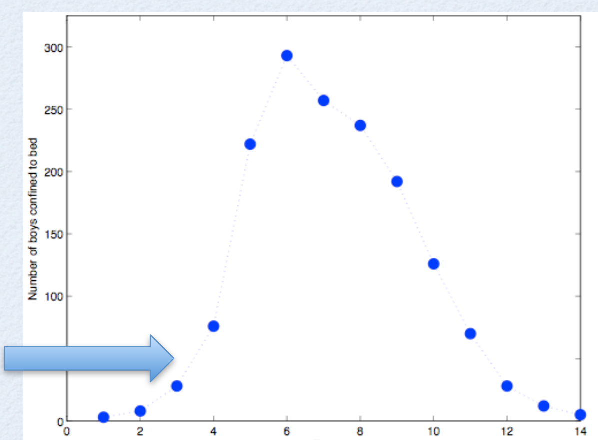
$$\log(I_{SIR}) = \log(I(0)) + (R_0 - 1)\gamma t$$

So, regression slope will give $R_0$




# 3. Epidemic take-off

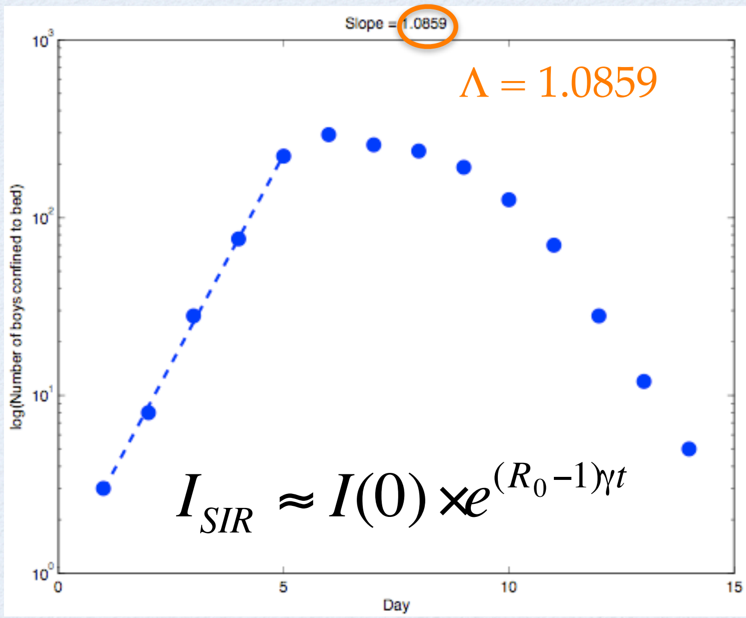
- Back to the school boys




Influenza epidemic in a British boarding school in 1978

Looks like classic exponential take-off

# Epidemic take-off

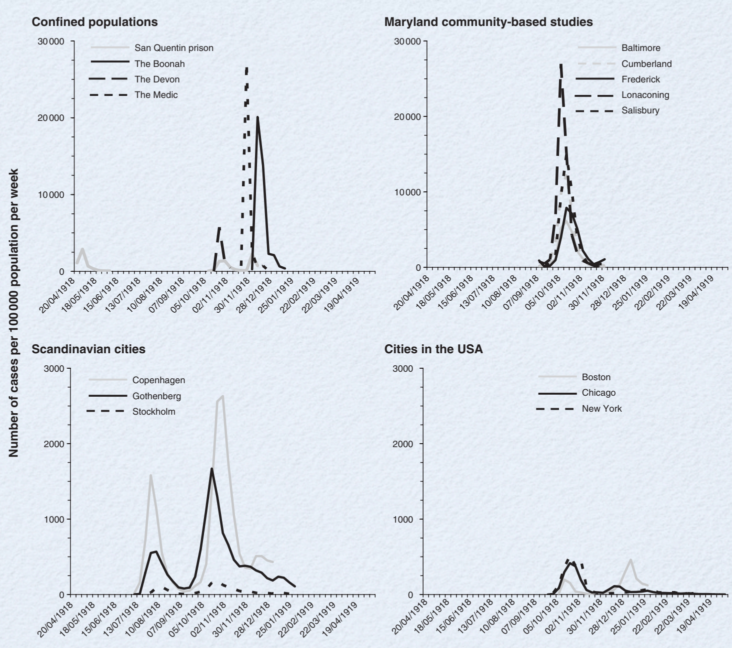



Our (guessed?) value for 'flu incubation period

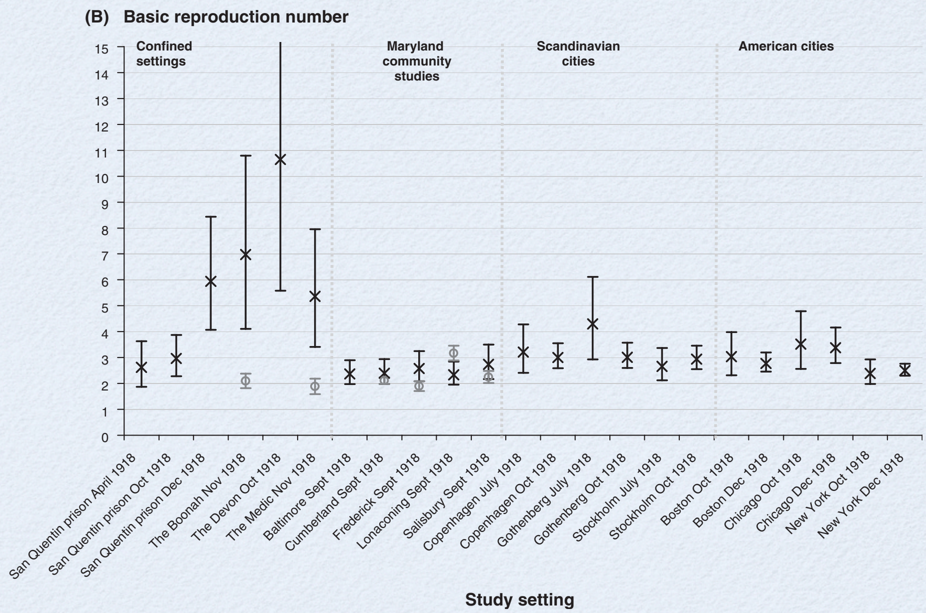
So, $R_0$=1.0859*2.5+1

= 3.71

# Vynnycky et al. (2007)



# Vynnycky et al. (2007)




# Variants on this theme

- Recall

  $$\log(I_{SIR}) = \log(I(0)) + (R_0 - 1)\gamma t$$

- Let $T_d$ be 'doubling time' of outbreak
- Then,
  - $R_0 = \log(2)/(T_d \gamma) + 1$

# 4. Likelihood & Estimation

- Given some epidemiological data, a model and some parameter values, "likelihood" is a measure of model's appropriateness as a descriptor of reality
- L(model | data) = Pr(data | model)

- Assume we have data, D, and model output, M (both are vectors containing state variables). Model predictions are generated using set of parameters, θ.

# 4. Likelihood & Estimation

- Data, D
- Model output, M
- Parameters, θ

- If (we assume) errors are normally distributed, with mean μ and variance $\sigma^2$ then

$$L(M(\theta) \mid D) = \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(D_i - M_i)^2}{2\sigma^2}}$$

Also assumes likelihood of sequential observations independent – sensible?

# 4. Likelihood & Estimation

- Data, D
- Model output, M
- Parameters, θ

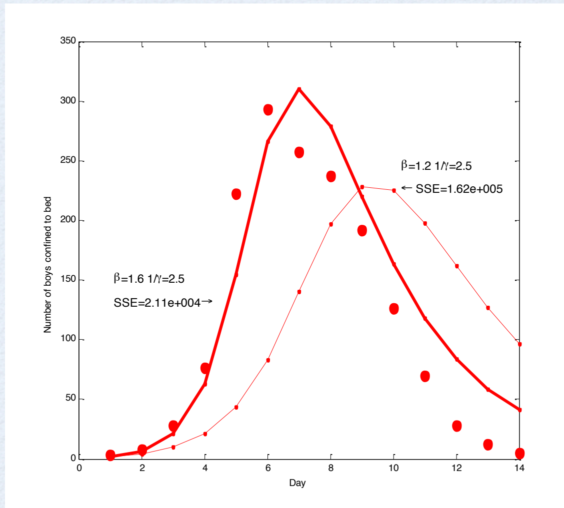
- Often easier to deal with Log-likelihoods:

$$\log\left(L(M(\theta)\mid D)\right) = -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_i (D_i - M_i)^2$$

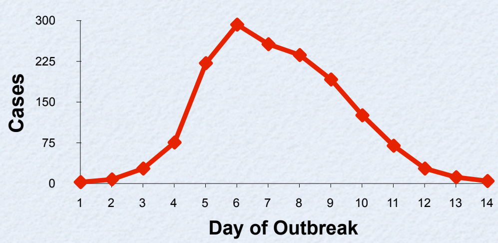
# 4. Likelihood & Estimation

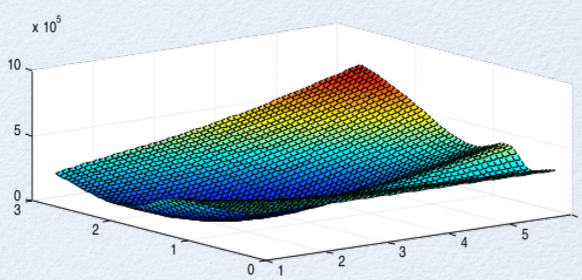
- Want parameter set θ that gives predictions with smallest deviation from data
- Can quantify goodness of fit using least square errors
  - $SSE = \Sigma(D_i - M_i)^2$
- Then miminise SSE to arrive at best parameter estimates

# Curve Fitting



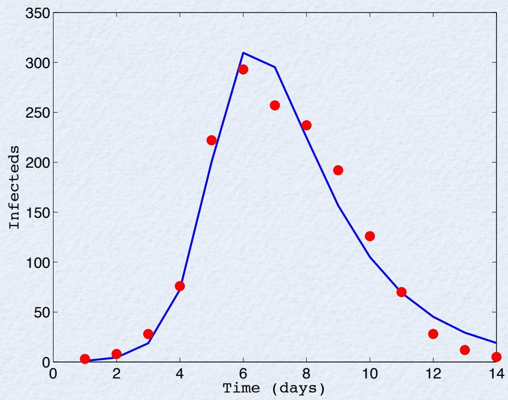

# Impact on $R_0$ Estimation: Influenza Outbreak in a School (1978)

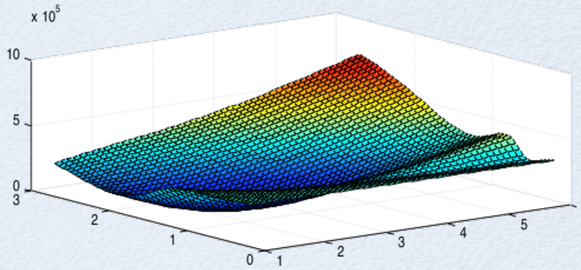



- Systematically vary β and γ, calculate SSE
- Parameter combination with lowest SSE is ‘best fit’

## Impact on $R_0$ estimation: Influenza outbreak in a school (1978)

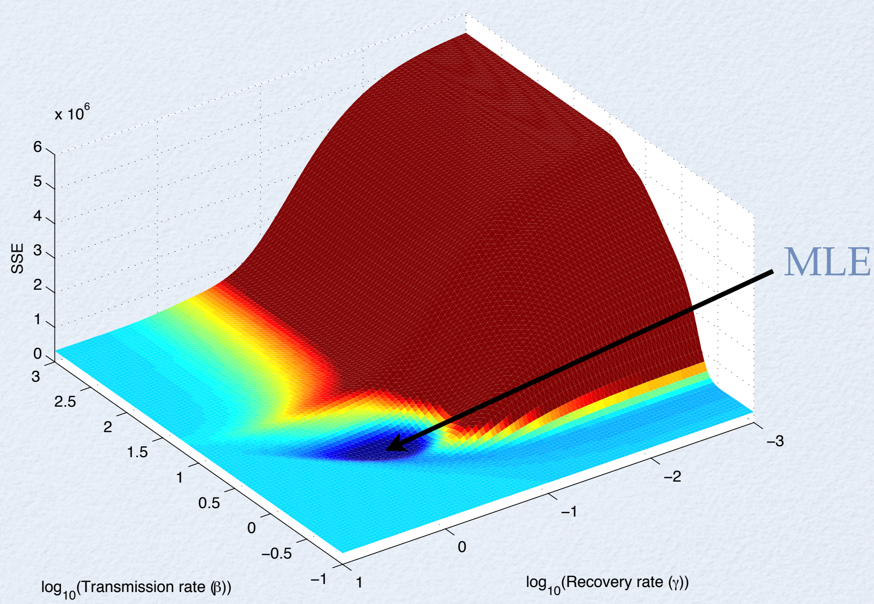



Best fit parameter values:

1. $\beta = 1.7$ (per day)
2. $1/\gamma = 2.2$ days
3. $R_0 \sim 3.74$

## Likelihood surface



# Caveat

- In boarding school example, data represent number of boys sick ~ $I(t)$
- Typically, data are 'incidence' (newly detected or reported infections)
- May not correspond to any model variables
- May need to 'construct' new information:
  - $dC/dt = \gamma I$ diagnosis at end of infectiousness
  - $dC/dt = \beta SI$
- Set $C(t+\Delta t) = 0$ where $\Delta t$ is sampling interval of data

# Lecture Summary ...

- $R_0$ can be estimated from epidemiological data in a variety of ways
  - Final epidemic size
  - Mean age at infection
  - Outbreak exponential growth rate
  - Curve Fitting
- In principle, variety of unknown parameters may be estimated from data
- Big issue we've ignored: reporting bias!