Model-Data Interface

Model Calibration I: *Ad hoc* estimation

Parameter estimation

- We've seen that basic reproductive ratio, R_0, is a very important quantity
- How do we calculate it?
- In general, we might not know (many) model parameters. How do we achieve parameter estimation from epidemiological data?
- Review some simple methods

The death of an epidemic

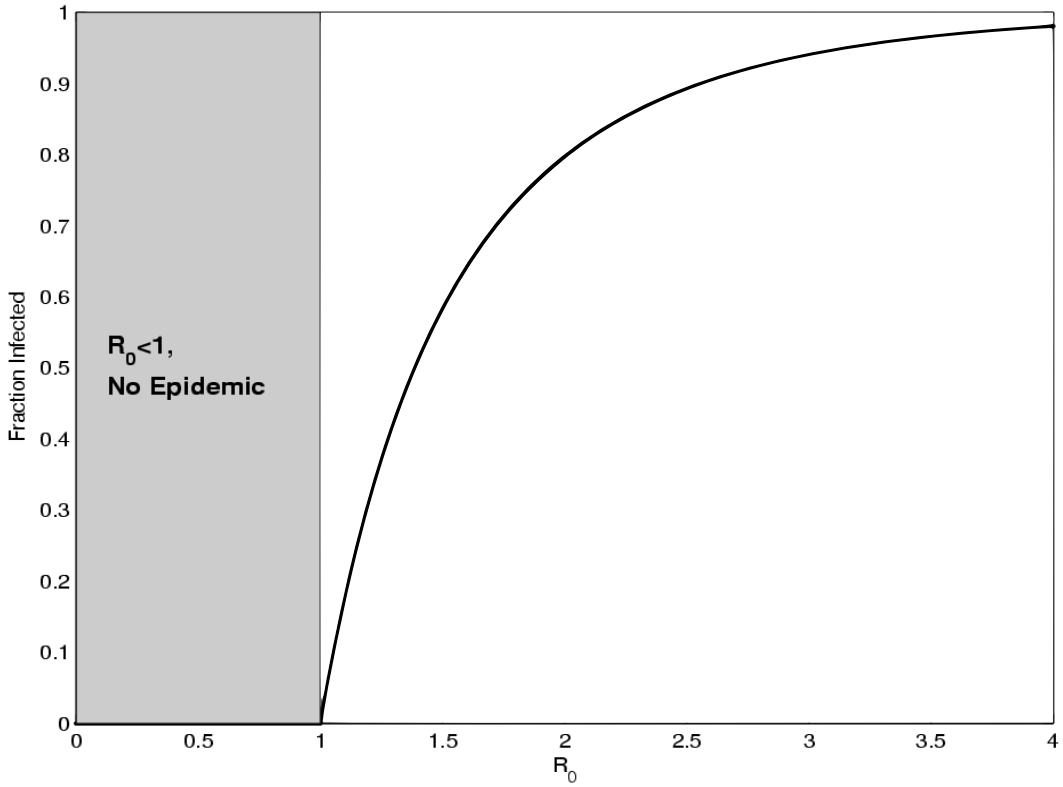
- In SIR equations, let's divide equation for dS/dt by dR/dt:
  - $dS/dR = - (\beta SI)/(\gamma I)$
  - $= - R_0 S$
- Integrate with respect to R
  - $S(t) = S(0)\, e^{-R(t)\, R_0}$
- When epidemic is over, by definition, we have $S(\infty)$, $I(\infty)$ $(=0)$, and $R(\infty)$
- $S(\infty) = 1 - R(\infty) = S(0)\, e^{-R(\infty)\, R_0}$

The death of an epidemic

Epidemic dies out because there are too few infectives, not because of too few susceptibles

Kermack & McKendrick (1927)

- So, $1 - R(\infty) - S(0)\, e^{-R(\infty) R_0} = 0$
- Solve this numerically ('transcendental' equation)



1a. Final outbreak size

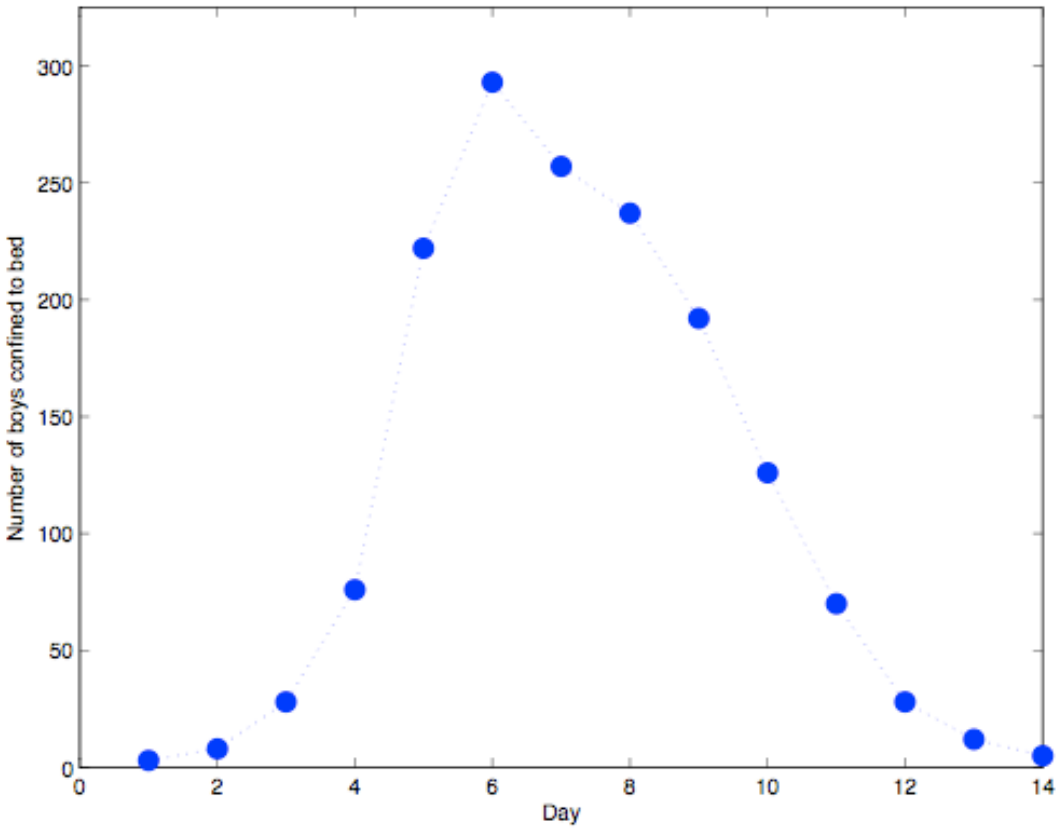
- So, if we know population size (N) , initial susceptibles (to get S(0)), and total number infected (to get R(∞)), we can calculate R_0

$$R_0 = -\frac{\log(1 - R(\infty))}{R(\infty)}$$

- 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(∞) ~ 512

R_0 ~ 1.65

1b. Final outbreak size

- Becker 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

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

2. Independent data

- 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 leads to

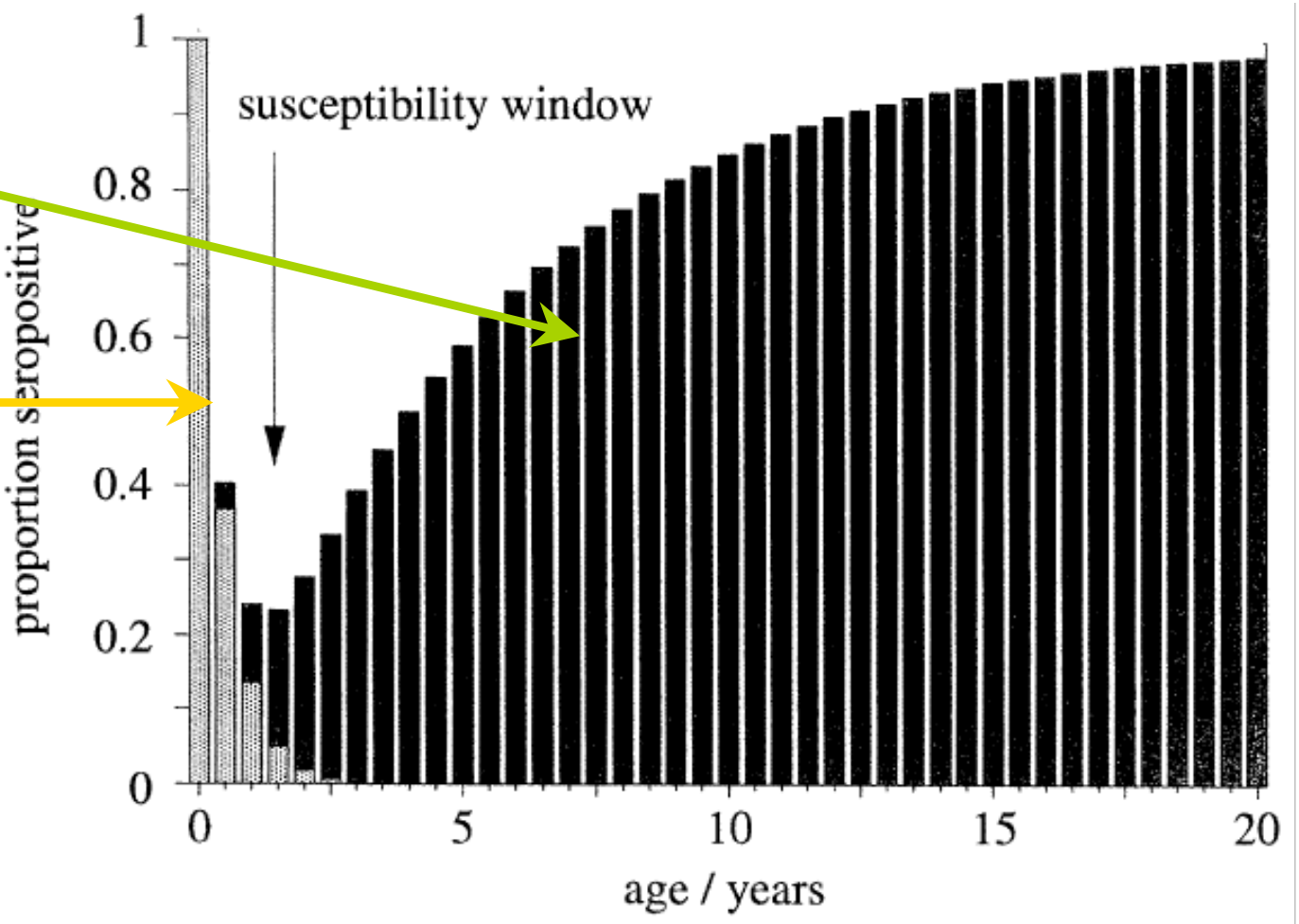
$$A = \left(\frac{1}{\mu(R_0 - 1)}\right) = \left(\frac{1}{\mu}\right)\left(\frac{1}{(R_0 - 1)}\right)$$

  - This can be written as $R_0 - 1 \approx L/A$ (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

Infection-derived immunity

Maternally-derived antibodies



Mean age at infection (A) is ~4.5 years

Assume L~75, so R_0 ~ 17.6

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

A slightly more common approach is to study the epidemic take off

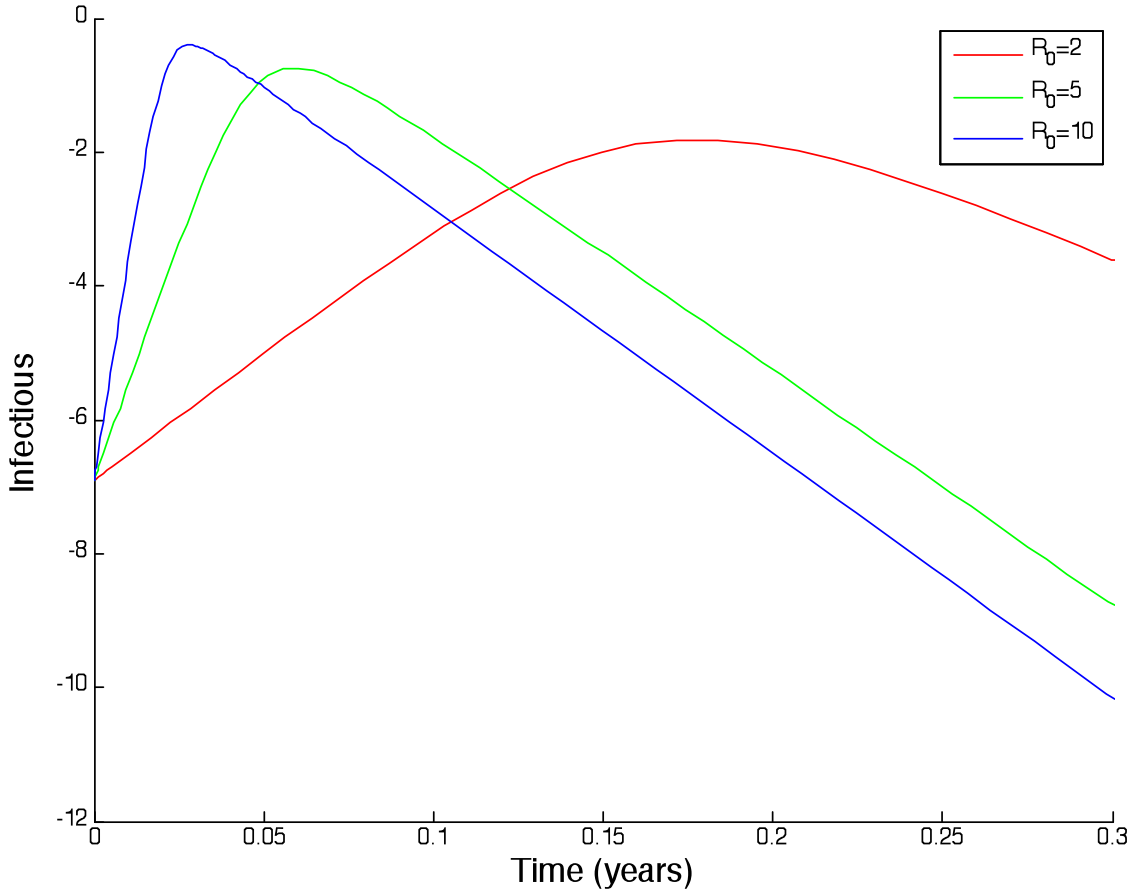
Recall from linear stability analysis that

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

Take logarithms

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

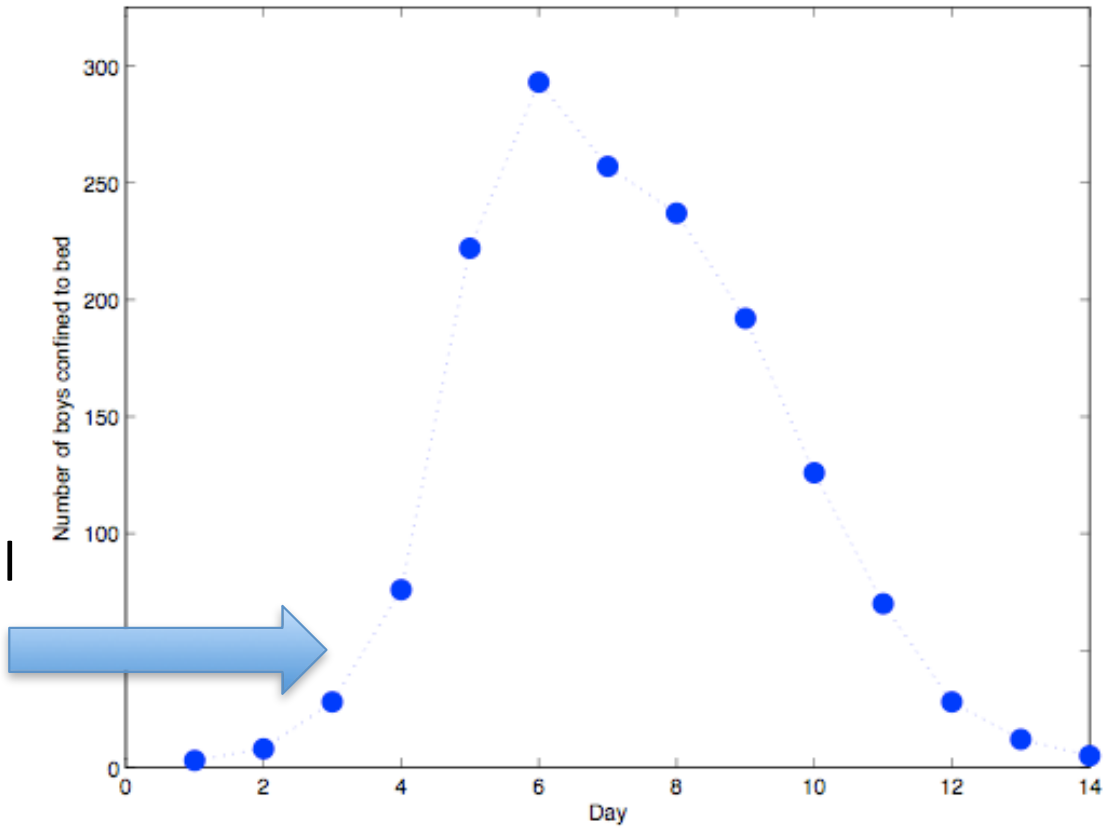
So, regression slope will give R_0



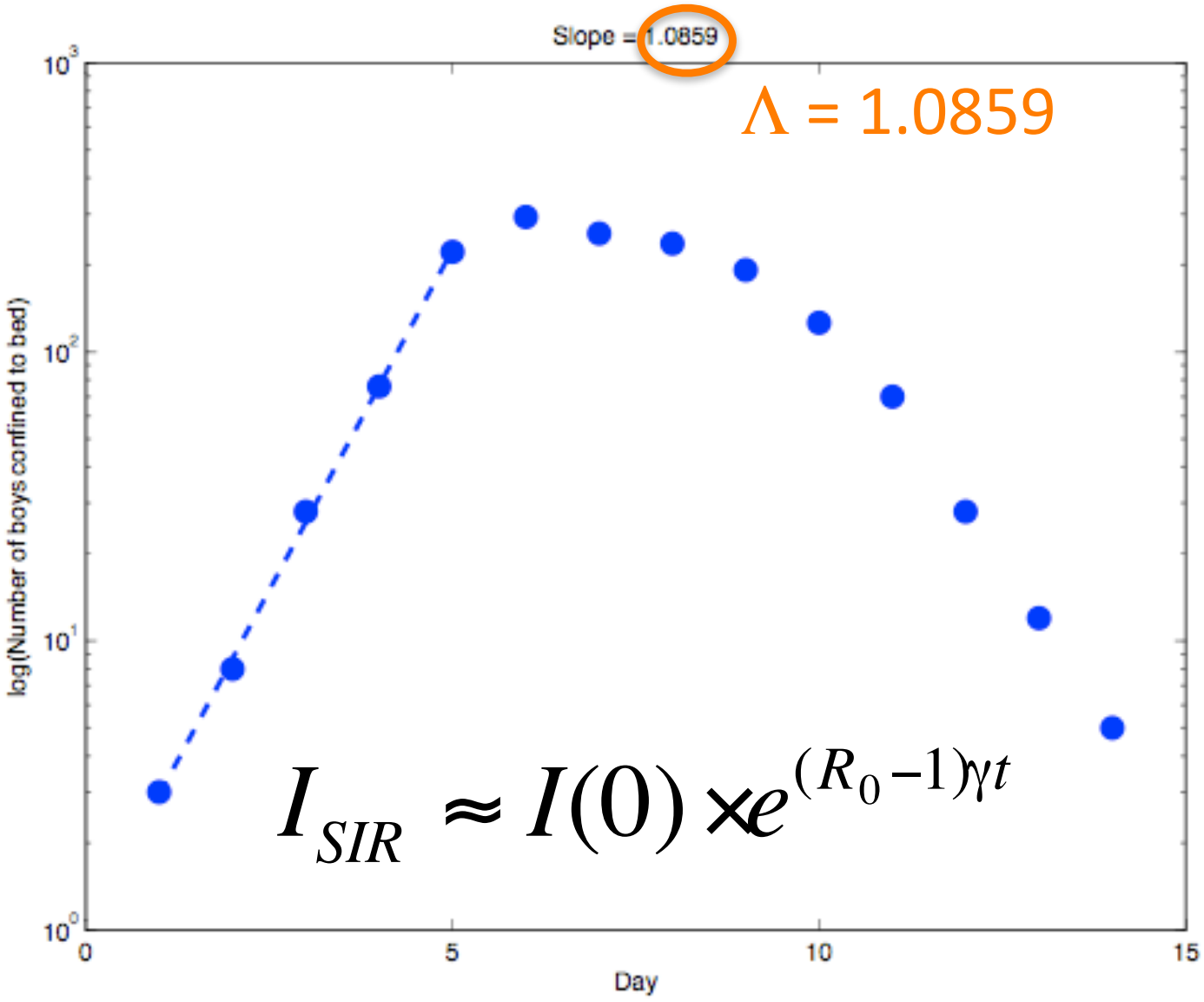
3. Epidemic take-off

- Back to school boys

Looks like classic exponential take-off



Epidemic take-off



$\Lambda = 1.0859$

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

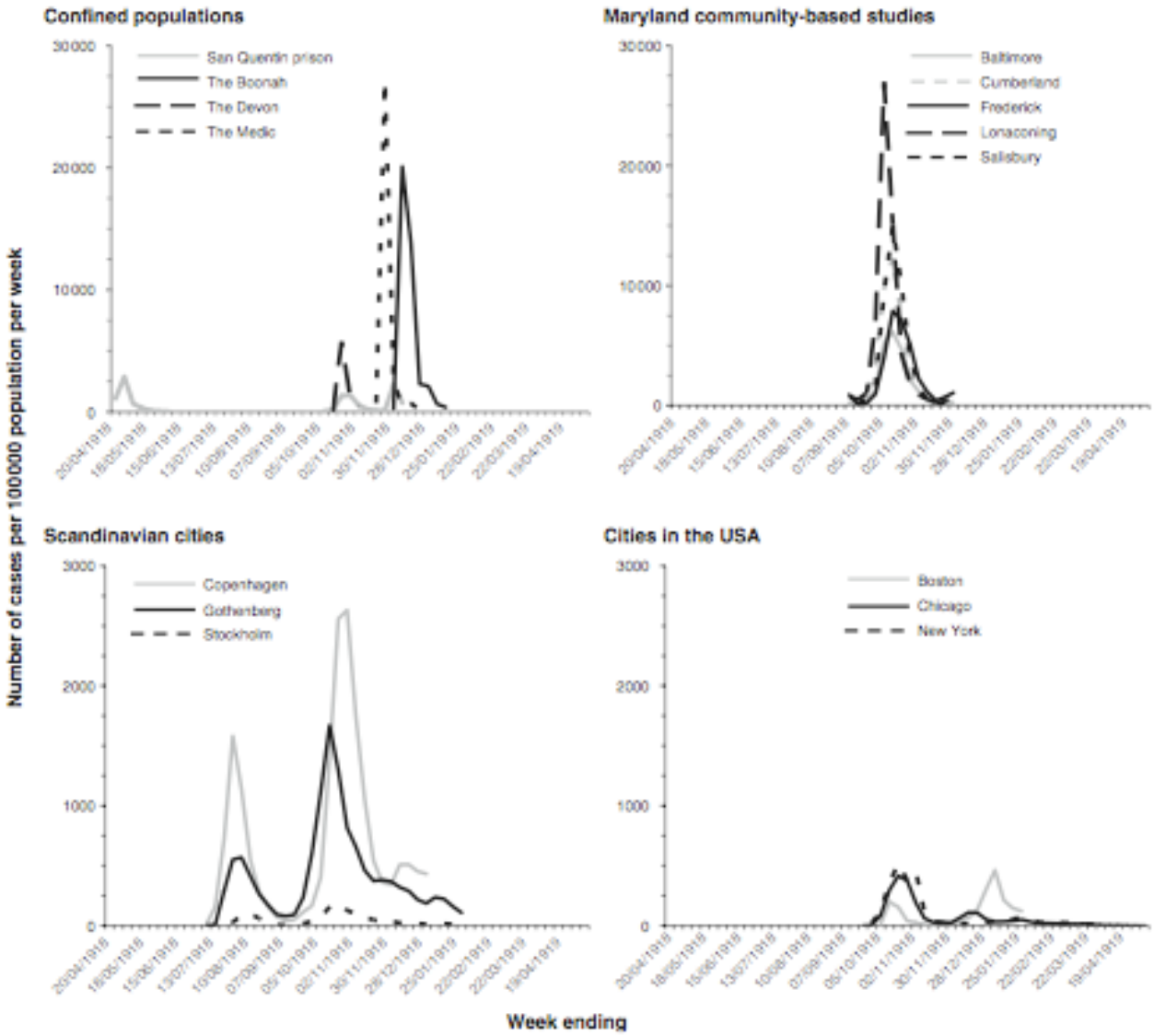
Our value for 'flu infectious period, $1/\gamma$

So,

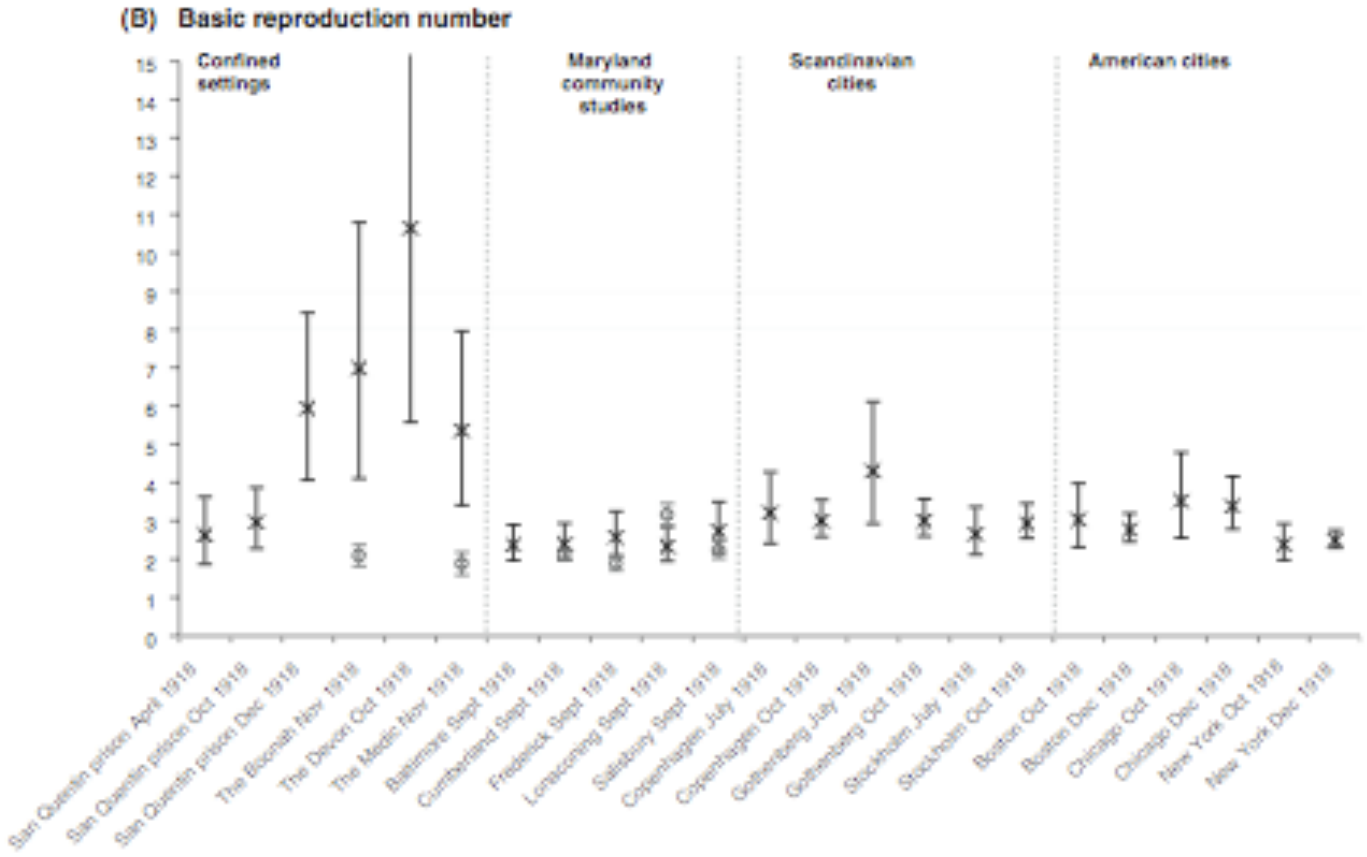
$R_0 = 1.0859 * 2.5 + 1$

$= 3.7$

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 & inference

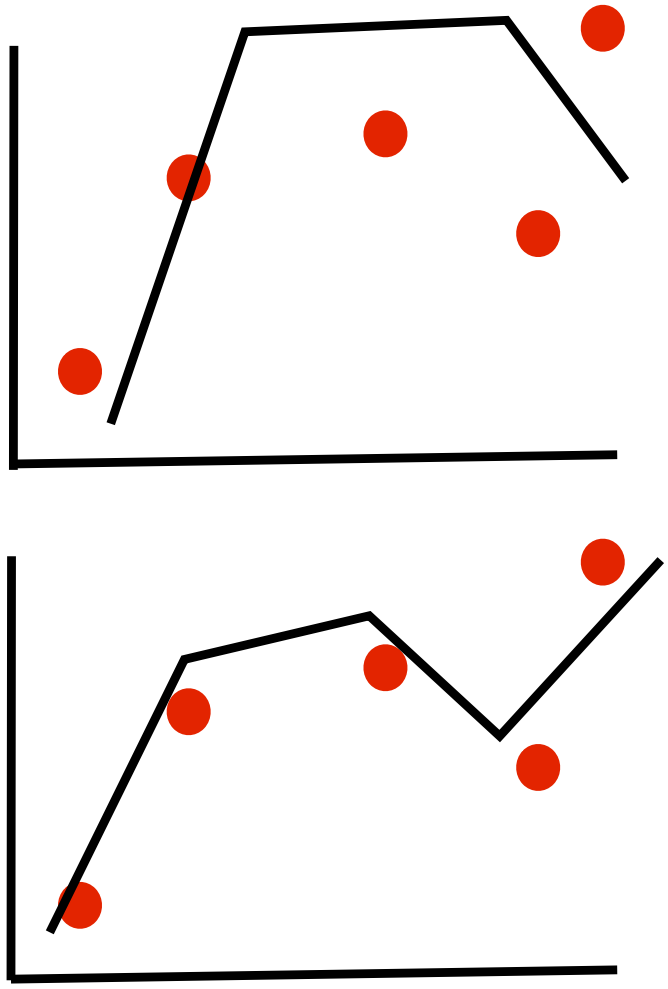
- We focus on random process that (putatively) generated data
- A model is explicit, mathematical description of this random process
- "The likelihood" is probability data were produced given model and its parameters:
- $$L(\text{model} \mid \text{data}) = Pr(\text{data} \mid \text{model})$$
- Likelihood quantifies (in some sense optimally)

4. Likelihood & estimation

- Assume we have data, D, and model output, M (both are vectors containing state variables). Model predictions generated using set of parameters, θ
- Observed dynamics subject to
  - "process noise": heterogeneity among individuals, random differences in timing of discrete events (environmental and demographic stochasticity)
  - "observation noise": random errors made in measurement process itself

4. Likelihood & estimation

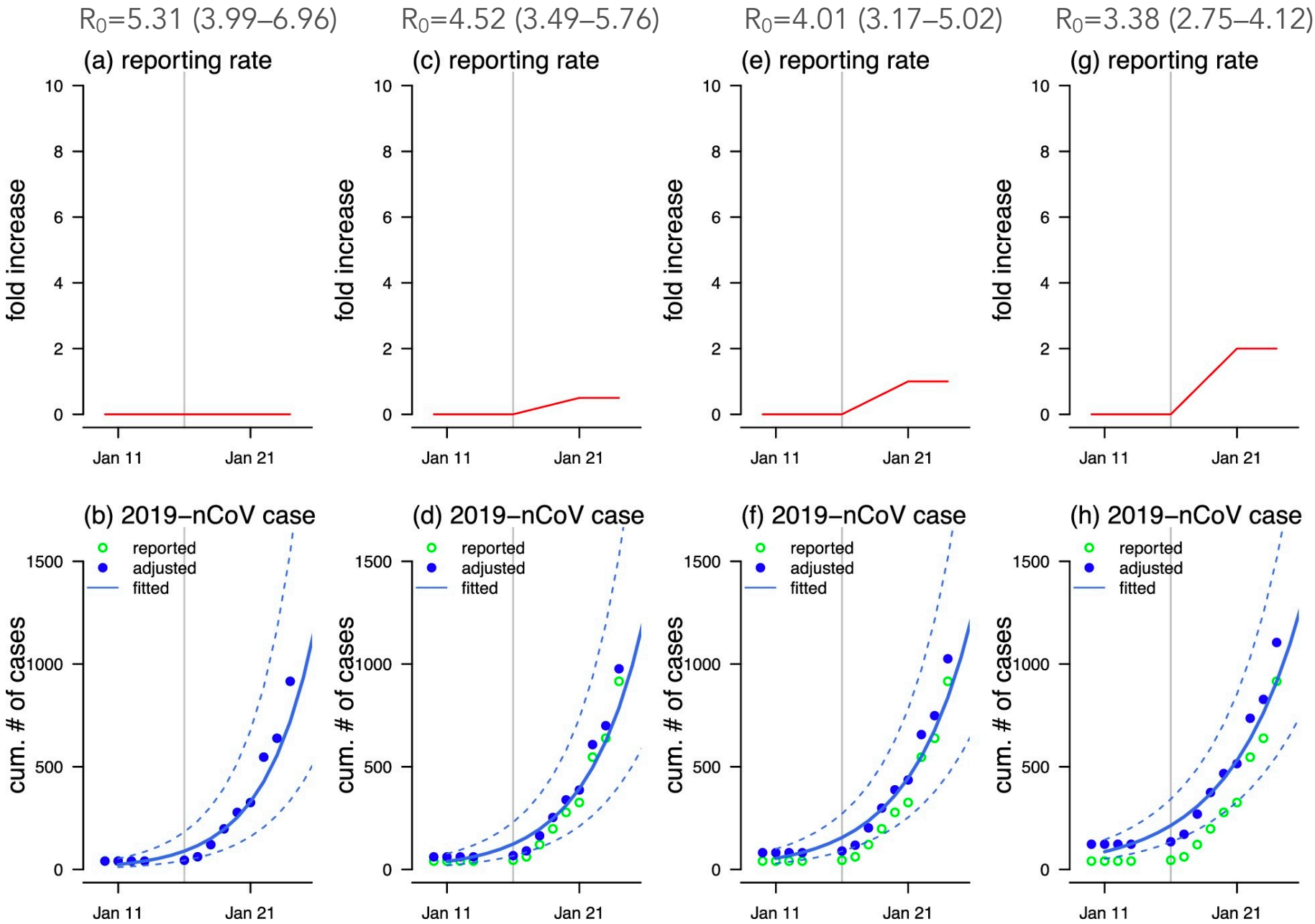
- If we ignore process noise, then model is deterministic and all variability attributed to measurement error
- Observation errors assumed to be sequentially independent
- Maximizing likelihood in this context is called ‘trajectory matching’



4. Likelihood & estimation

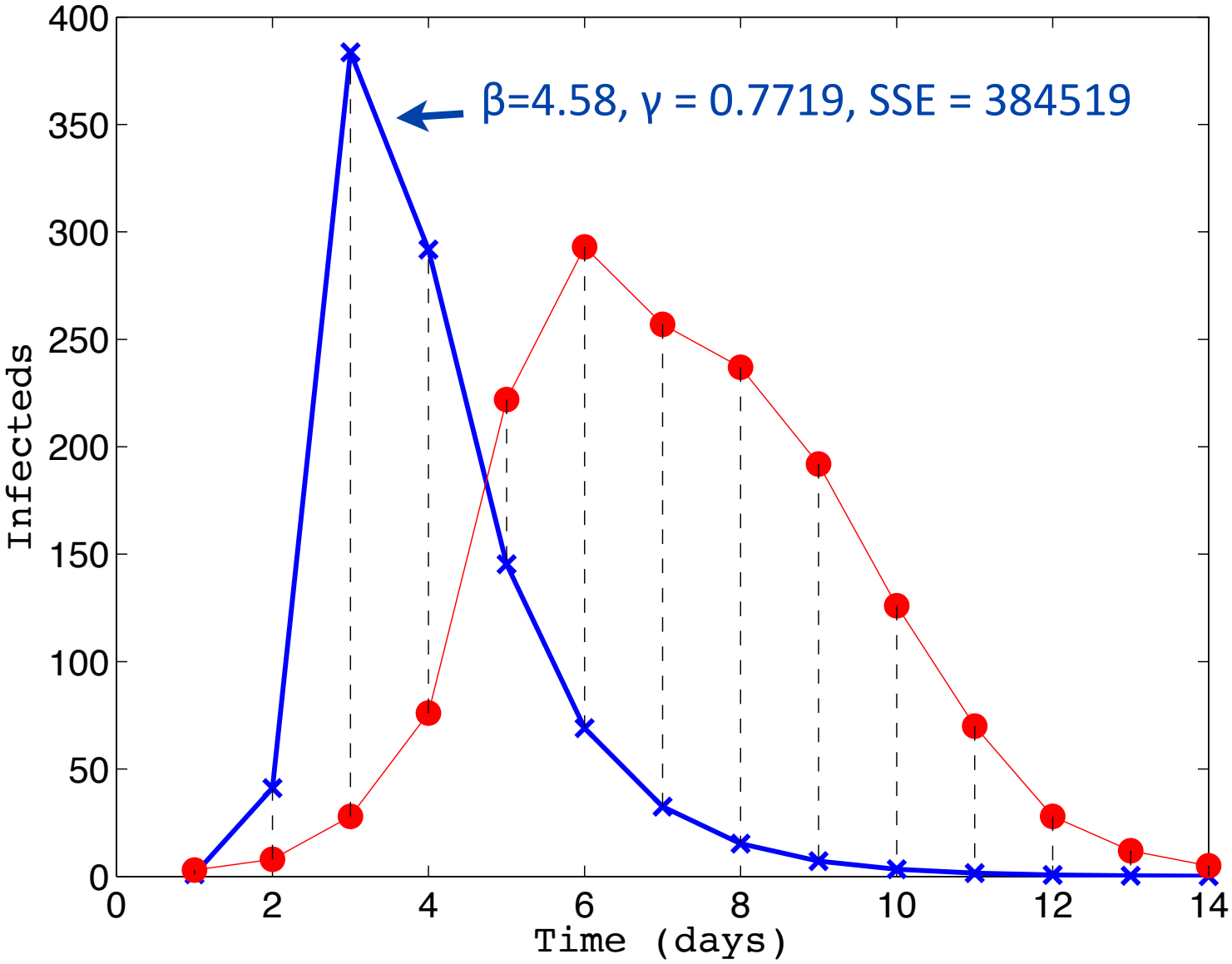
- Under such conditions, Maximum Likelihood Estimate, MLE, is simply parameter set with smallest deviation from data
- Equivalent to using least square errors, to decide on goodness of fit
  - Least Squares Statistic = SSE = $\Sigma(D_i - M_i)^2$
- Then, minimize SSE to arrive at MLE

COVID-19 fitting

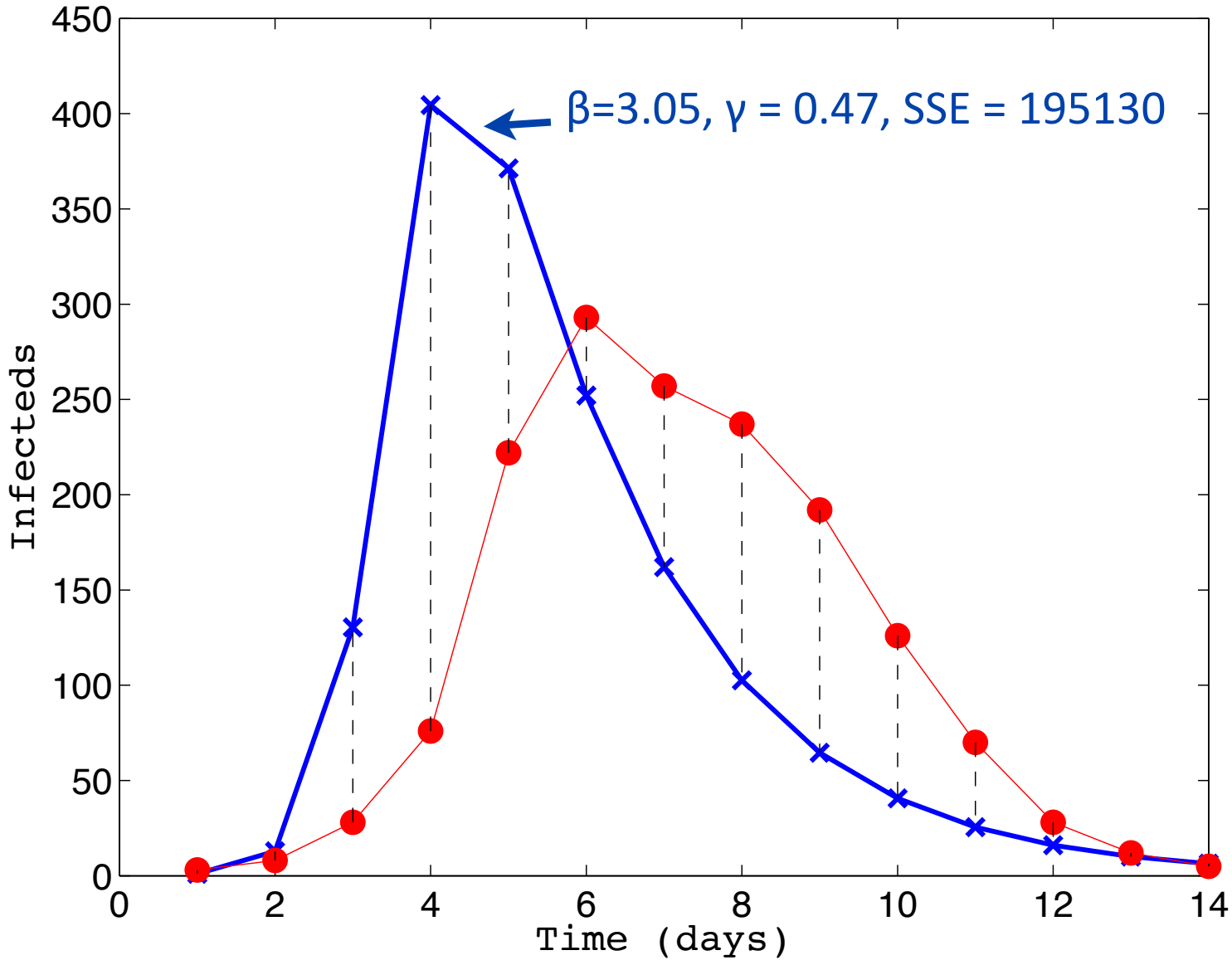


Zhao *et al*. (2020; *Int. J. Inf. Dis.*)

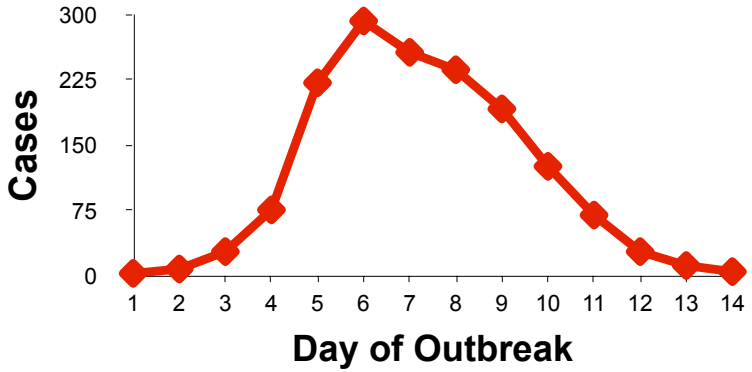
Trajectory matching

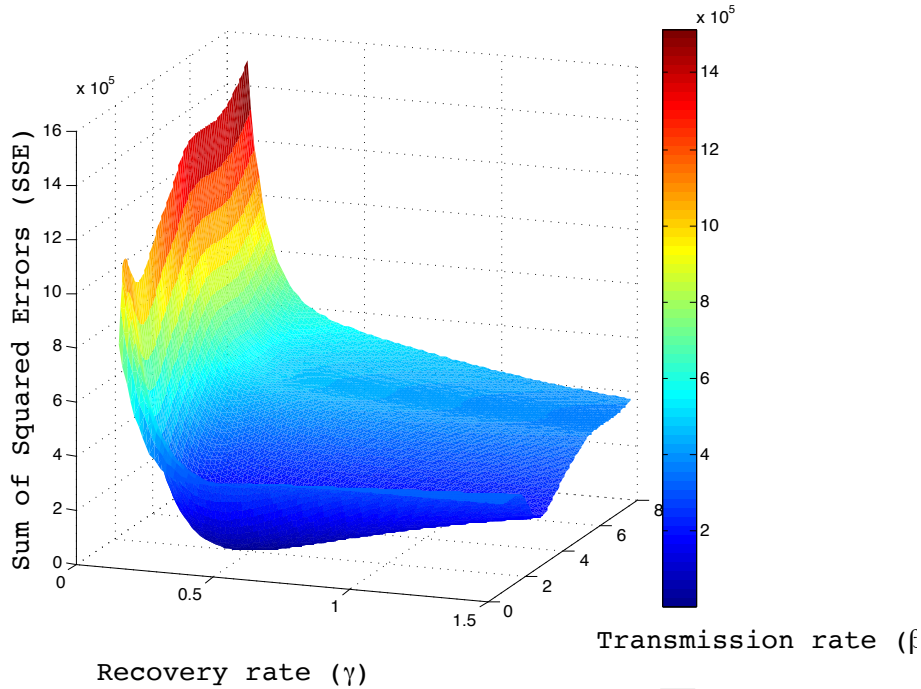


Trajectory matching

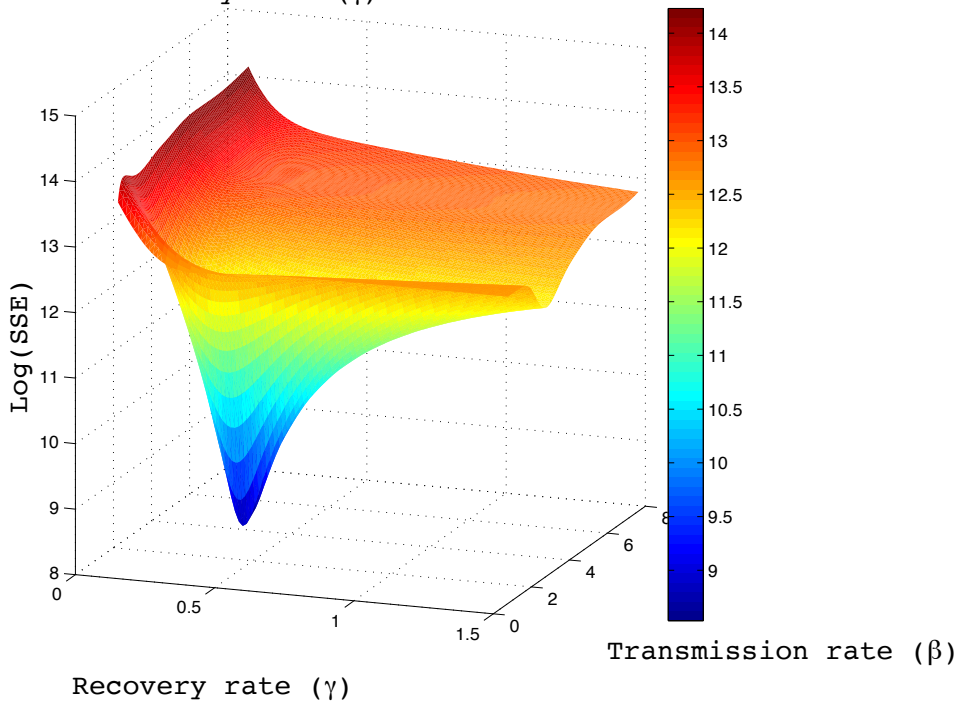


Model estimation: Influenza outbreak

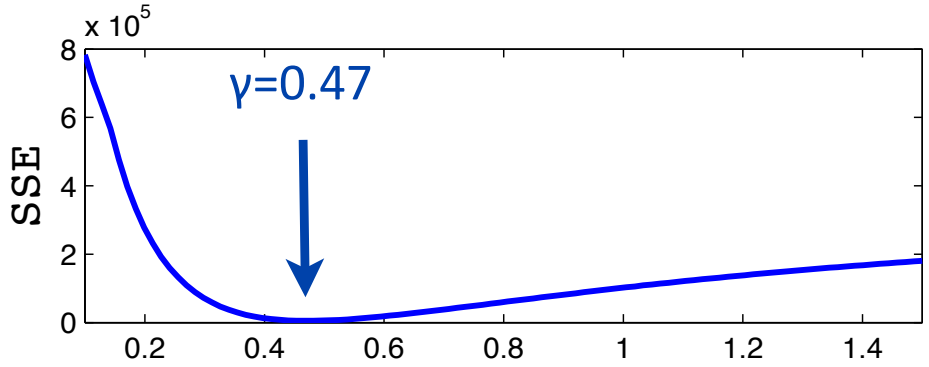




- Systematically vary β and γ, calculate SSE
- Parameter combination with lowest SSE is 'best fit'



Model estimation: Influenza outbreak

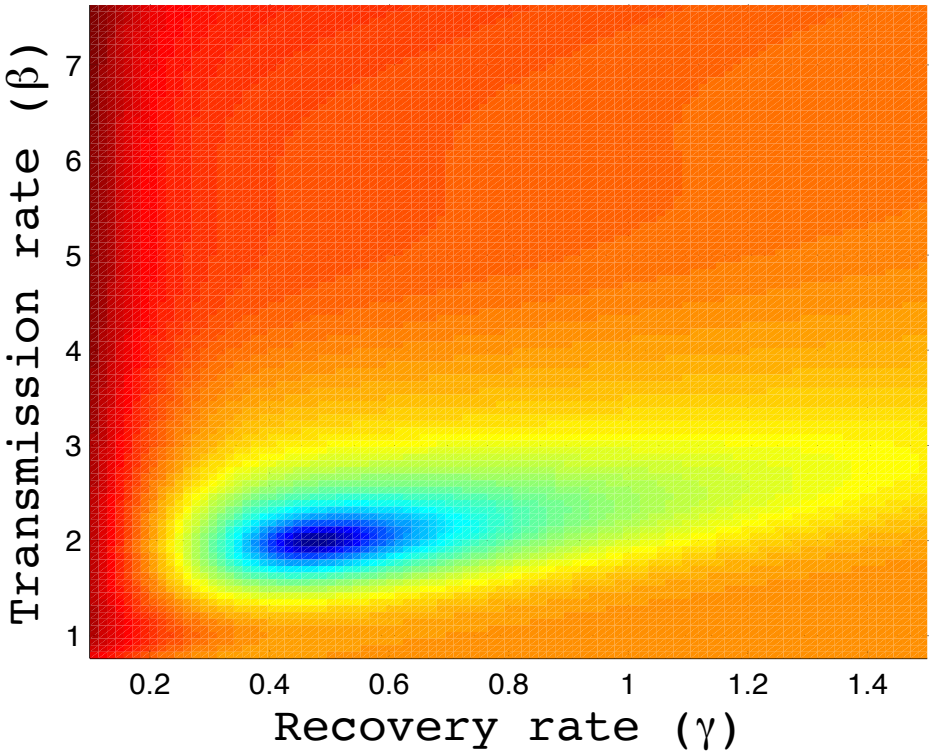


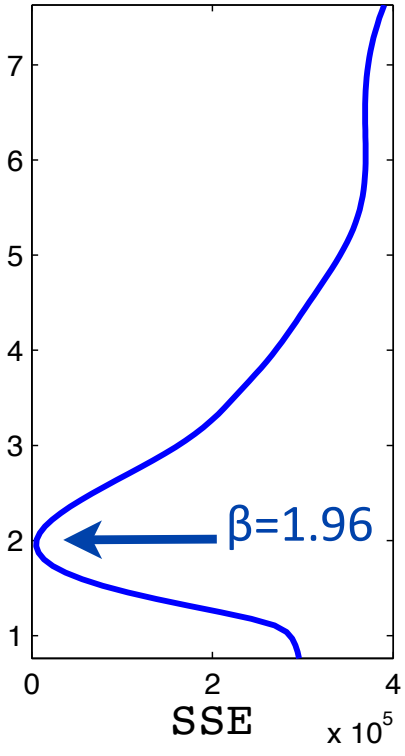
Best fit parameter values:

β = 1.96 (per day)

$1/\gamma$ = 2.1 days

$R_0 \sim 4.15$

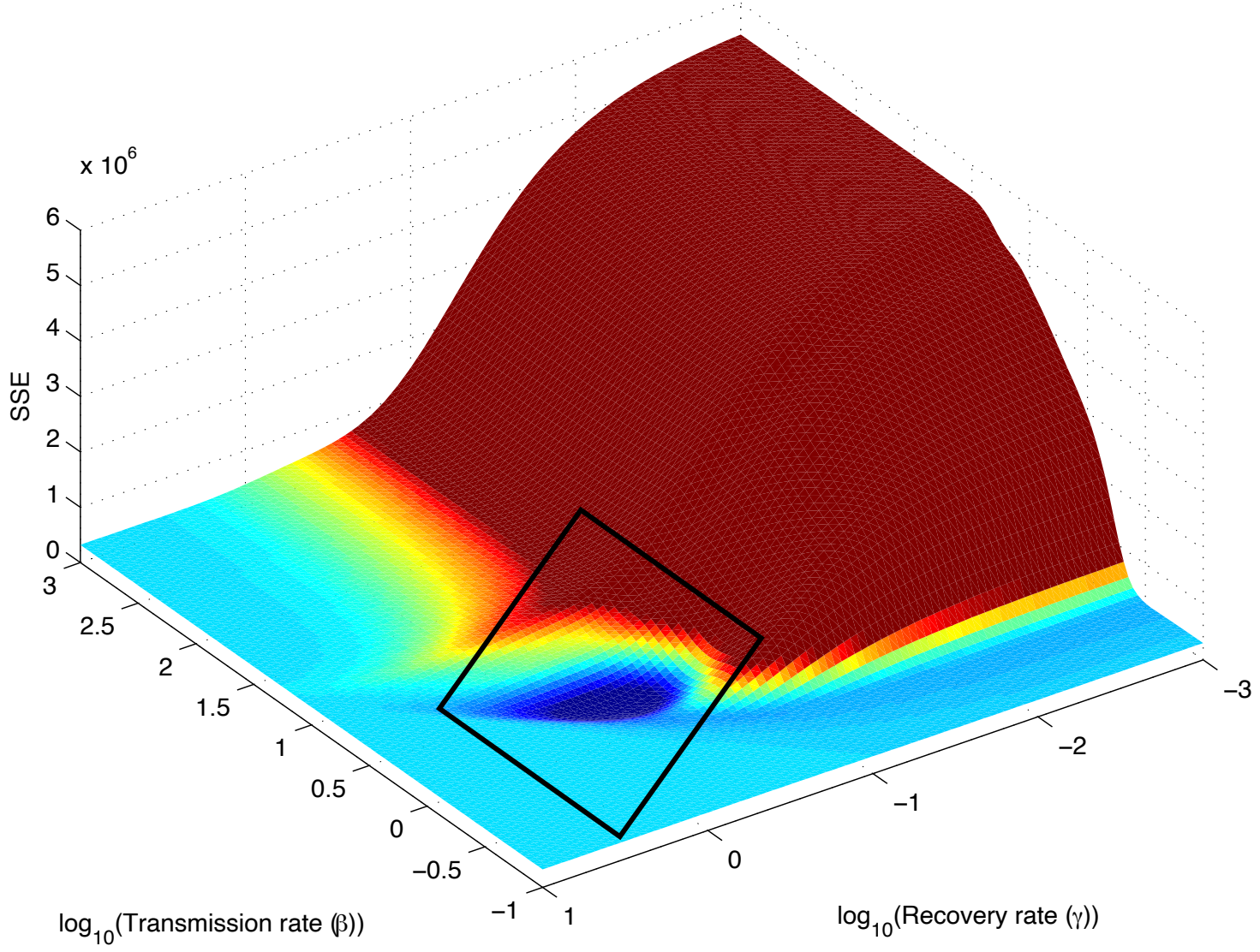




Generally, may have more parameters to fit, so grid search not efficient

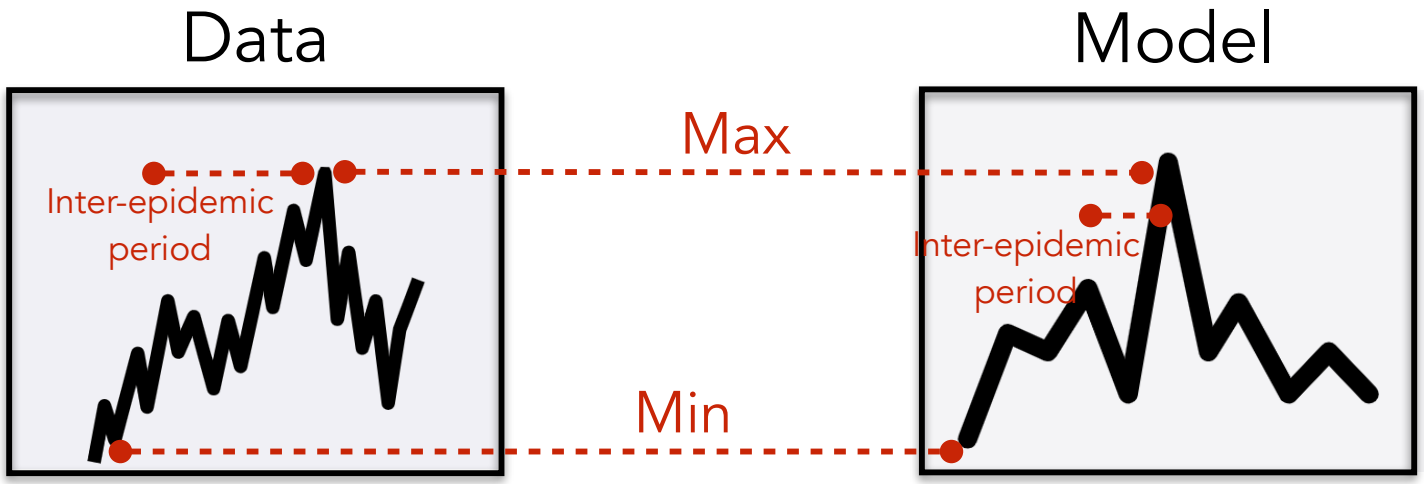
Nonlinear optimization algorithms (eg Nelder-Mead) would be used

Likelihood surface



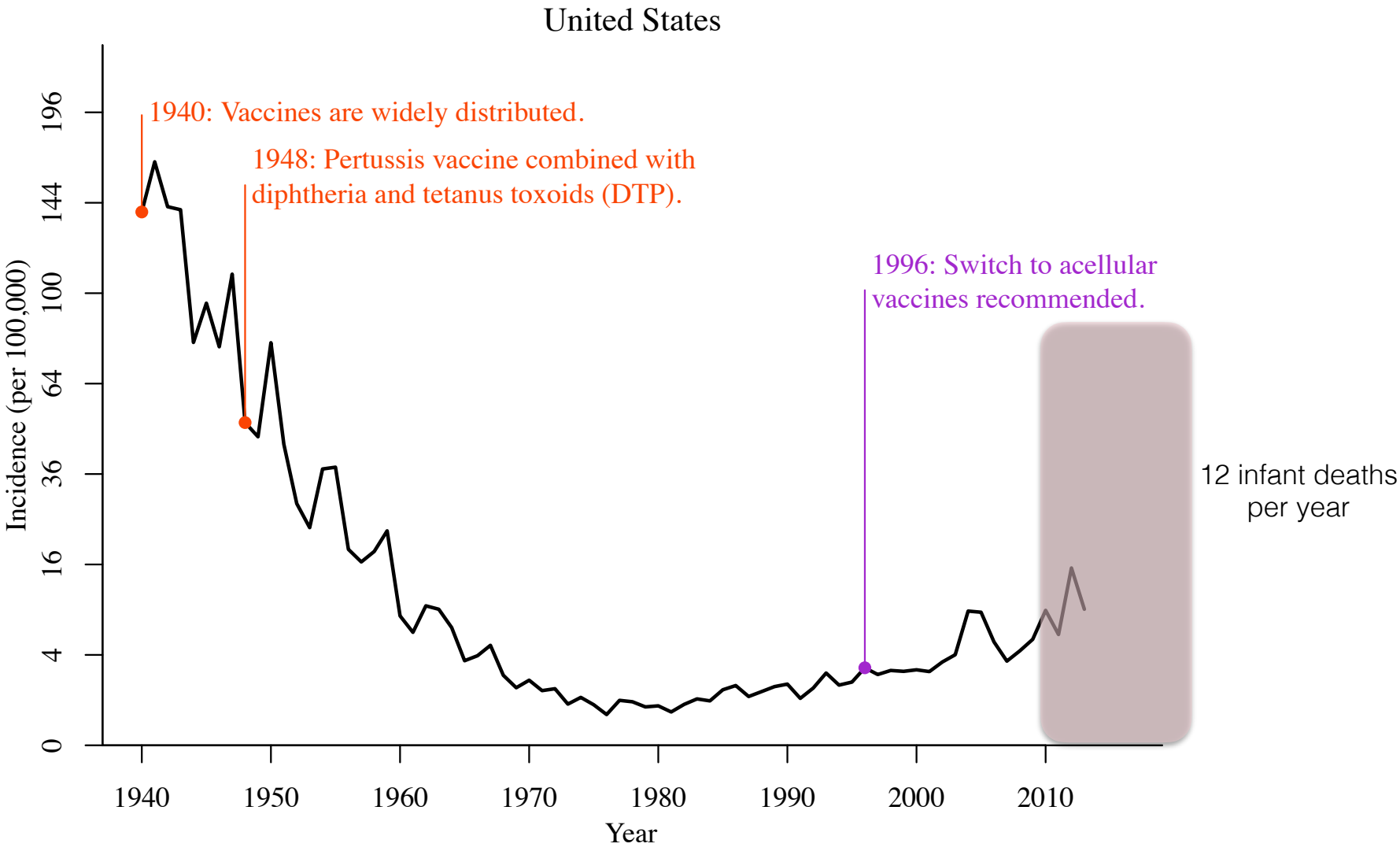
When likelihood surface is somewhat complex, success of estimation using gradient-based optimization algorithms (eg Nelder-Mead) will depend on providing a good initial guess

Other approaches



Compare model output with data, based on statistical "features" (or "probes") rather than raw numbers

Historical patterns of pertussis



PUZZLE: Some countries with high vaccine coverage have experienced resurgence

CIDRAP Center for Infectious Disease Research and Policy

News & Perspective | Infectious Disease Topics | Antimicrobial Stewardship | Ongoing Programs

FEATURED NEWS TOPICS COVID-19 Flu Vaccines Roadmap MERS-CoV Chronic Wasting Disease

Whooping cough cases tied to waning vaccine protection

Filed Under: Pertussis; Childhood Vaccines

Stephanie Soucheray | News Reporter | CIDRAP News | Jun 10, 2019

Share Tweet LinkedIn Email Print & PDF

A child who has never been vaccinated against pertussis, or whooping cough, is 13 times more likely to suffer from an infection of *Bordetella pertussis* than is a child who is up-to-date on his or her vaccines.

But new evidence from a decade-long study at Kaiser Permanente shows that vaccinated children were five times more likely to suffer from whooping cough if it had been more than 3 years since their last vaccine dose. The research was published today in *Pediatrics*.

Ran Kyu Park / iStock

s J. 2005 May;24(5 Suppl):S58-61.

.59160.41.

unity against pertussis after natural ination

s Van Rie, Stefania Salmaso, Janet A Englund

/01.inf.0000160914.59160.41

ation coverage, pertussis has remained endemic and reemerged as a public health problem in many countries in the past 2 decades. Waning of vaccine-induced nd. A review of red immunity

Q: Do pertussis vaccines protect for a lifetime?

A: Pertussis vaccines are effective, but not perfect. within the first 2 years after getting the vaccine, bu health experts call this 'waning immunity.' Similarly a few years.

In general, DTaP vaccines are 80% to 90% effective schedule, effectiveness is very high within the year are fully protected. There is a modest decrease in of 10 kids are fully protected 5 years after getting t kids are partially protected – protecting against se

Journal List › Cold Spring Harb Perspect Biol › v.9(12); 2017 Dec › PMC5710106

CSH PERSPECTIVES Cold Spring Harbor Perspectives in Biology

Cold Spring Harb Perspect Biol. 2017 Dec; 9(12): a029454.

doi: 10.1101/cshperspect.a029454

What Is Wrong with Pertussis Vaccine Immunity?

The Problem of Waning Effectiveness of Pertussis Vaccines

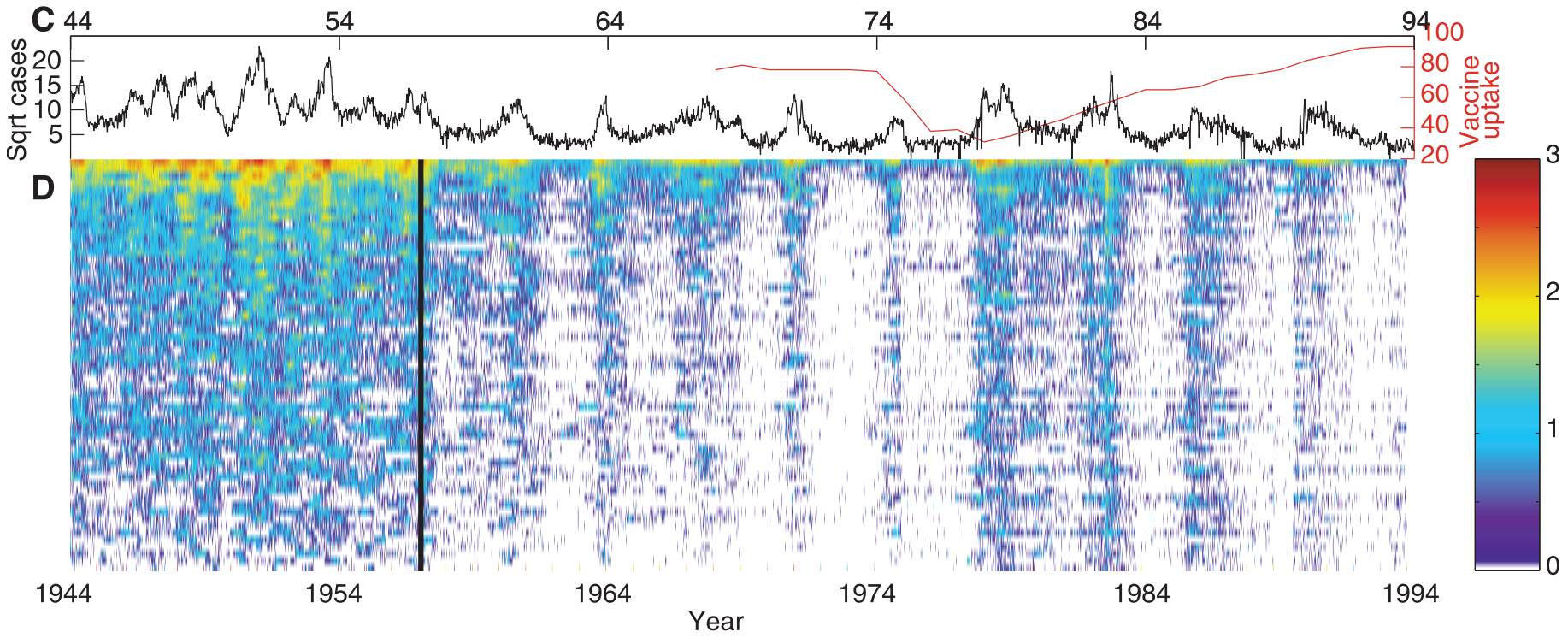
Nicolas Burdin,[1] Lori Kestenbaum Handy,[2] and Stanley A. Plotkin[3]

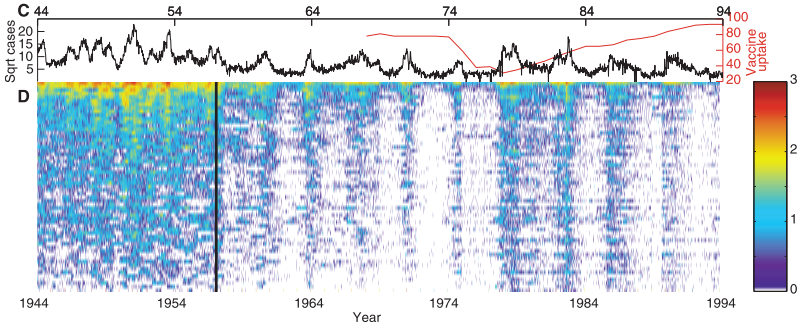
Author information › Copyright and License information Disclaimer

Cold Spring Har

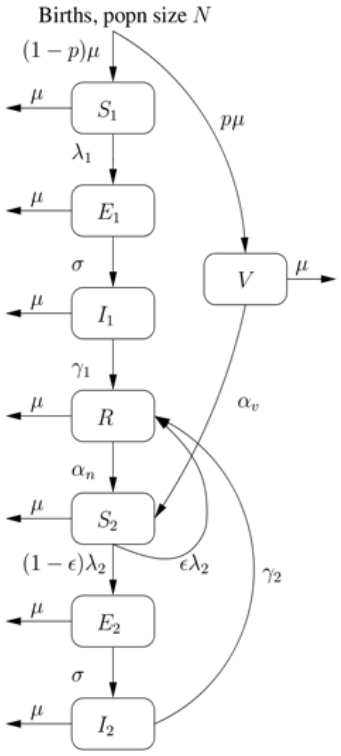
Pertussis!

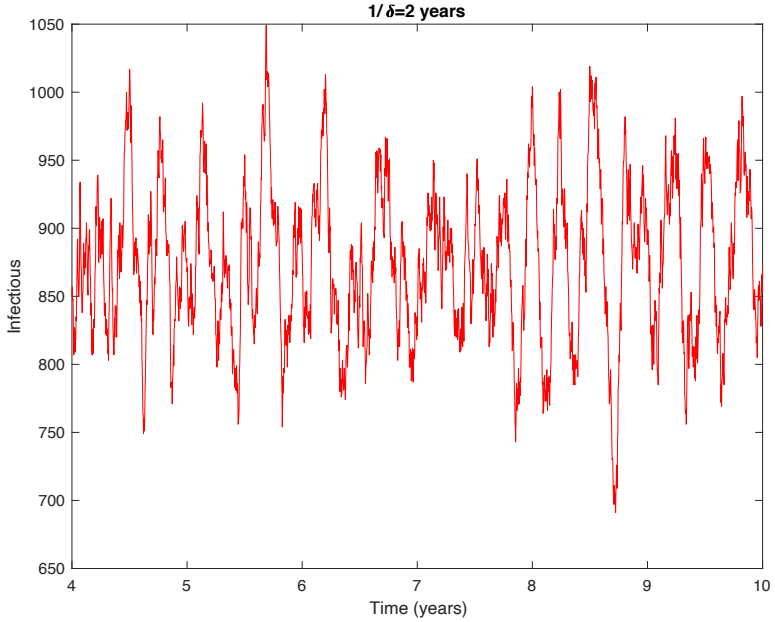
- Problem: No identified serological marker for protection
- How to infer protective duration following infection or vaccination?

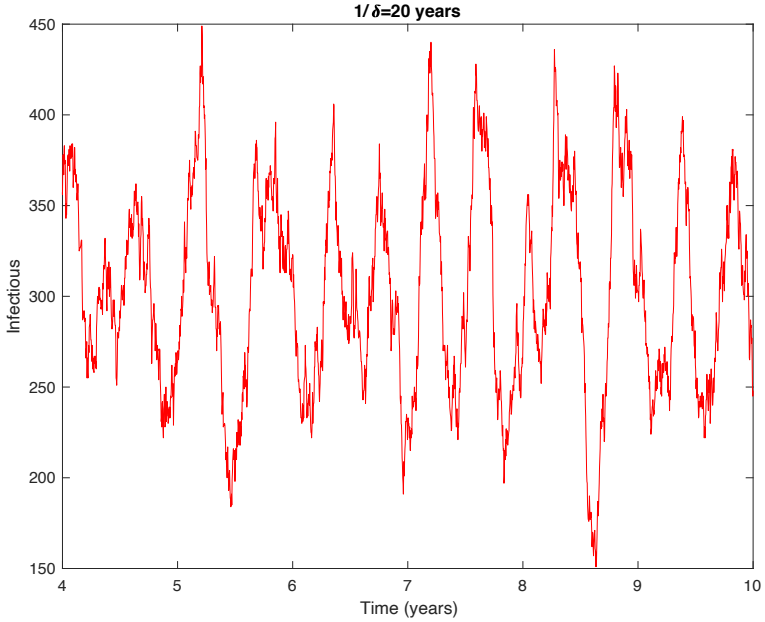




- Simulate model using Gillespie's Direct Method
- Include birth rates and population size drawn from England & Wales
- Quantitatively contrast model output with data
  Inter-epidemic period
  Fade-out (extinction) frequency

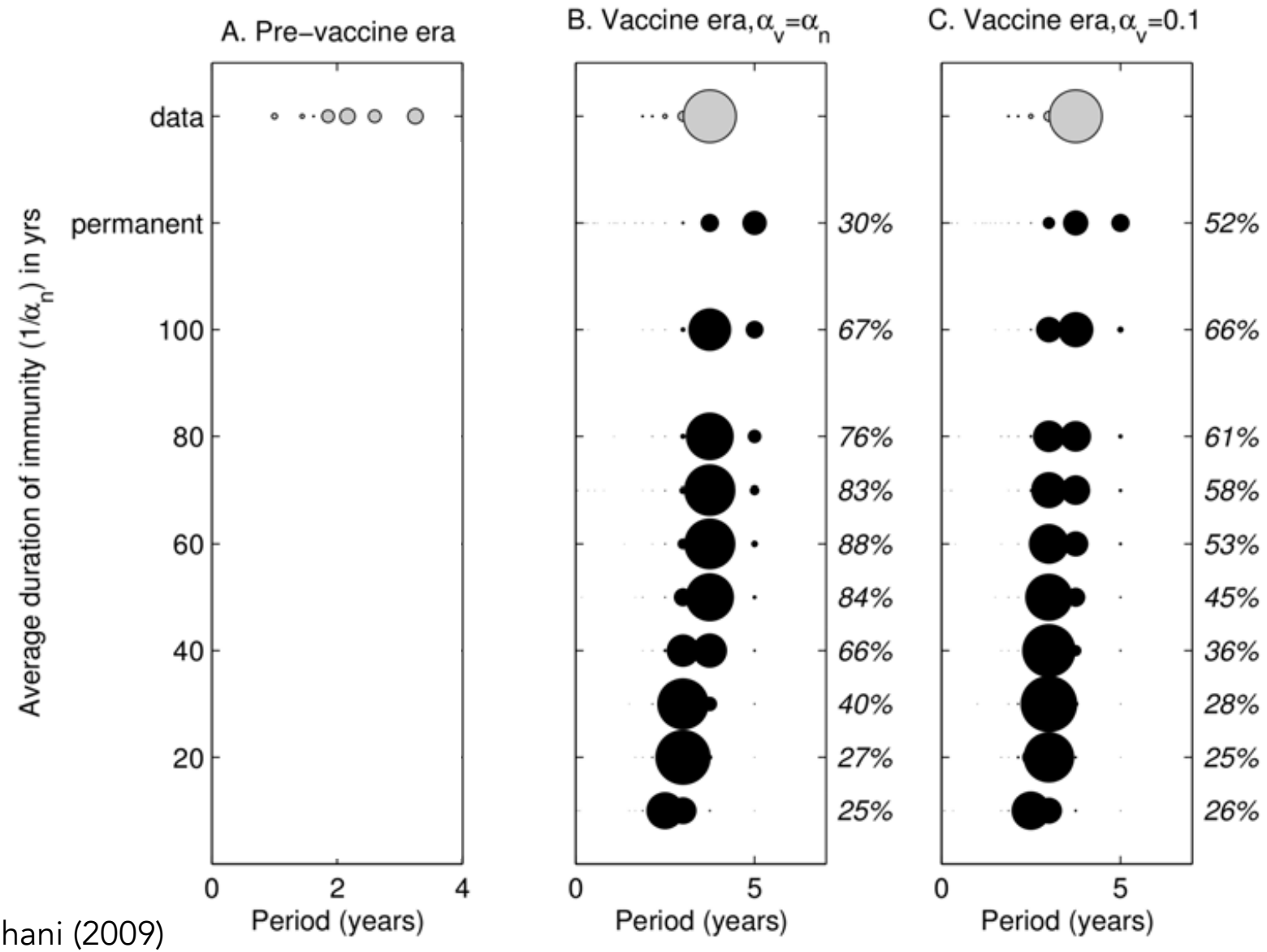






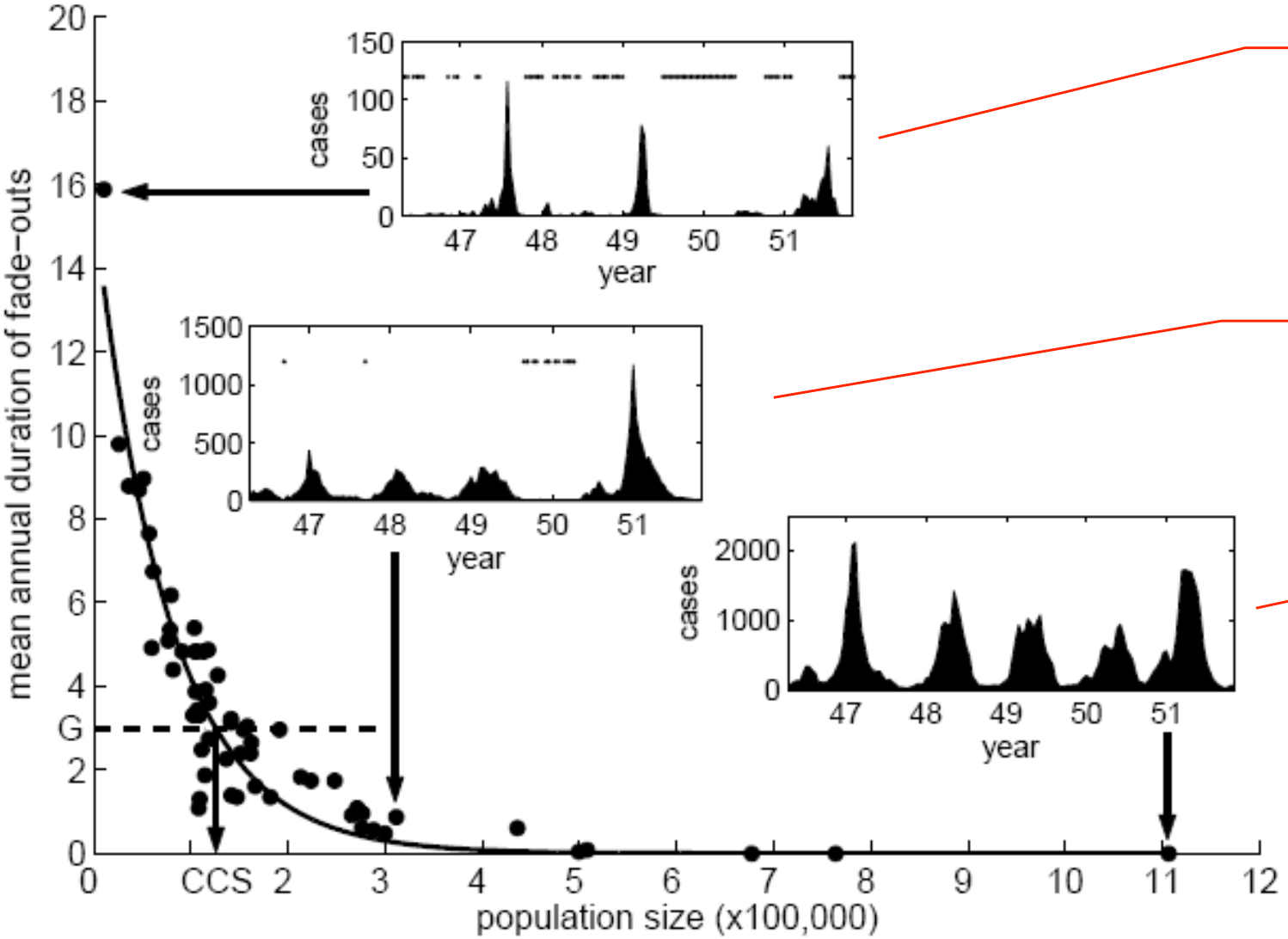
Wearing & Rohani (2009)

Inter-epidemic period



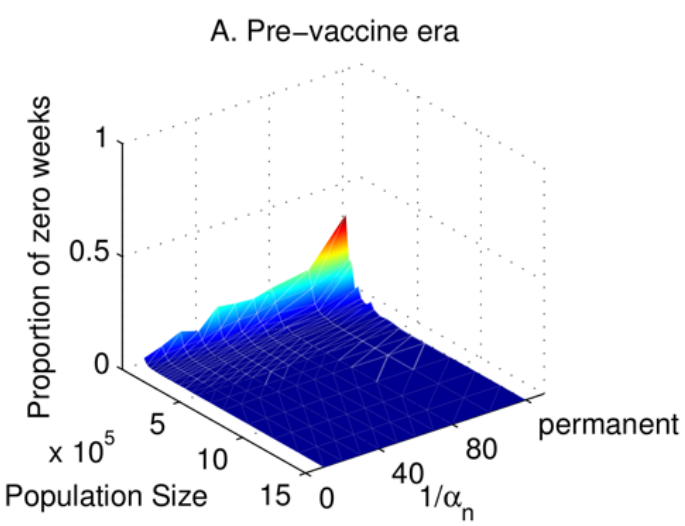
Wearing & Rohani (2009)

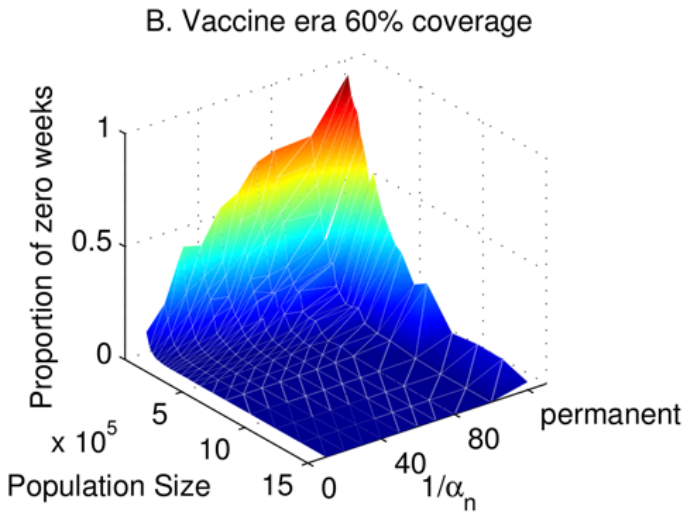
Different kinds of epidemics

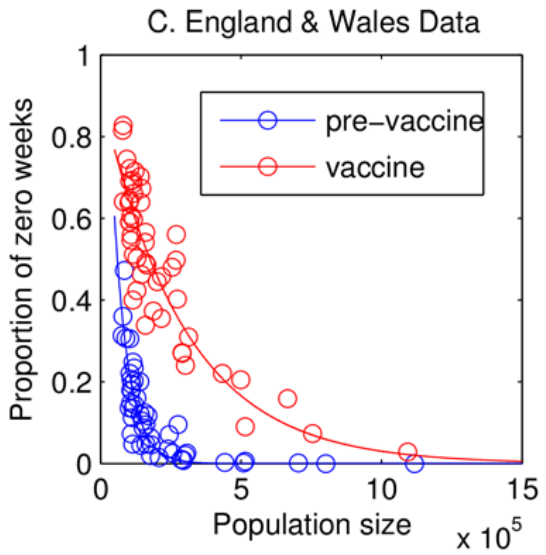


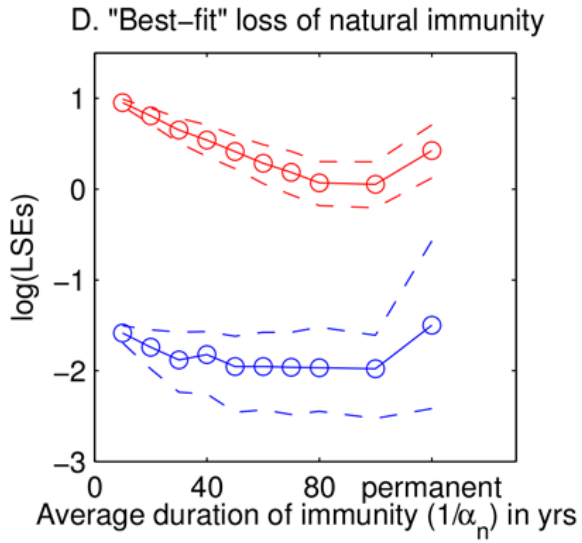
Bartlett (1957)

Fade-out frequency









Synthetic "likelihood"

- Idea formalized by Wood (2010)
- Take data ***y*** and convert to statistics ***s*** (*eg* coefficients of autocovariance function, mean incidence of # zeros)
- Choice of ***s*** allows us to define <u>what</u> matters about dynamics, but not <u>how much</u> it matters
- Use model to simulate N_r replicate data sets (y_1^*, y_2^*, …)
- Convert to replicate statistics vectors (s_1^*, s_2^*, …) exactly as *y* was converted to *s*
- Evaluate

$$\hat{\mu}_\theta = \sum_i \frac{s_i^*}{N_r} \quad \text{And} \quad S = (s_1^* - \hat{\mu}_\theta, s_2^* - \hat{\mu}_\theta, \ldots)$$

Synthetic "likelihood"

$$\hat{\mu}_\theta = \sum_i \frac{s_i^*}{N_r} \quad \text{And} \quad S = (s_1^* - \hat{\mu}_\theta, s_2^* - \hat{\mu}_\theta, \ldots)$$

- So $$\hat{\Sigma}_\theta = SS^T/(N_r - 1)$$

- The synthetic likelihood is therefore

$$l_s(\theta) = \frac{1}{2}(s - \hat{\mu}_\theta)^T \hat{\Sigma}_\theta^{-1}(s - \hat{\mu}_\theta) - \frac{1}{2}\log|\hat{\Sigma}_\theta|$$

- Measures consistency of parameter values with observed data

Caveat

- In boarding school example, data represent number of boys sick ~ C(t)
- Typically, data are 'incidence' (newly detected or reported infections)
- Don't directly correspond to any model variables
- May need to 'construct' new information:
  - $dP/dt = \gamma Y$ diagnosis at end of infectiousness
  - $dP/dt = \beta XY/N$
- Set $P(t+\Delta t) = 0$ where Δt 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

Further, ...

1. Include uncertainty in initial conditions
   - We took I(0) = 1. Instead could estimate I(0) together with β and γ (now have 1 fewer data points)
2. Explicit observation model
   - Implicitly assumed measurement errors normally distributed with fixed variance, but can relax this assumption
   - Sometimes, better to use log-normal distribution
3. What is appropriate model?
   - SEIR model? (latent period before becoming infectious)
   - SEICR model? ("confinement to bed")
   - Time varying parameters? (e.g. action taken to control spread)

Further, ...

4. Assumed model deterministic -- how do we fit a stochastic model?
   - Use a 'particle filter' to calculate likelihood
5. Can we simultaneously estimate numerous parameters?
   - More complex models have more parameters… estimate all from 14 data points? ⇒ identifiability
6. More complex models are more flexible, so tend to fit better
   - How do we determine if increased fit justifies increased complexity? ⇒ information criteria