McMaster University

Mathematics and Statistics

$$\int_M d\omega = \int_{\partial M} \omega$$

Mathematics 747 / 5GT3
Topics in Mathematical Biology

Instructor: David Earn

Lecture 1
Epidemic Modelling Intro
Thursday 17 September 2020

Course information

- The course web site:
  http://davidearn.github.io/tmb2020

- Office hours are by appointment (online only):
  E-mail earn@math.mcmaster.ca

Software

- **ASAP**, install the software discussed on the Software page on the course web site:
  - LaTeX

  - R
  - RStudio
  - Emacs

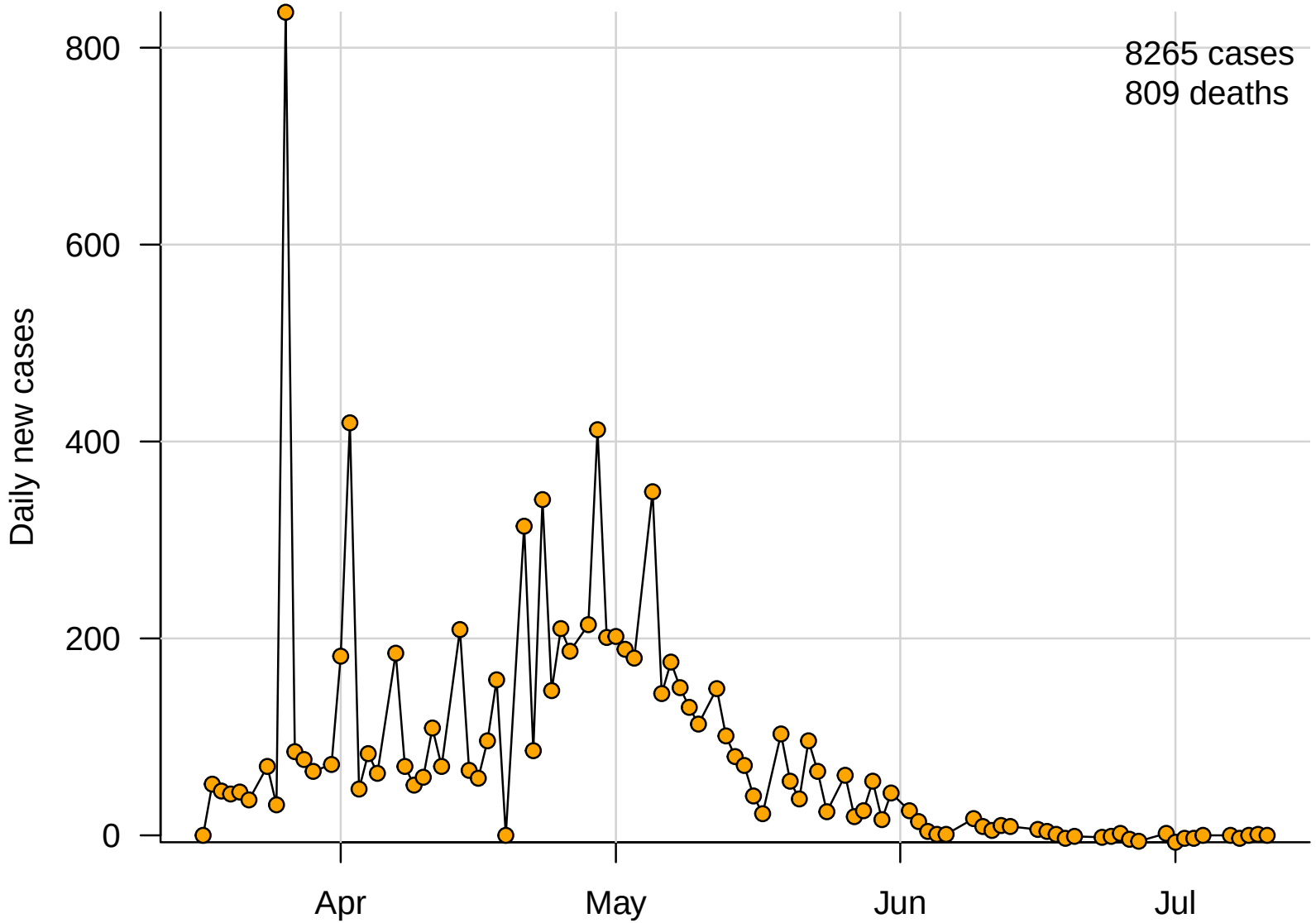
- *Note:* the Software page also contains some info about spell-checking and counting words in LaTeX documents.

Attendance

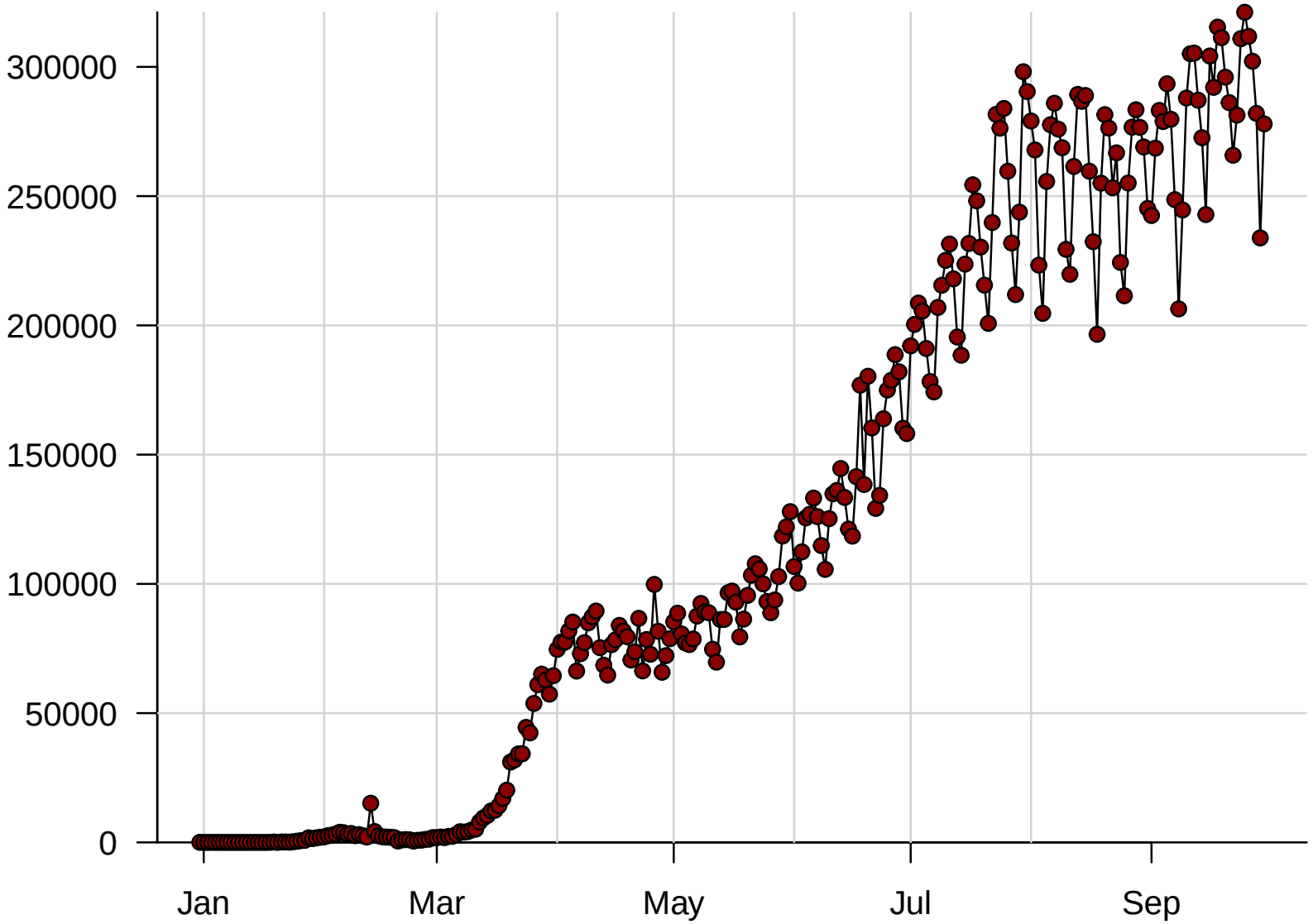
Who is here?

Epidemic Modelling

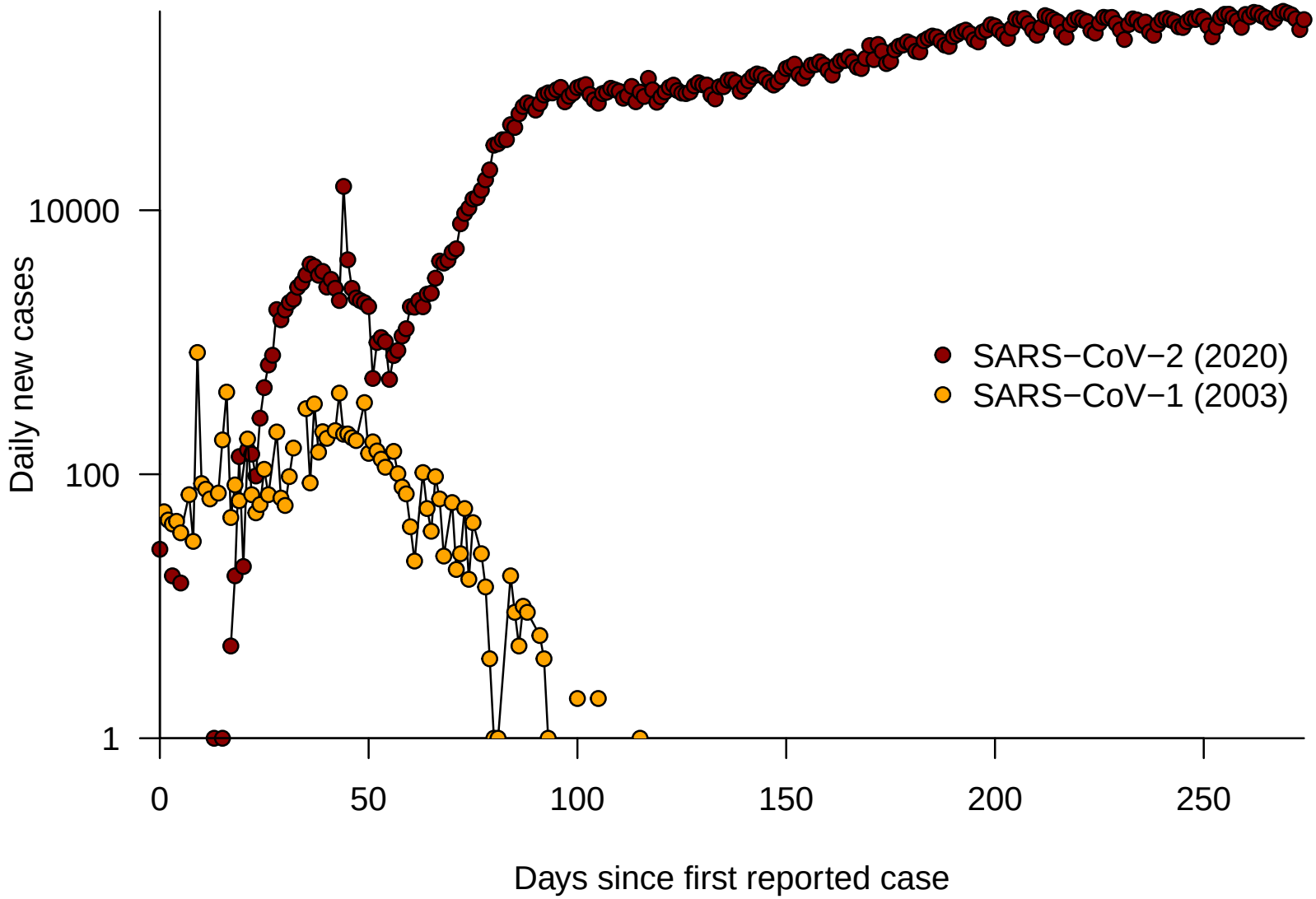
Daily SARS-CoV-1 in 2003 (Worldwide)



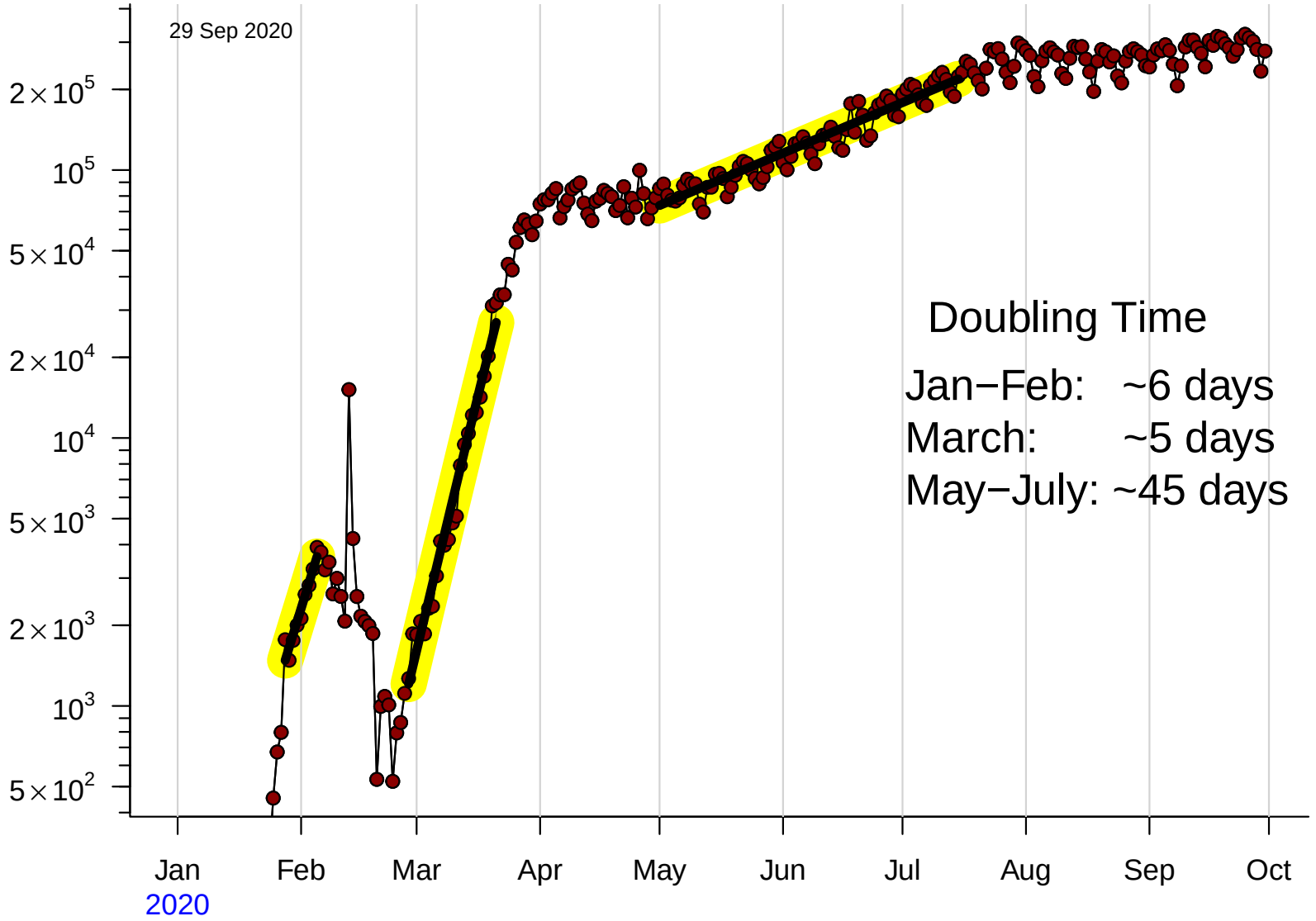
Daily SARS-CoV-2 in 2020 (Worldwide)



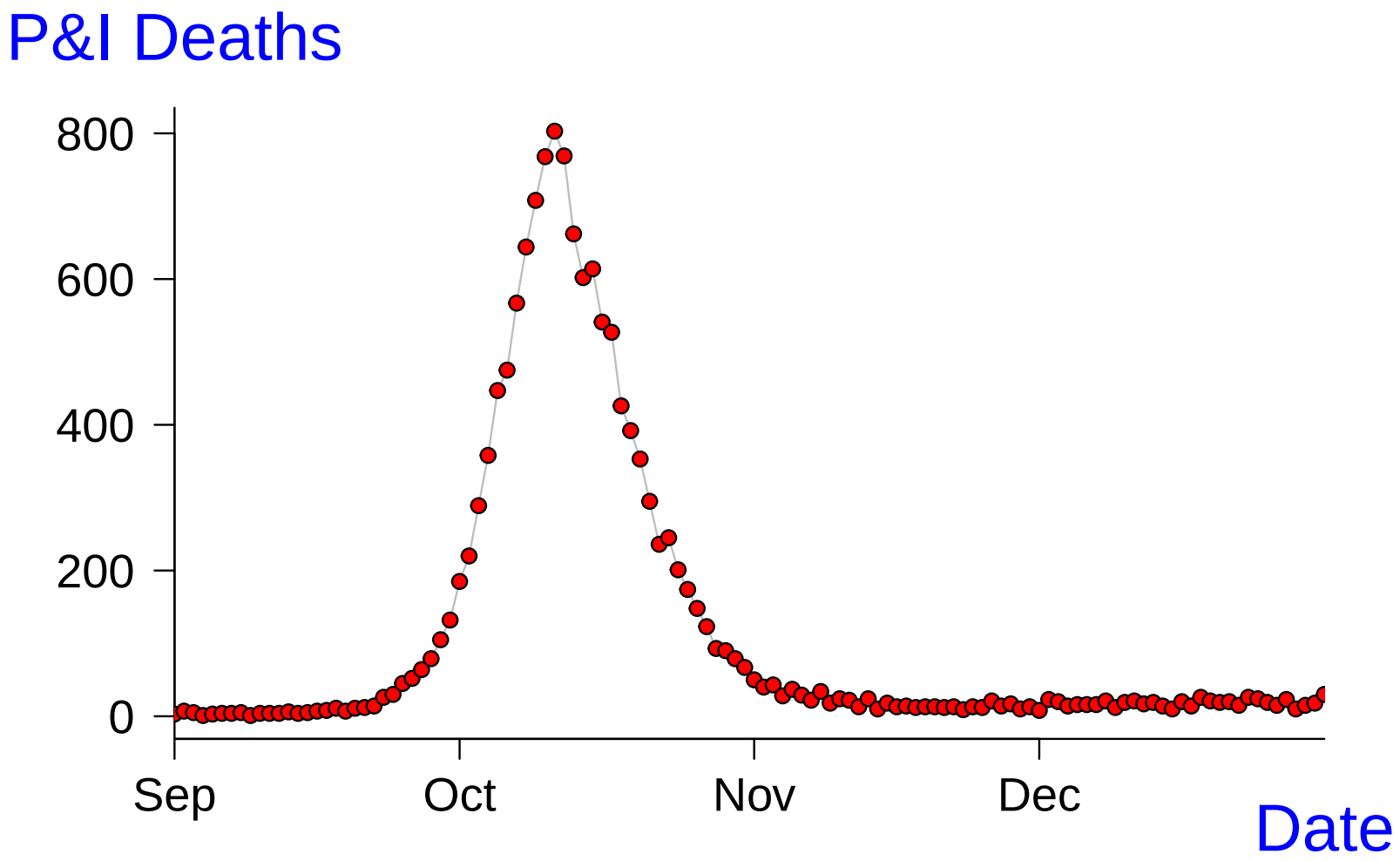
Daily SARS-CoV-1 vs SARS-CoV-2 (Worldwide)



Daily SARS-CoV-2 (Worldwide) exponential growth fits



Pneumonia & Influenza Mortality, Philadelphia, 1918



Modelling challenge

Develop a model that helps us understand the graph on the previous slide, based on mechanisms of disease spread.

- Only one variable is observed (P&I deaths per day) so construct a model containing only one variable.
- Think about how disease spreads and express your thoughts with mathematical notation.
- Derive a differential equation that models the process of disease transmission.
- Analyze the model and determine its strengths and weaknesses/limitations.

Make (Biological) Assumptions Clear

1. *Assume* the disease is transmitted by contact between an infected individual and a susceptible individual.
2. *Assume* the latent period (delay between being infected and becoming infectious) is so short that it can be ignored (technically assume it is zero).
3. *Assume* all members of the population are identical and respond identically to the disease. In particular, all susceptible individuals are equally susceptible and all infected individuals are equally infectious.
4. *Assume* the population size is fixed during the epidemic, *i.e.*, ignore births, migration, and deaths from causes other than the disease, and count individuals who have died from the disease as part of the population.

About Assumptions. . .

- Note that the first assumption on the previous slide is actually correct.
- The other assumptions are wrong, but are reasonable approximations.
- It is best to *start as simple as possible and add complexity later,* in order to:
  - obtain a model that actually succeeds in explaining the data with as few ingredients as possible;
  - identify model features that are most important;
  - check that inferences we draw from our model(s) are robust to the inclusion of more biological details/realism.

What variables should we include in our model?

- Independent variable: time (t)
- Dependent variable: Many options, *e.g.*,
  - Incidence (number of new infections per unit time)
  - Prevalence (total current number of infected individuals)
  - Death rate (number of deaths per unit time)
  - Death toll (number of deaths so far)
- So, what would be best?
- Not deaths, because whether or not you die may be unrelated to how much you transmit.
- But deaths are what we observe! What to do?!?
  - Even when we have case counts (*e.g.*, SARS-CoV-2), deaths may be more useful. Why?
- Make another assumption. . .

Additional assumption(s)

- We actually want to know incidence or prevalence, but we observe deaths.
- Under what circumstances would daily deaths be a good estimate of incidence? (*i.e.*, What must we assume in addition to the assumptions we have already made.)
  5. *Assume* that the time from infection to death is exactly the same (a certain number of days) for every individual who dies.
  6. *Assume* that the probability of dying from the disease is the same for every individual who is infected.
- Then daily death counts are proportional to daily incidence a certain number of days in the past, *i.e.*, the "mortality curve" that we observe is a translated and scaled version of the "epidemic curve" (new cases per day).

So. . . what variables should we include in our model?

- Independent variable: time (t)
- Dependent variable: one of:
  - Incidence (number of new infections per unit time)
  - Prevalence (total current number of infected individuals)
- Which one?
- Choose prevalence (I) because anybody who is currently infectious can infect others, so it will probably be easier to formulate a transmission model based on prevalence. (Try not to lose sight of underlying biological mechanisms.)
- But our mortality curve is related to incidence, not prevalence!?! Argh. What to do?!?
- Let's work with prevalence and see how it works out. Maybe we'll be able to derive the incidence curve from a model based on prevalence.

Notational note

- We use I for prevalence because prevalence is the number of infected individuals.

- So, let's try to write down a model...

A first (naïve) attempt at an epidemic model

- Variables: time t, prevalence $I(t)$
- How does I increase?
- Start with I_0 infected individuals at time $t = 0$. What happens for $t > 0$.
- Let $B =$ average number of contacts with susceptible individuals that lead to a new infective per unit time per infective in the population (and suppose B is constant). Then

$$I(t + \Delta t) \simeq I(t) + B\, I(t)\Delta t$$

- In the limit $\Delta t \to 0$, we have

$$\frac{dI}{dt} = BI \quad \Longrightarrow \quad I(t) = I_0 e^{Bt}$$

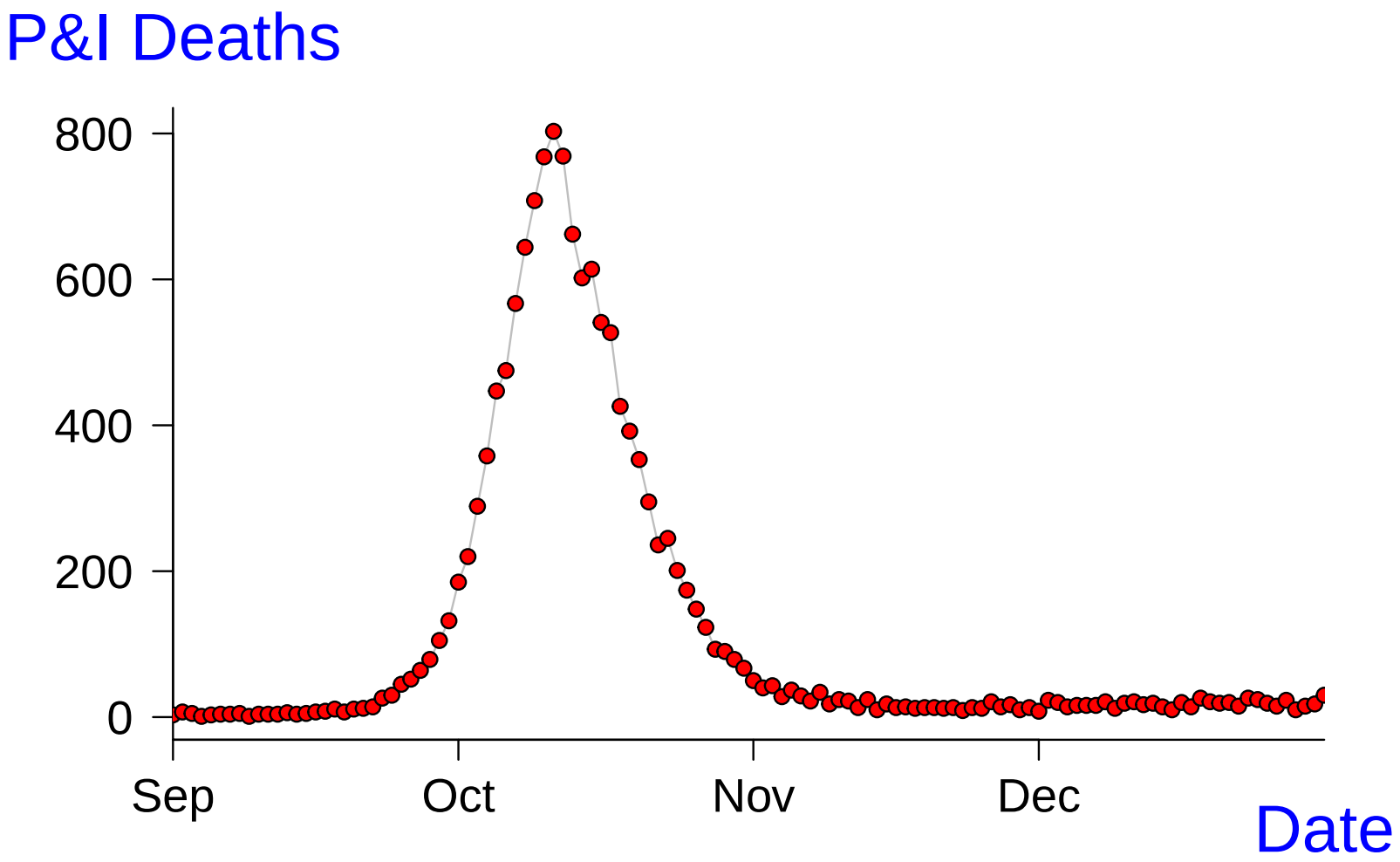
Beware: implicit assumptions that should be explicit

- Ignored discrete nature of individuals when taking limit.
- *Ignored finite infectious periods!*
  - Sometimes it isn't obvious that we've made some assumptions until after we see what the model predicts.

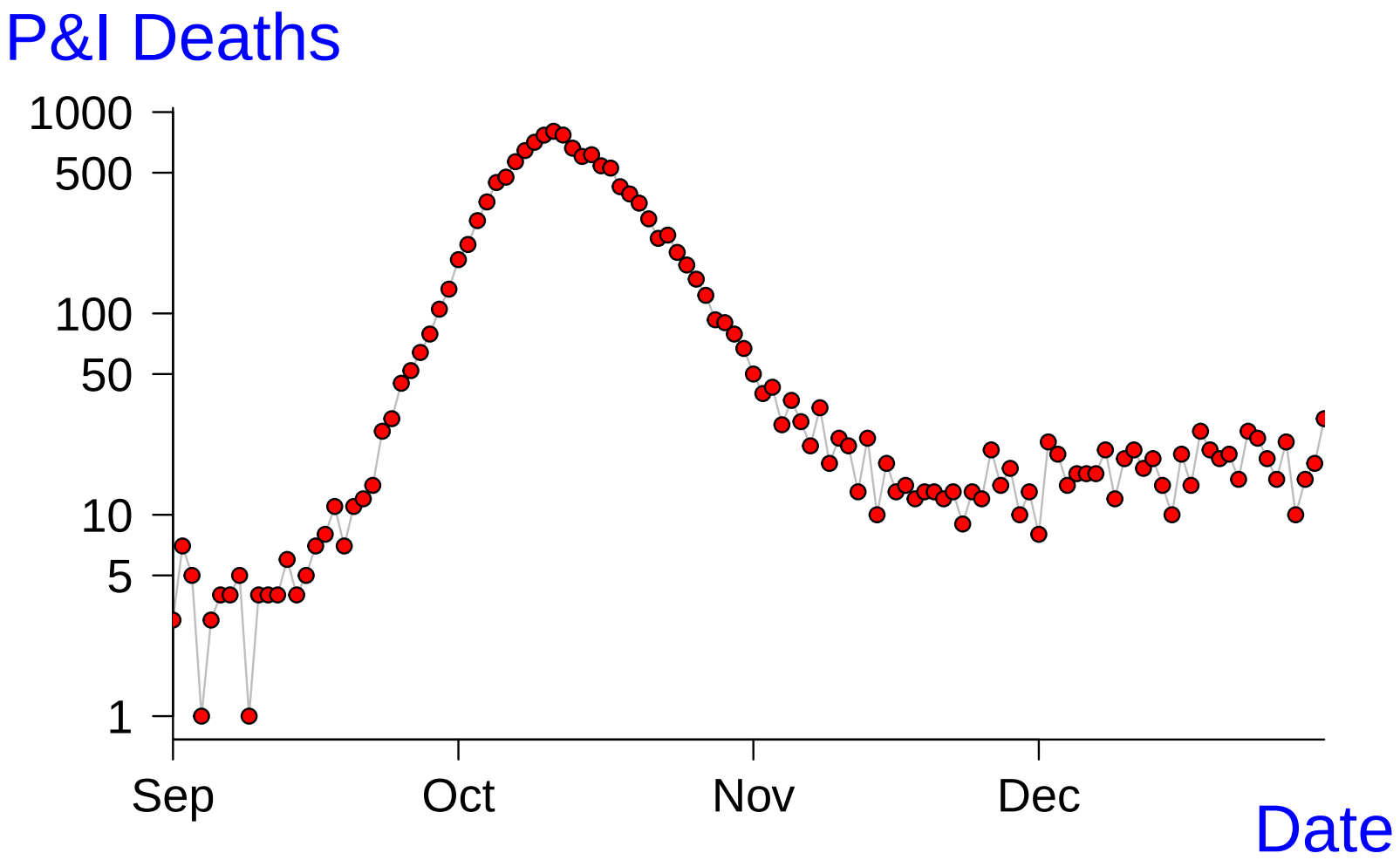
How can we tell if our model is good?

- Compare model predictions with data.
- What is the best way to do that?
- Depends on what predictions we're trying to test.
- Model predicts exponential growth.
  How should we test that prediction?
- Transforming the data might help.

Original data: P&I Mortality, Philadelphia, 1918



Logarithmic scale: P&I Mortality, Philadelphia, 1918



Parameter estimation

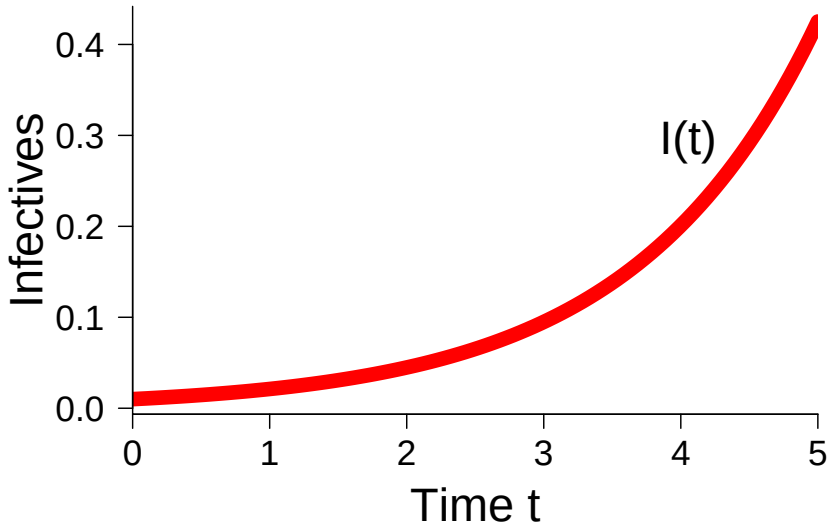
How can we estimate the model parameters, I_0 and B, from the P&I data?

- Fit a straight line through the part of the logarithmic mortality curve that looks straight.
- The slope of the line is B.
- The "intercept" is $\log I_0$.
  - "Intercept" in quotes because we need to define $t = 0$ as the time when exponential growth begins.
- *Note:* Parameter estimation is, in general, a very tricky business and deserves a great deal of attention (beyond the scope of this course).

Naïve epidemic model

- Variables: time t, prevalence $I(t)$
- Parameter B = average number of contacts with susceptible individuals that lead to a new infective per unit time per infective in the population

$$\frac{dI}{dt} = BI \quad \Longrightarrow \quad I(t) = I_0 e^{Bt}$$



Naïve model: the good and the bad

- Good:
  - Makes clear predictions
  - Predictions can be tested
  - Estimation of parameter (B) is easy
    - B is the slope of the straight portion of the epidemic curve on the log scale. (Why?)
    - Remember we are imagining that the mortality curve is equivalent to the epidemic curve after translation and scaling.
- Bad:
  - Model is consistent only with exponential growth phase.
  - Absurd long-term prediction: unbounded growth in $I(t)$
    - Implicitly assumed that population size $N = \infty$.

How can we improve our model?

- Insist that population size is finite ($N < \infty$).
- Keep track of both infectives $I(t)$ and susceptibles $S(t)$.
- Assume individuals who are *not infected* are susceptible:

$$I(t) + S(t) = N = \text{constant}.$$

New model parameter(s)?

- B = average number of contacts with susceptible individuals that lead to a new infective *per* unit time *per* infective
- In the naïve model, we assumed B = constant.
  Is B really constant?
- B depends on how many susceptibles there are.
- $B = \beta S(t)$
- β = average number of contacts between susceptibles and infectives that lead to a new infective
  per unit time
  per infective
  per susceptible
- β is called the ***transmission rate***.

Revised epidemic model: "SI model"

$$\frac{dI}{dt} = \beta S(t) I(t)$$

- Two state variables. One equation. Problem? No:

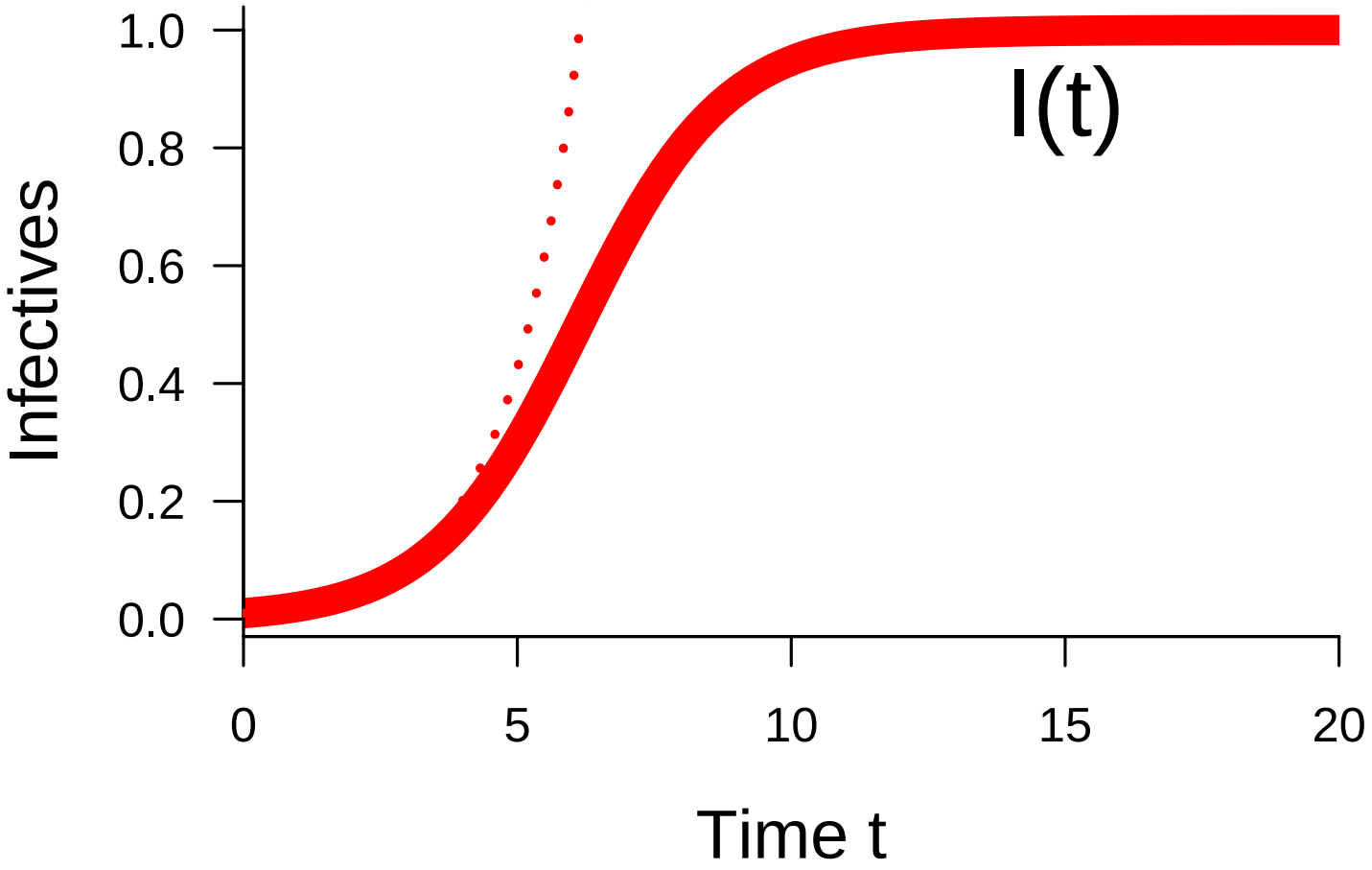
$$\frac{dS}{dt} = -\beta S(t) I(t)$$

- But $S(t) = N - I(t)$ $\quad \Longrightarrow \quad$ $I(t)$ is still the only variable:

$$\frac{dI}{dt} = \beta I(N - I)$$

- Is this a better model?
- What does it predict?

SI model: Example solution



SI model: Analysis

- We can find the exact solution. How?

$$I(t) = \frac{I_0 e^{N\beta t}}{1 + (I_0/N)(e^{N\beta t} - 1)}$$

- But exact solution is not particularly enlightening.
- Qualitative analysis:
  - Initially $I \ll N$. What does the model predict then? Exponential growth. Great!
  - As I grows, growth rate slows. Why? Fewer and fewer susceptibles to infect.
  - Asymptotic behaviour? Equilibria? Periodic orbits? (periodic orbit = recurrent epidemics)
  - (Non-trivial) periodic orbits impossible in one dimension (existence-uniqueness theorem).
  - Consider equilibria...

SI model: Equilibrium Analysis

$$\frac{dI}{dt} = \beta I(N - I)\,, \qquad I \in [0, N]$$

- Two equilibria:
  - $I = 0$ Disease Free Equilibrium (**DFE**)
  - $I = N$ Endemic Equilibrium (**EE**)
- Stability:
  - DFE is unstable ($0 < I < N \implies dI/dt > 0$)
  - EE is locally asymptotically stable (**LAS**)
  - EE is globally asymptotically stable (**GAS**)
    (stability of EE follows from $0 < I < N \implies dI/dt > 0$)
    (GAS requires a little more analysis...)
  - *Note:* In one dimension, global analysis always easy.
    In higher dimensions, often try to find Lyapunov function.
    (Lyapunov function for EE of SI model?...)
- Conclusions identical for any $\beta > 0$.

SI model: Biological Inferences

- For *any* transmission rate β:
  - Initially, exponential growth of cases.
  - Eventually, convergence to equilibrium (EE) at which *everyone* in the population is infective. hmmm...
- Is this model better than our first naïve model?
  YES.
  - Still correctly predict initial exponential growth.
  - Can match epidemic curve for longer (up to the peak).
  - Does not predict absurd unbounded growth in infective population.
  - But this model cannot explain the decline of the epidemic.
- What should we do? Two obvious options:
  1. Get depressed, drop the course.
  2. Try to improve the model.

Recall motivating data: 1918 flu in Philadelphia

- Mortality curve (linear scale)
- Mortality curve (log scale)

How can we improve on the SI model?

- Include a key biological fact:
  Individuals do not stay infectious with flu forever
- Either they *recover* and are immune thereafter, or they *die* (doesn't matter which) (well, maybe to them it does)
- Why don't we care if someone recovers or dies?
  (*i.e.*, Why doesn't it affect our dynamical inferences?)
- Because in either case the individual is *removed* from the transmission process, hence cannot affect the future pattern of the epidemic.

The SIR model

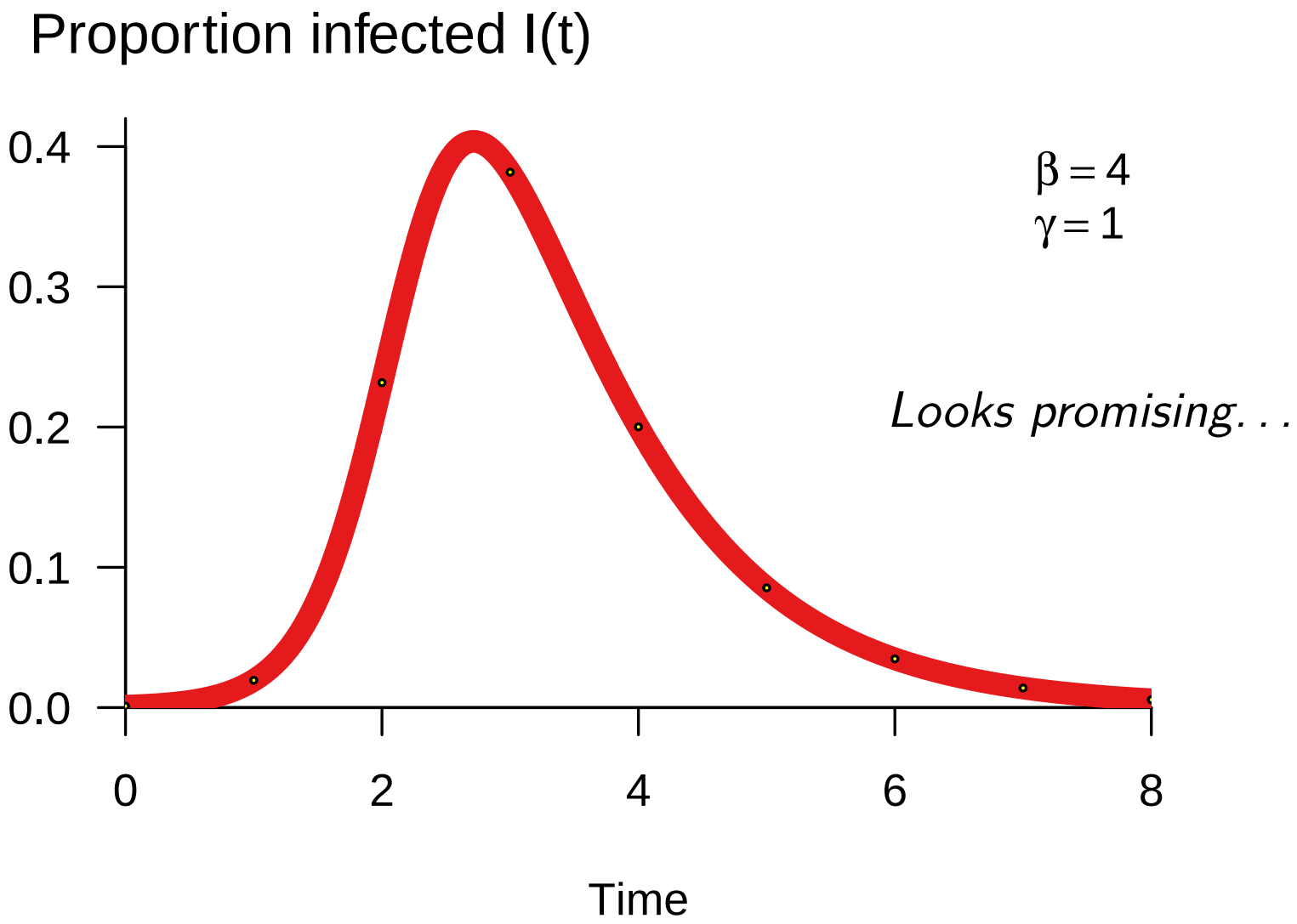
Introduce new class of removed individuals:

- $R(t)$ = number of individuals who have either recovered and are now immune or have died
- Let γ = rate of removal from the infective class (via recovery or death)

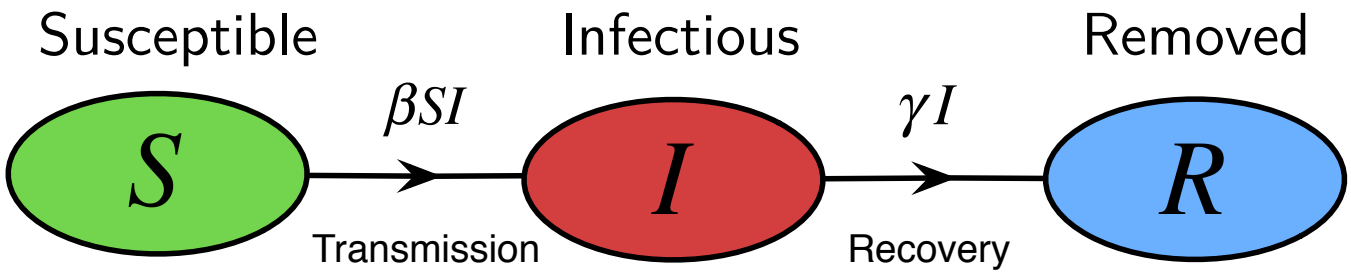
$$\frac{dS}{dt} = -\beta SI$$
$$\frac{dI}{dt} = \beta SI - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$

- Note: $\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0 \implies S + I + R = N = \text{constant}$
- Convenient to rescale variables by N and interpret S, I, R as *proportions* of the population in each disease state.

The SIR model: Example numerical solution



The SIR model: Flow Chart and Parameters



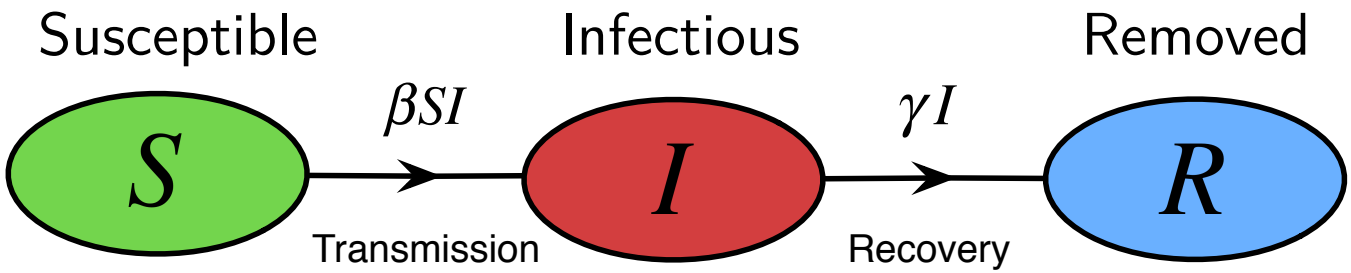
$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

- **Parameters:**
  - Transmission rate β
  - Recovery rate γ (or Removal rate)

The SIR model: Derived Parameters



$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

- **Derived Parameters:**
  - Initial growth rate $\beta - \gamma$
  - Mean infectious period $\frac{1}{\gamma}$
  - Basic Reproduction Number

$$\mathcal{R}_0 = \frac{\beta}{\gamma}$$

The SIR model: Derived parameters

The initial growth rate

- How do we calculate the initial growth rate from our model?
- Consider change in prevalence initially (when $I \ll 1$):

$$\begin{aligned}\frac{dI}{dt} &= \beta SI - \gamma I \\ &= \Big(\beta S - \gamma\Big) I \\ &\approx \Big(\beta - \gamma\Big) I \qquad \text{if } S \sim 1 \text{ initially.}\end{aligned}$$

- $\therefore$ Initially $\quad I(t) \approx I_0 e^{(\beta-\gamma)t}$.
- $\therefore$ Initial slope of logged prevalence curve is $\quad \beta - \gamma$.

The SIR model: Derived parameters

The mean infectious period

- How do we calculate the mean infectious period from our model?
- Suppose at time 0 there are I_0 infectious individuals, and suppose we can *prevent contact with susceptibles*.
- The equation for I then simplifies to

$$\frac{dI}{dt} = -\gamma I\,, \qquad I(0) = I_0$$

- We can solve this immediately to find

$$I(t) = I_0 e^{-\gamma t}$$

The SIR model: Derived parameters

The mean infectious period, continued...

- Thus, after time t, the number of people still infectious is reduced by a factor $e^{-\gamma t}$
- *i.e.*, the proportion of individuals who have an infectious period shorter than t is $1 - e^{-\gamma t}$
- *i.e.*, the cumulative distribution of the infectious period is $C(t) = 1 - e^{-\gamma t}$.
- Therefore, the probability density of the infectious period is $p(t) = C'(t) = \frac{d}{dt}\left(1 - e^{-\gamma t}\right) = \gamma e^{-\gamma t}$
- The mean of this exponential probability distribution is $\int_0^\infty t\, p(t)\, dt = \int_0^\infty t\, \gamma e^{-\gamma t}\, dt = \frac{1}{\gamma}$

The SIR model: Derived parameters

The basic reproduction number $\mathcal{R}_0$

$$\begin{aligned}\mathcal{R}_0 &= \beta \cdot \frac{1}{\gamma} \\ &= (\text{transmission rate}) \\ &\qquad \times (\text{mean infectious period})\end{aligned}$$

- $\mathcal{R}_0$ is dimensionless
- $\mathcal{R}_0$ is the average number of secondary cases caused by a primary case (in a wholly susceptible population).
- We must have $\mathcal{R}_0 > 1$ to have an epidemic. Why?
  - $\frac{dI}{dt} = \beta SI - \gamma I = \big(\mathcal{R}_0 S - 1\big)\gamma I$
  - $\therefore \mathcal{R}_0 \leq 1 \implies \frac{dI}{dt} \leq 0$ for all $(S, I) \in [0,1]^2 \implies$ no growth

The SIR model: Analysis (biological well-posedness)

$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \gamma I$$

Be careful:
Is this a sensible biological model?

- We need S, I and R all non-negative at all times.
- Does $0 \leq S(0) + I(0) \leq 1$ imply $0 \leq S(t) + I(t) \leq 1$ for all $t > 0$?
  - $S = 0 \implies S' = 0$, so $S(0) \geq 0 \implies S(t) \geq 0 \ \forall t > 0$.
  - $I = 0 \implies I' = 0$, so $I(0) \geq 0 \implies I(t) \geq 0 \ \forall t > 0$.
  - $(S + I)' = S' + I' = -\gamma I \leq 0$
    $\implies S + I$ is always non-increasing
    $\implies S(t) + I(t) \leq S(0) + I(0) \leq 1$.

The SIR model: Analysis (equilibria *etc.*)

$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \gamma I$$

- **Equilibria:** $(S, I) = (S_0, 0)$ for any $S_0 \in [0, 1]$
  - Continuum of equilibria.
  - Does this make sense?
  - Biological meaning of equilibria?

- **Linearization:**
  - $DF_{(S,I)} = \begin{pmatrix} -\beta I & -\beta S \\ \beta I & \beta S - \gamma \end{pmatrix}$
  - $DF_{(S_0,0)} = \begin{pmatrix} 0 & -\beta S_0 \\ 0 & \beta S_0 - \gamma \end{pmatrix}$
  - All equilibria are *non-hyperbolic*.

- **Periodic orbits:**
  - $(S + I)' = -\gamma I$
    $\Longrightarrow$ no periodic orbits. Why?
    - If $I(0) = 0$ then equilibrium.
    - If $I(0) > 0$ then $(S + I)' < 0$, so cannot increase back to initial state.
  - Also follows from Index Theorem (cannot enclose any equilibria).

Recap what we've done so far...

- Began analysis of standard SIR model.
- Showed SIR model:
  - is biologically well-posed
  - has a continuum of (disease-free) equilibria, all of which are non-hyperbolic
  - does not have any periodic solutions

The SIR model: Analysis

$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \gamma I$$

Nullclines:

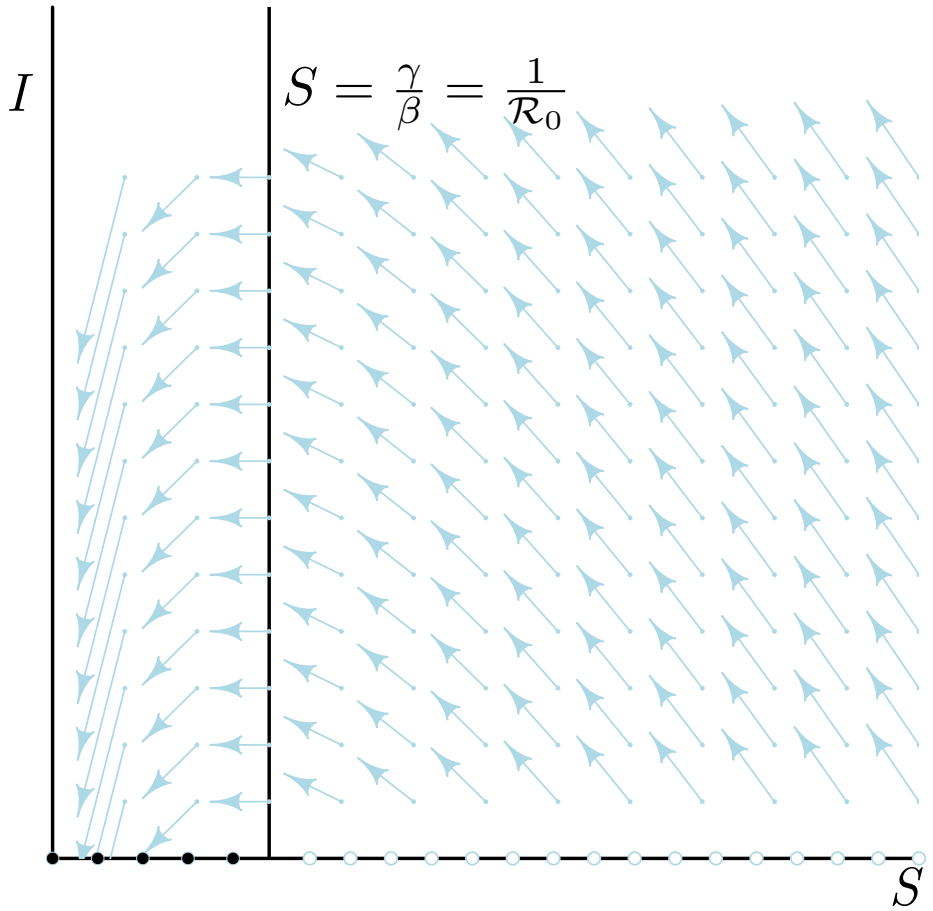
- $S' = 0 \implies S = 0$ or $I = 0$
  - S nullclines: both coordinate axes
- $I' = 0 \implies I = 0$ or $S = \gamma/\beta$
  - I nullclines: S axis and vertical line at $S = 1/\mathcal{R}_0$
  - Is the I nullcline at $S = 1/\mathcal{R}_0$ always relevant?
    - If, and only if, $\mathcal{R}_0 > 1$.
    - If $\mathcal{R}_0 < 1$ then $S = 1/\mathcal{R}_0$ is outside the biologically relevant region of the (S, I) phase plane.

The SIR model: Analysis

$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \gamma I$$

Nullclines and Direction Field ($\mathcal{R}_0 = 4$):



The SIR model: Analysis

$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \gamma I$$

Phase Portrait:

- We cannot find solutions $S(t)$ and $I(t)$ for this system.
- We *can* find exact analytical solution for the phase portrait!

- *i.e.*, we can find an expression $I(S)$ for solution curves in the (S, I) phase plane.
- Slope of $I(S)$ depends only on S:

$$\frac{dI}{dS} = \frac{dI/dt}{dS/dt} = -1 + \frac{1}{\mathcal{R}_0 S} \qquad (*)$$

- *Note:* Slope is flat for $S = 1/\mathcal{R}_0$, so max or min of $I(S)$ occurs on I nullcline if $\mathcal{R}_0 > 1$
- Easy to integrate $(*)$: $\int_{I_0}^{I} dI = \int_{S_0}^{S} \left(-1 + \frac{1}{\mathcal{R}_0 S}\right) dS$
- $I - I_0 = -(S - S_0) + \frac{1}{\mathcal{R}_0} \log(S/S_0)$

The SIR model: Analysis

Model Equations:

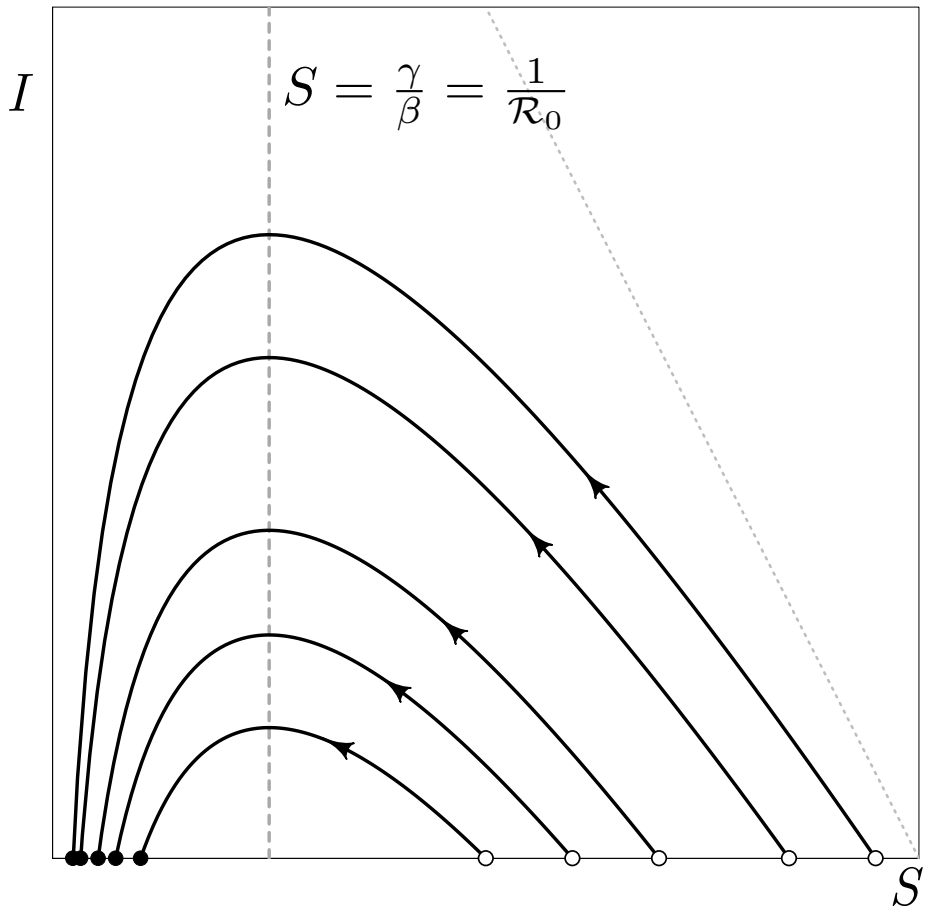
$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \gamma I$$

Solution Curves in Phase Plane:

$$I + S - (I_0 + S_0) = \frac{1}{\mathcal{R}_0} \log\left(S/S_0\right)$$

Phase Portrait ($\mathcal{R}_0 = 4$):



The SIR model: Analysis

Model Equations:

$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \gamma I$$

Solution Curves in Phase Plane:

$$I + S - (I_0 + S_0) = \frac{1}{\mathcal{R}_0} \log (S/S_0)$$

Final Size of Epidemic:

- As $t \to \infty$ we have $(I_\infty + S_\infty) - (I_0 + S_0) = \frac{1}{\mathcal{R}_0} \log S_\infty / S_0$
- But for a newly invading pathogen: $S_0 \simeq 1$, $I_0 \simeq 0$, $I_\infty = 0$
- In the limit $I_0 \to 0$, we have $(S_\infty - 1) = \frac{1}{\mathcal{R}_0} \log S_\infty$
- Define "Final Size" $Z = 1 - S_\infty$
- $\therefore -Z = \frac{1}{\mathcal{R}_0} \log (1 - Z)$, *i.e.*,

$$Z = 1 - e^{-\mathcal{R}_0 Z}$$

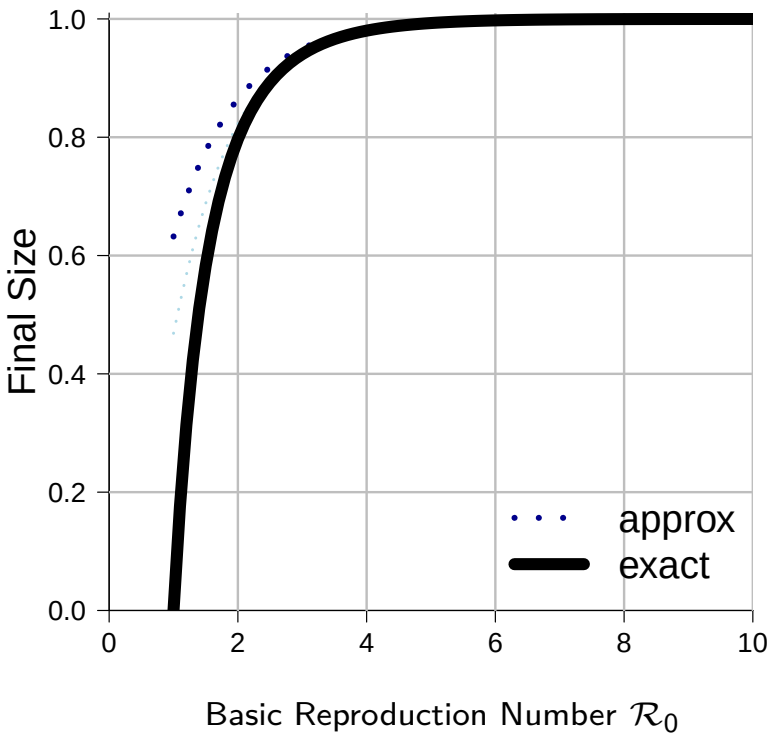
- This is a famous formula, derived by Kermack and McKendrick in 1927.

The SIR model: Analysis

Final Size Formula:

$$Z = 1 - e^{-\mathcal{R}_0 Z}$$

- Final size is determined entirely by $\mathcal{R}_0$
- Final size is never the whole population ($Z < 1$)
- Formula is valid for much more realistic models (Ma & Earn, 2006)



- For 1918 flu: $1.5 \lesssim \mathcal{R}_0 \lesssim 2$
- Proportion of world population infected?
- $\sim$60–80%

From Final Size to Reproduction Number

- The final size relation allows us to estimate the proportion of the population that will be infected *given* an estimate of $\mathcal{R}_0$.
- But we can turn it around: if we know the final size Z then we can easily estimate $\mathcal{R}_0$:

$$Z = 1 - e^{-\mathcal{R}_0 Z} \quad \Longrightarrow \quad \mathcal{R}_0 = -\frac{1}{Z}\log(1-Z)$$

- This is useful *post-hoc* only (*after* an epidemic).

The SIR model: Non-dimensionalization

- Often helpful to use dimensionless parameters.
- How do we identify "the right" dimensionless parameters?
- $\frac{\beta}{\gamma}$? $\frac{\gamma}{\beta}$? $\frac{\beta}{\beta+\gamma}$? $\frac{\beta^2}{\beta^2+\gamma^2}$?
- We choose β/γ because it has a natural interpretation.
- But we are still left with γ as a second parameter.
- Can we simplify the model somehow?
- γ defines a time scale ($1/\gamma$ is the mean infectious period).
- If time unit is mean infectious period, then $\gamma = 1$.
- So in these "natural" time units, the SIR model is

$$\frac{dS}{dt} = -\mathcal{R}_0 SI, \quad \frac{dI}{dt} = \mathcal{R}_0 SI - I$$

- There is really only one parameter in the model. The other is just a time scale and does not affect the *qualitative* dynamics.

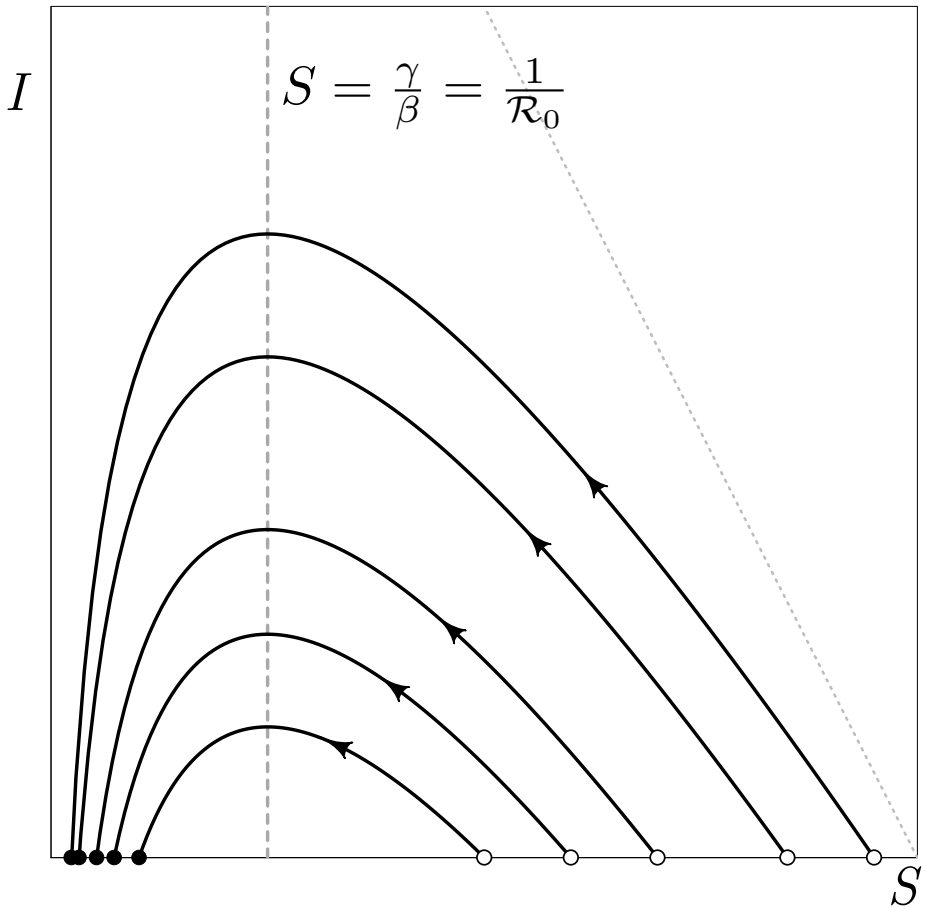
The SIR model: Results so far

Mathematical Results:

- Model is biologically well-posed
  - $0 \leq S(0) + I(0) \leq 1 \implies 0 \leq S(t) + I(t) \leq 1 \quad \forall t > 0$
- No periodic orbits.
- Continuum of equilibria.
- Stability of equilibria:
  - Linearization useless (all equilibria non-hyperbolic).
  - Further analysis necessary.
- Exact solution for phase portrait:
  $$I(S) = I_0 + (S_0 - S) + \frac{1}{\mathcal{R}_0} \log(S/S_0)$$
- Final size formula: $Z = 1 - e^{-\mathcal{R}_0 Z}$

The SIR model: Stability of equilibria

Phase Portrait ($\mathcal{R}_0 = 4$):



Model Equations:

$$\frac{dS}{dt} = -\beta SI, \quad \frac{dI}{dt} = \beta SI - \gamma I$$

Which equilibria are:

- Unstable?
  - $S_0 > 1/\mathcal{R}_0$
- Stable?
  - $S_0 \leq 1/\mathcal{R}_0$
- Asymptotically stable?
  - None!
- How do we prove these facts?

The SIR model: Effects of Control Measures

- How could the final size of the 1918 pandemic have been reduced?
- Any intervention that reduces $\mathcal{R}_0$ reduces the final size.
- What could have been done to reduce $\mathcal{R}_0$?
- Masks? Quarantine? Isolation?
- Ideally a vaccine, but no such luck in 1918.
- Even in 2009, it took months to mass-produce vaccine.
- But suppose there had been a vaccine immediately...
- What proportion (p) of the population do we need to vaccinate to eradicate an infectious disease?

The SIR model: Effects of Control Measures

Suppose a proportion (p) of the population is vaccinated before an epidemic starts. Then:

- At the start of the epidemic, the proportion of the population that is susceptible is $S_0 = 1 - p$.
- $\therefore$ Initially (at time $t = 0$) the rate of change of prevalence is

$$\begin{aligned} \left.\frac{dI}{dt}\right|_{t=0} &= \Big((\mathcal{R}_0 S - 1)I\Big)\Big|_{t=0} = (\mathcal{R}_0 S_0 - 1)I_0 \\ &= (\mathcal{R}_0(1-p) - 1)I_0 \quad < 0 \iff \mathcal{R}_0(1-p) < 1 \end{aligned}$$

- $\therefore$ An epidemic will be prevented if

$$p > p_{\text{crit}} = 1 - \frac{1}{\mathcal{R}_0}$$

- $\therefore$ Public Health Agency will ask you to estimate $\mathcal{R}_0$.

The SIR model: Results so far

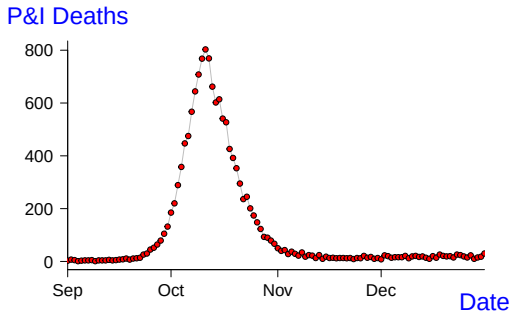
Biological inferences:

- $\mathcal{R}_0$ is extremely important to estimate in practice!
- Epidemic occurs if and only if $\mathcal{R}_0 > 1$.
- Single epidemic, then disease disappears.
- Can prevent epidemic by vaccinating (or otherwise removing) a proportion $1 - \frac{1}{\mathcal{R}_0}$ from the transmission process.

  Note: It doesn't matter whether we remove people from the susceptible pool by vaccination, isolation, or other means. What matters is the proportion of the population who are removed from the transmission process.

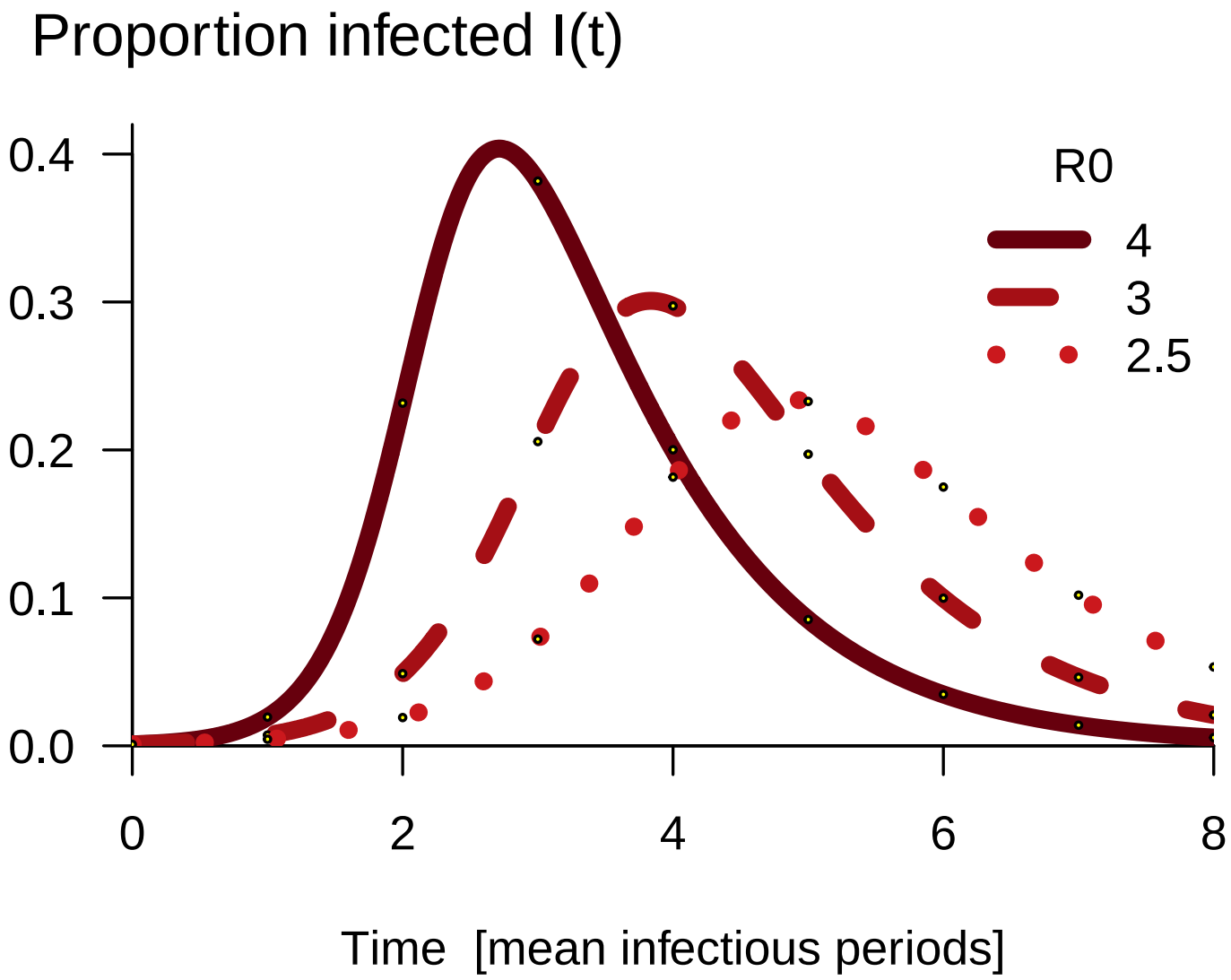
The SIR model: Does it explain our data?

What about 1918 flu in Philadelphia?



- Does the SIR model explain these data?
- Can we fit the SIR model to the P&I time series?
- If so, are our estimated parameter values (for $\mathcal{R}_0$ and $1/\gamma$) biologically reasonable?

The SIR model: How solutions depend on $\mathcal{R}_0$



The SIR model: prevalence *vs.* incidence

- In the SIR model as we have defined it, prevalence is $I(t)$ and incidence is
$$i(t) = \beta S(t) I(t)\,,$$
so we can compute incidence $i(t)$ from solutions to the SIR model, and then compare predicted incidence with observed reports of cases or deaths.
- Could we have derived an equally sensible model starting from the closer-to-observable quantity (incidence i) ?
- The answer is YES,
$$\frac{\mathrm{d}S}{\mathrm{d}t} = -i(t)\,, \tag{1a}$$
$$i(t) = \mathcal{R}_0\, S(t) \int_0^\infty i(t-s)\, g(s)\, \mathrm{d}s\,, \tag{1b}$$
where $g(s)$ is the ***generation interval distribution***.
- How do solutions of this integro-differential equation differ from those of the SIR model as we have defined it?

If you are curious, see Champredon, Dushoff & Earn 2018.

McMaster University

Mathematics and Statistics

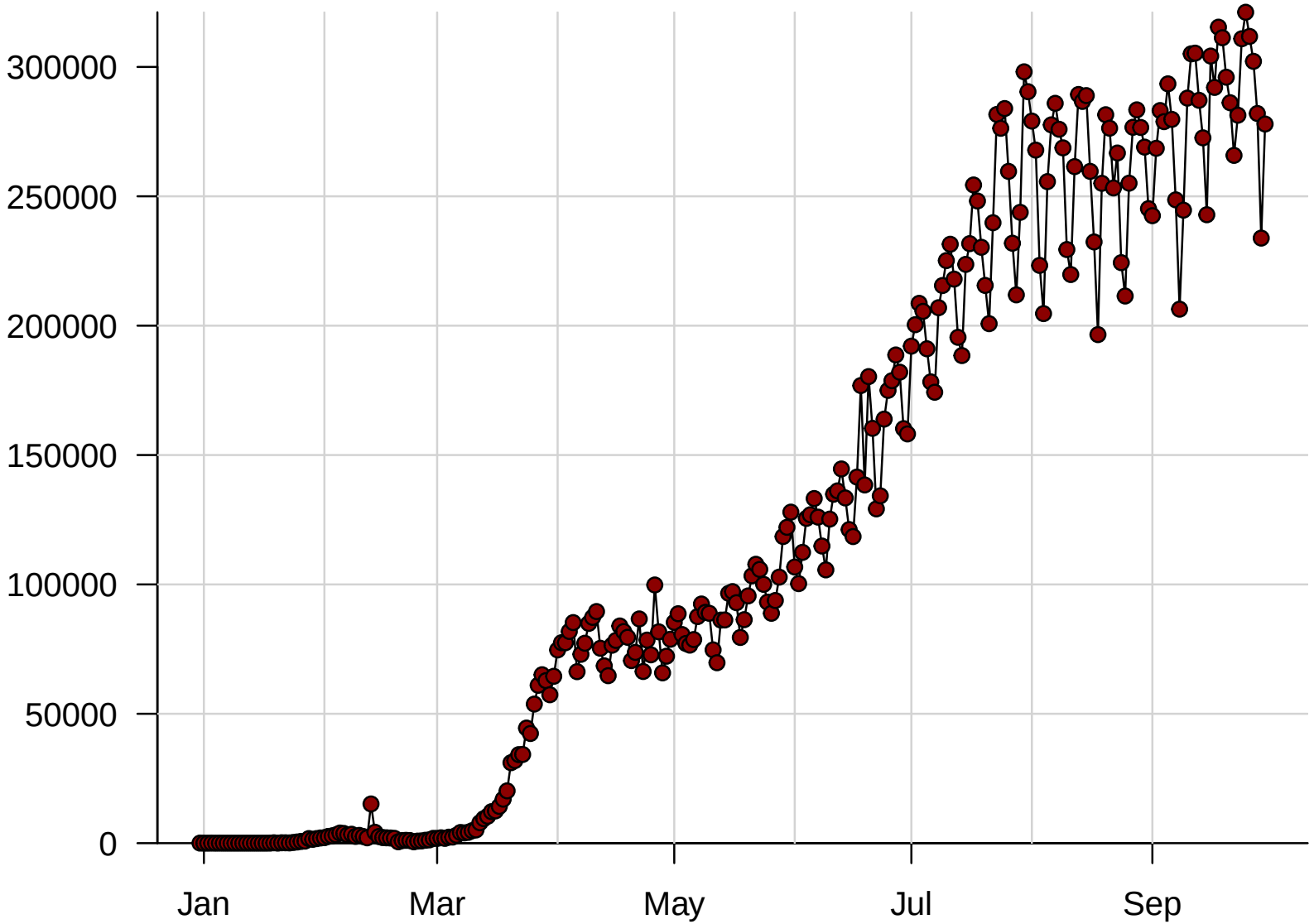
$$\int_M d\omega = \int_{\partial M} \omega$$

Mathematics 747 / 5GT3
Topics in Mathematical Biology

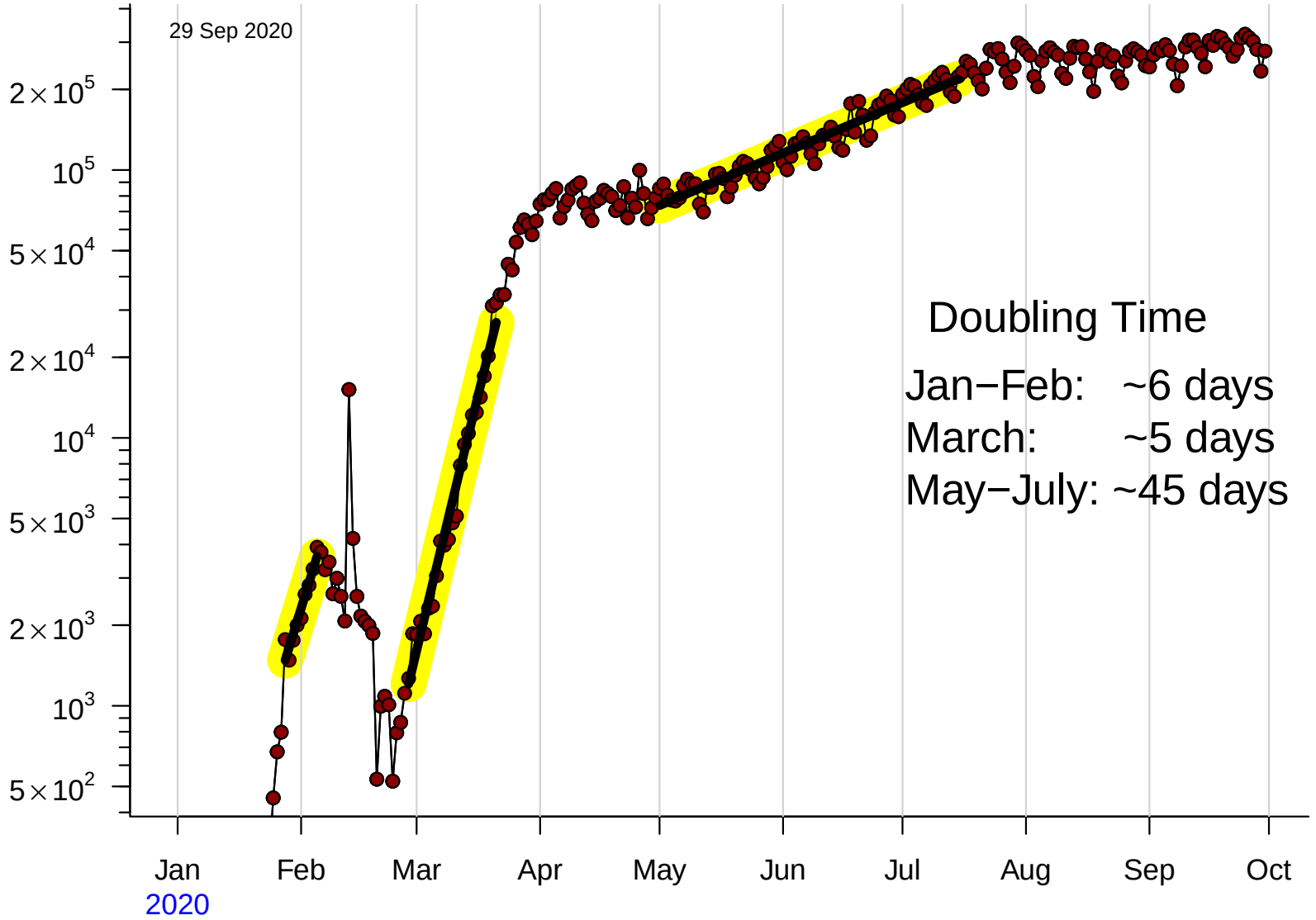
Instructor: David Earn

Lecture 2
Epidemic Modelling Intro 2
Thursday 24 September 2020

Daily SARS-CoV-2 in 2020 (Worldwide)



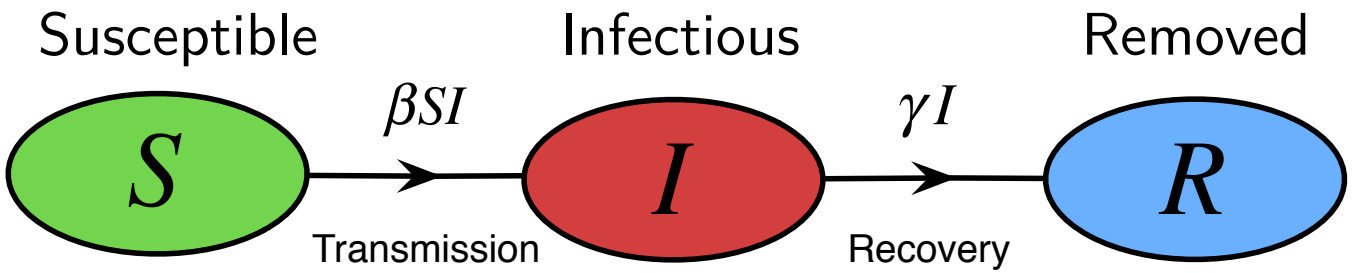
Daily SARS-CoV-2 (Worldwide) exponential growth fits



Mechanistic Epidemic Modelling: Principles

- Consider the biological mechanisms involved in disease transmission and spread
- Model mechanisms and infer their effects
- Start as simple as possible!
- Rule out simple models by comparing results with observed time series of incidence or mortality
- Add complexity one step at a time, so key mechanisms can be identified
- Ideally converge on simplest possible model that can explain observed patterns

The SIR model: Flow Chart and Parameters



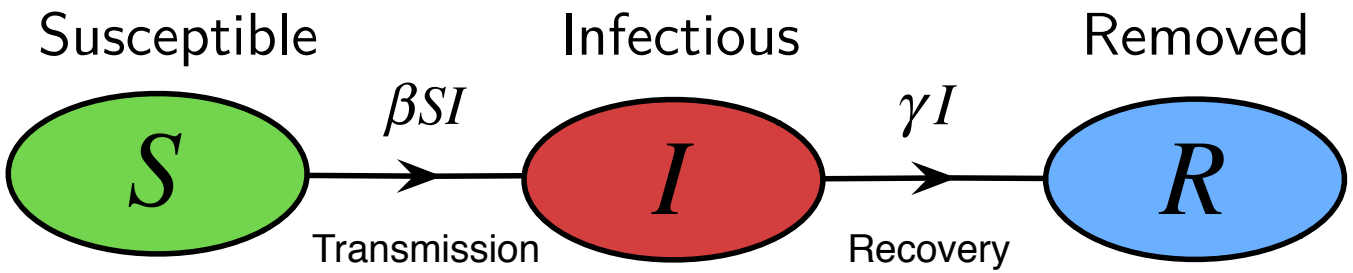
$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

- **Parameters:**
  - Transmission rate β
  - Recovery rate γ (or Removal rate)

The SIR model: Derived Parameters



$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

- **Derived Parameters:**
  - Initial growth rate $\beta - \gamma$
  - Mean infectious period $\frac{1}{\gamma}$
  - Basic Reproduction Number

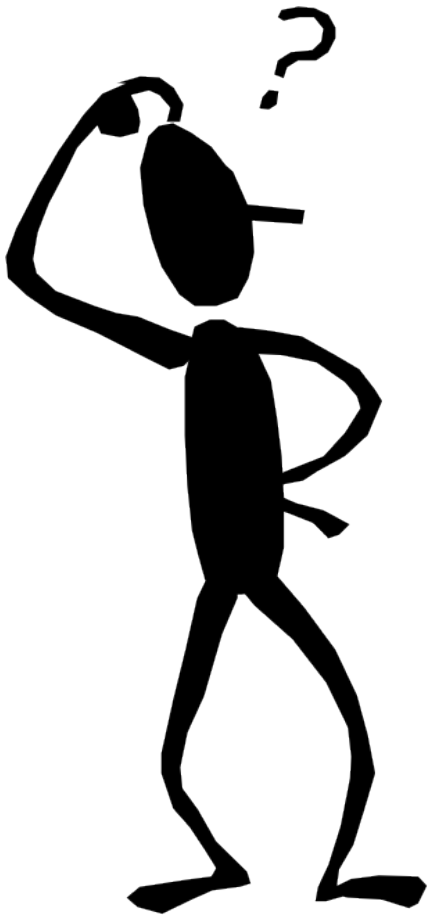
$$\mathcal{R}_0 = \frac{\beta}{\gamma}$$

Basic SIR Model: Important Results

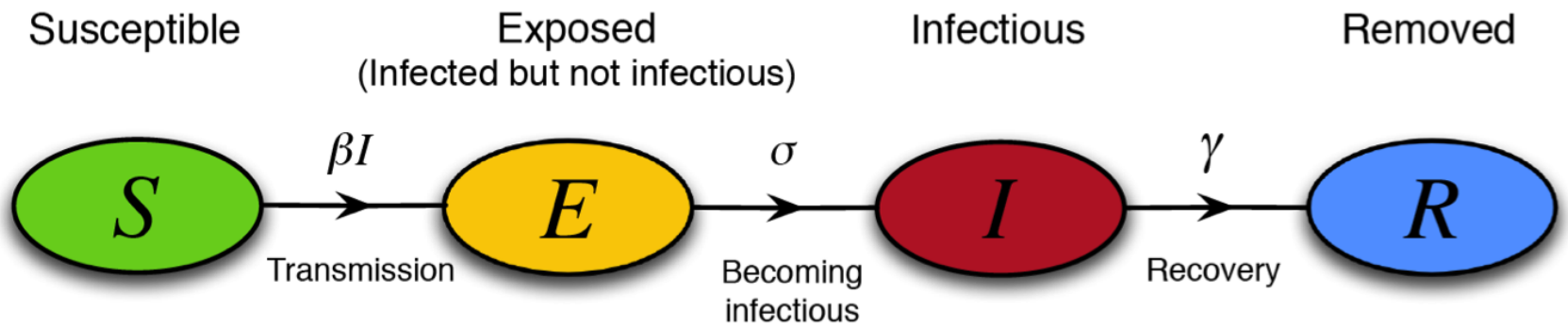
- Epidemic occurs if and only if $\mathcal{R}_0 > 1$
- Exact solution for phase portrait
- Single epidemic, then disease disappears
- Exact formula for final size as a function of $\mathcal{R}_0$

- Cannot explain diseases that persist
- Cannot explain recurrent cycles of epidemics

What are we missing?

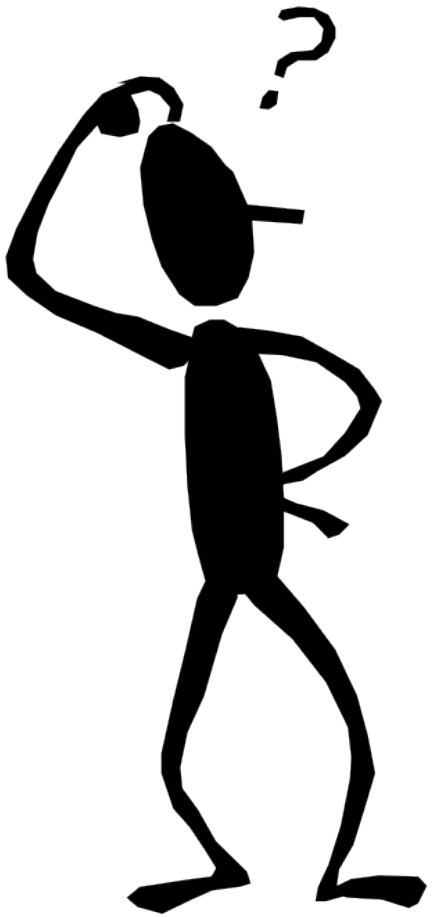


SEIR Model: flow chart

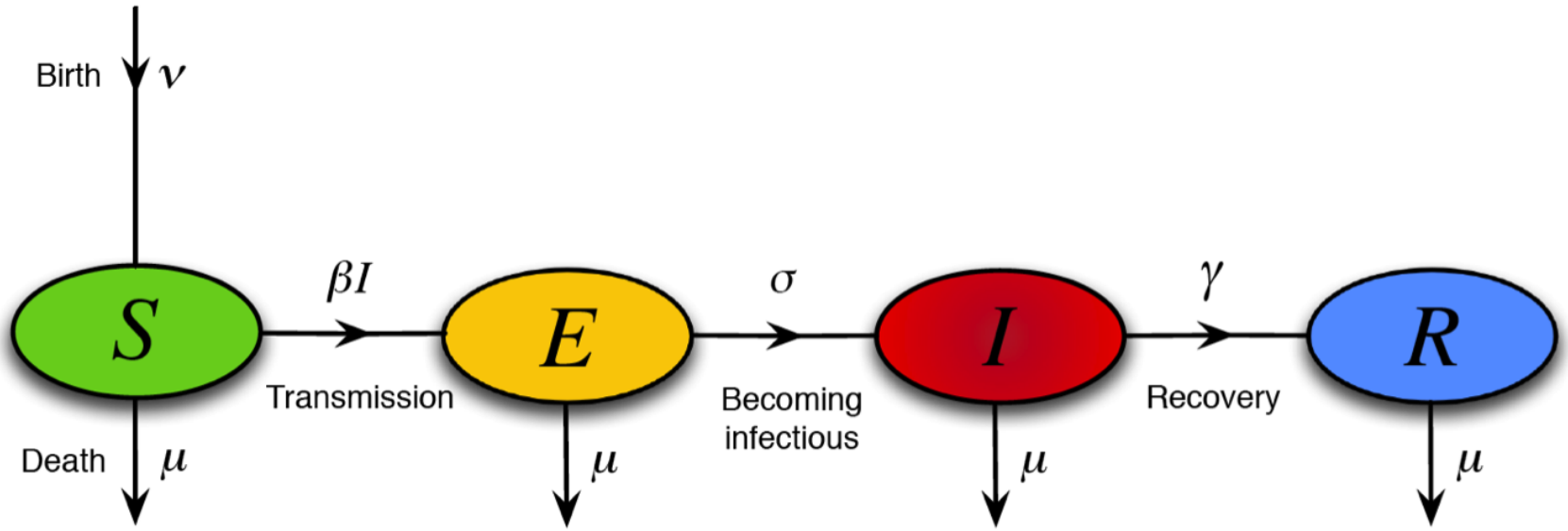


- Introduces only one new parameter (σ)
- Mean latent period ($1/\sigma$) can often be estimated
- Potentially important if there is a long latent period
- But... we still get only a single epidemic...

What are we still missing?



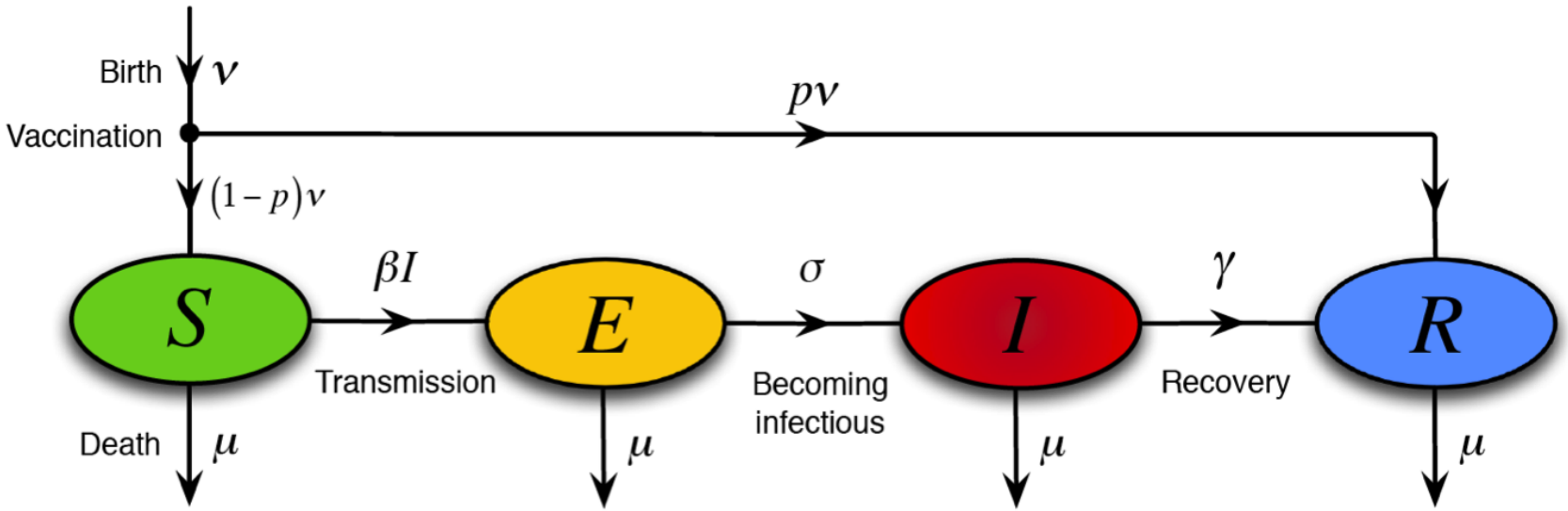
SEIR Model with vital dynamics: flow chart



New Parameters:

- Birth rate (ν for natality)
- Death rate (μ for mortality)
- Mean latent period ($1/\sigma$)
- What if we have a vaccine?

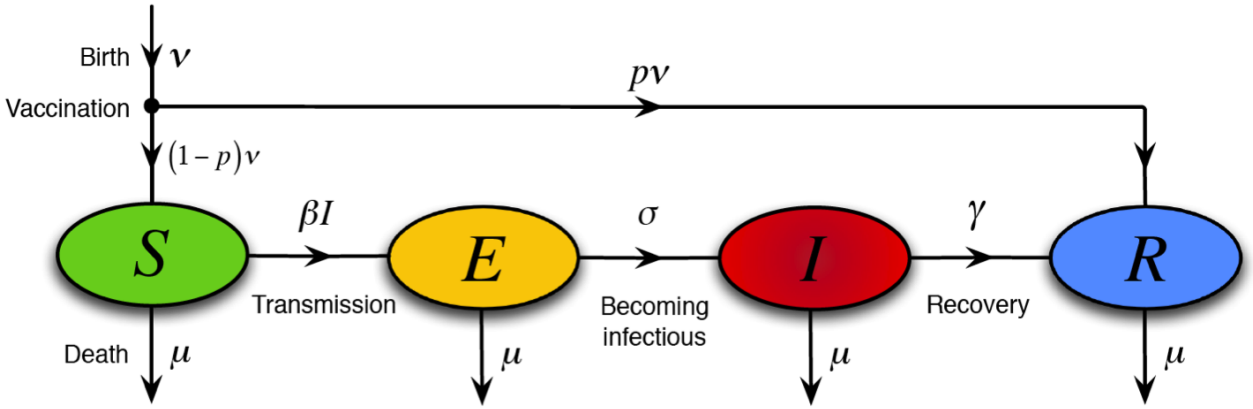
SEIR with vital dynamics and vaccination: flow chart



New Parameters:

- Birth rate (ν for natality)
- Death rate (μ for mortality)
- Mean latent period ($1/\sigma$)
- Proportion vaccinated (p)

SEIR with vital dynamics and vaccination: Equations



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

$$\frac{dE}{dt} = \beta SI - \sigma E - \mu E$$

$$\frac{dI}{dt} = \sigma E - \gamma I - \mu I$$

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

- Birth rate (ν for natality)
- Death rate (μ for mortality)
- Proportion vaccinated (p)
- Transmission rate (β)
- Mean latent period ($1/\sigma$)
- Mean infectious period ($1/\gamma$)

SEIR with vital dynamics and vaccination: Analysis

- $\mathcal{R}_0$?
  - Biological derivation: (assuming $\nu = \mu$ and $p = 0$)

$$\mathcal{R}_0 = \begin{pmatrix}\text{Transmission}\\ \text{rate}\end{pmatrix} \times \begin{pmatrix}\text{Probability of}\\ \text{surviving latency}\end{pmatrix} \times \begin{pmatrix}\text{Mean time}\\ \text{infectious}\end{pmatrix}$$

$$= \beta \times \frac{\sigma}{\sigma+\mu} \times \frac{1}{\gamma+\mu}$$

$$\simeq \frac{\beta}{\gamma} \qquad \because \frac{1}{\mu} \gg \max\left(\frac{1}{\sigma}, \frac{1}{\gamma}\right)$$

  - Mathematical derivation:
    $\mathcal{R}_0 = 1$ is stability boundary

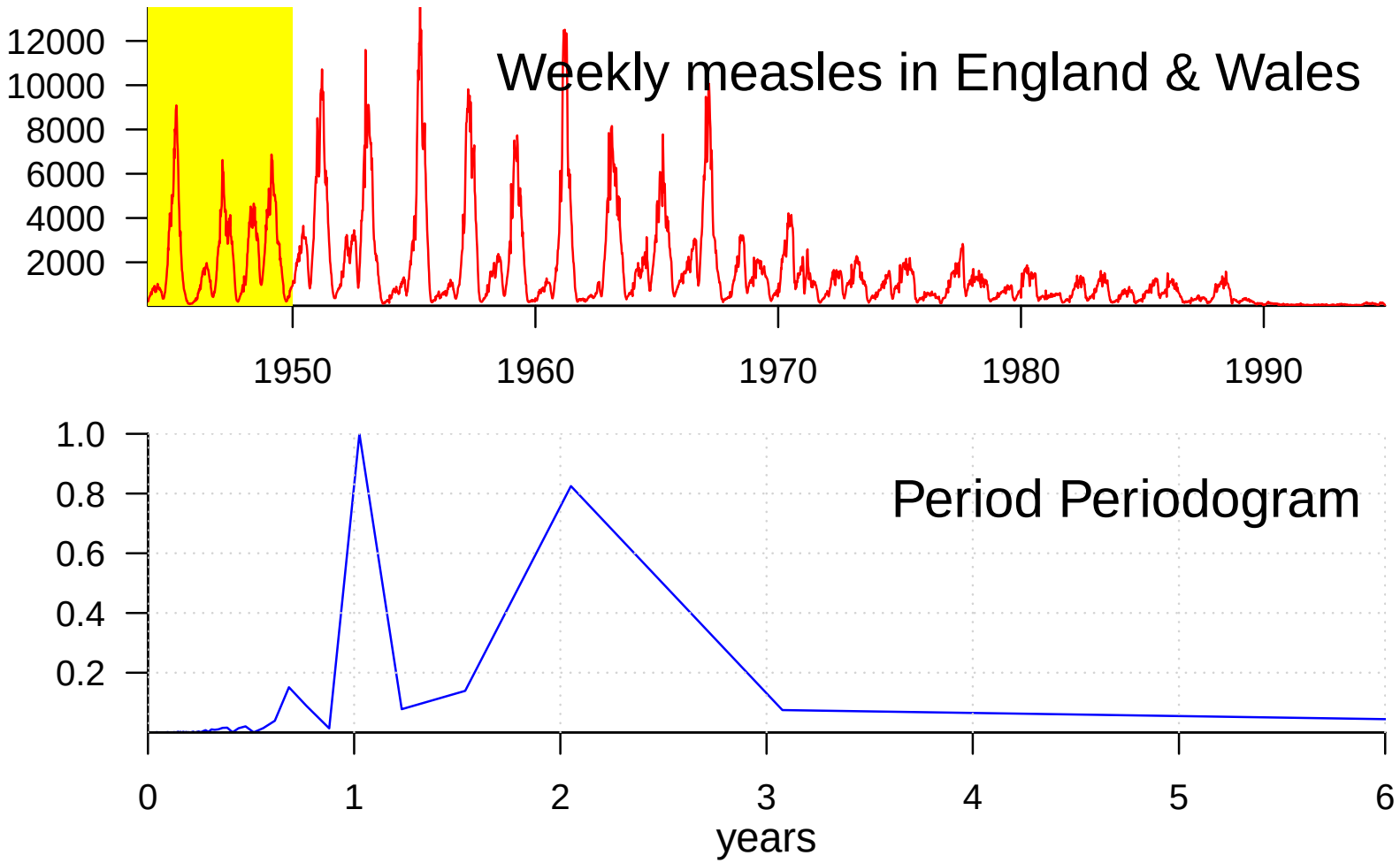
SEIR with vital dynamics and vaccination: Analysis

- Final size ?
  - If $\nu = \mu = 0$: same formula as for SIR model
  - Otherwise: not well defined ($\because$ continuous source of new susceptibles)
- Equilibria ?
  - Disease Free Equilibrium (DFE)
  - Endemic Equilibrium (EE)
  - That's all folks.
- Periodic solutions ? No.
- What else ? Chaos?

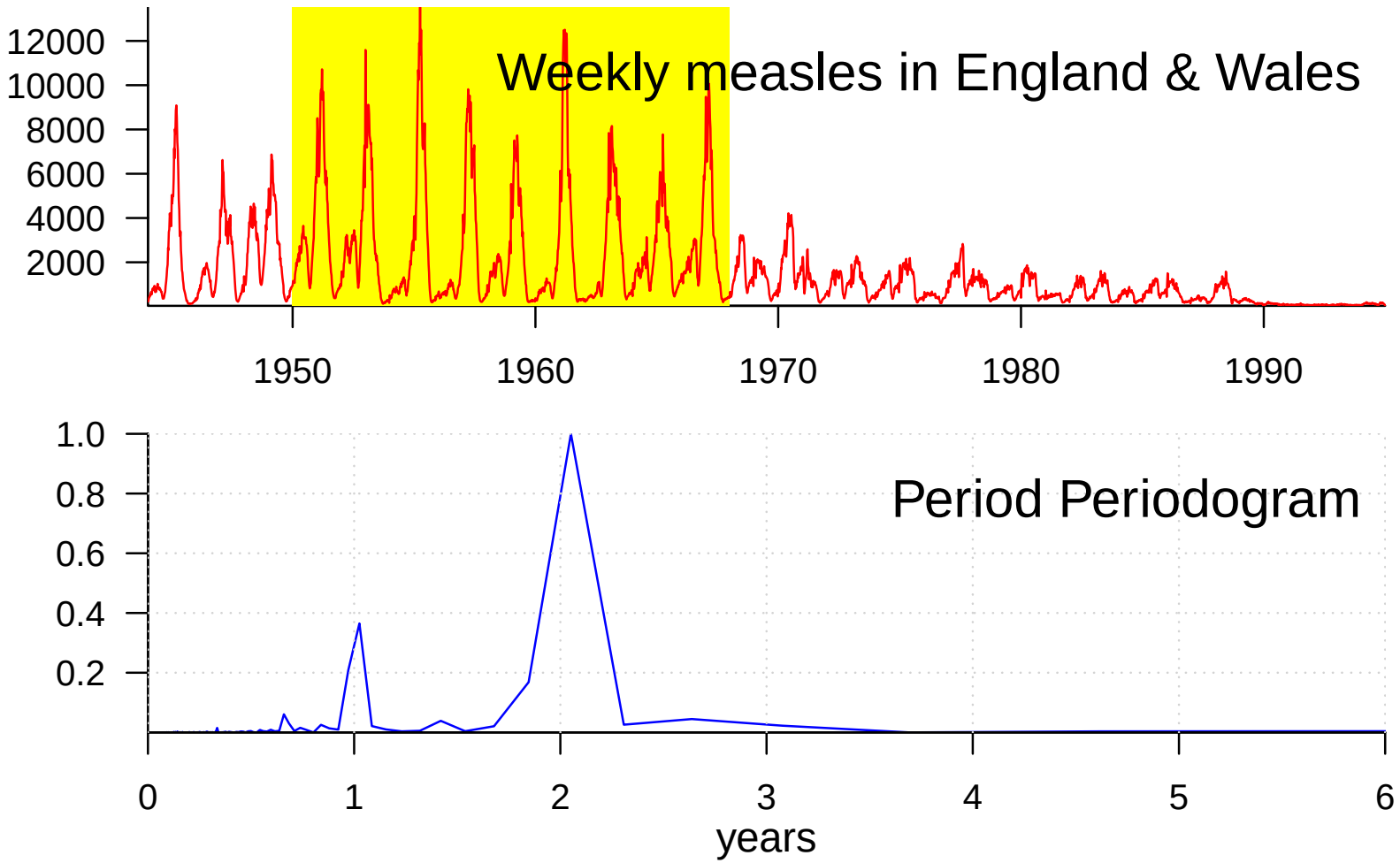
SEIR with vital dynamics and vaccination: Results

- $\exists$ Endemic Equilibrium $\iff \mathcal{R}_0(1-p) > 1$
  - EE is GAS in this case.
  - DFE is GAS otherwise.
- Eradication $\iff p > 1 - \frac{1}{\mathcal{R}_0}$ (herd immunity)
  - Smallpox: $\mathcal{R}_0 \sim 4 \implies p_{\text{crit}} \sim 75\%$
  - Measles: $\mathcal{R}_0 \sim 20 \implies p_{\text{crit}} \sim 95\%$
  - Covid-19: $\mathcal{R}_0 \sim 3 \implies p_{\text{crit}} \sim 66\%$
- Explains persistence of diseases (via births)
- No periodic solutions $\overset{?}{\implies}$ no recurrent epidemics
- GAS equilibrium $\implies$ no periodic solutions and no chaos
- Equilibrium approached by *damped oscillations* $\implies$ recurrent epidemics
- But typical epidemic patterns of persistent diseases show *undamped* oscillations...

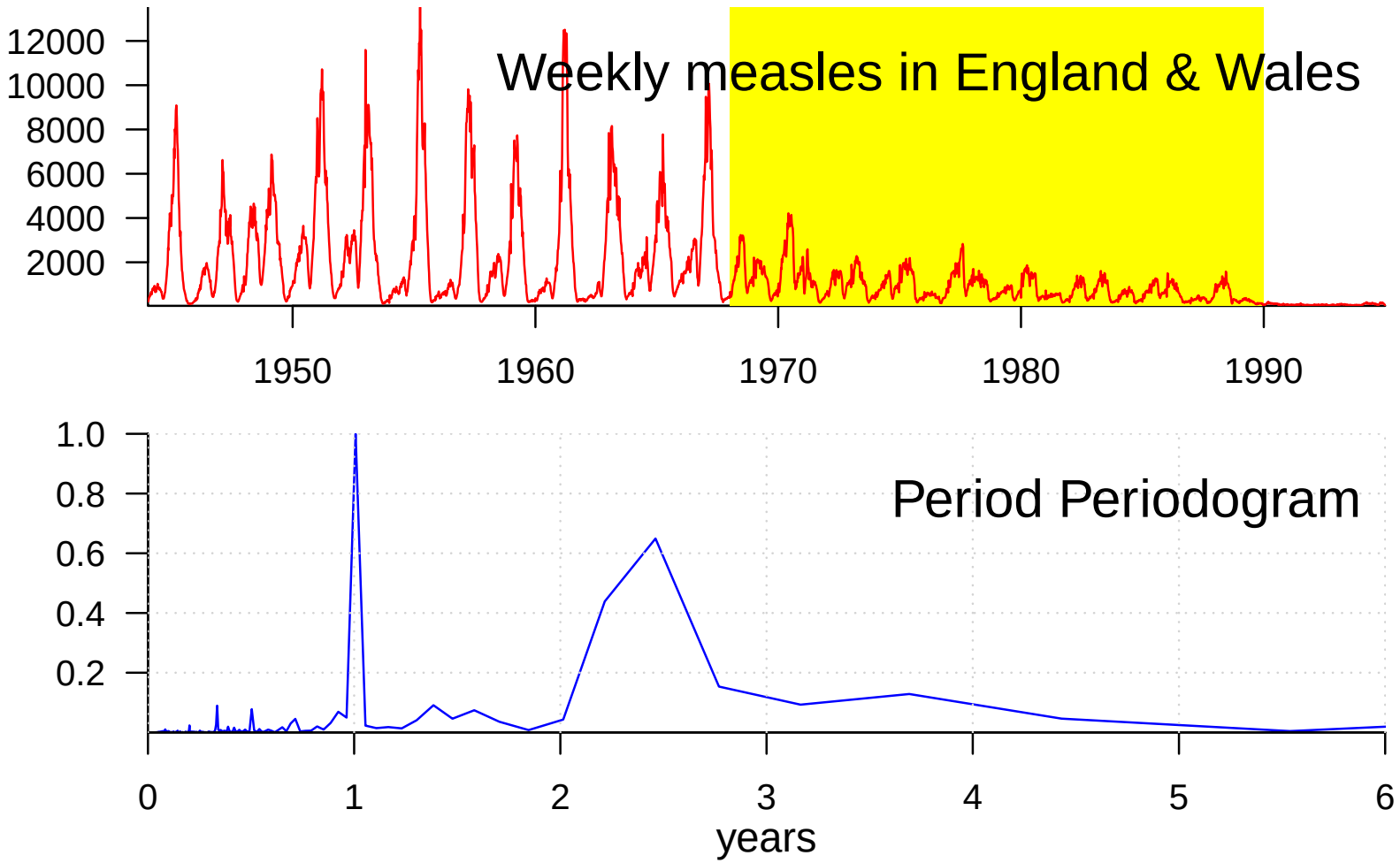
20th century measles epidemics in England and Wales



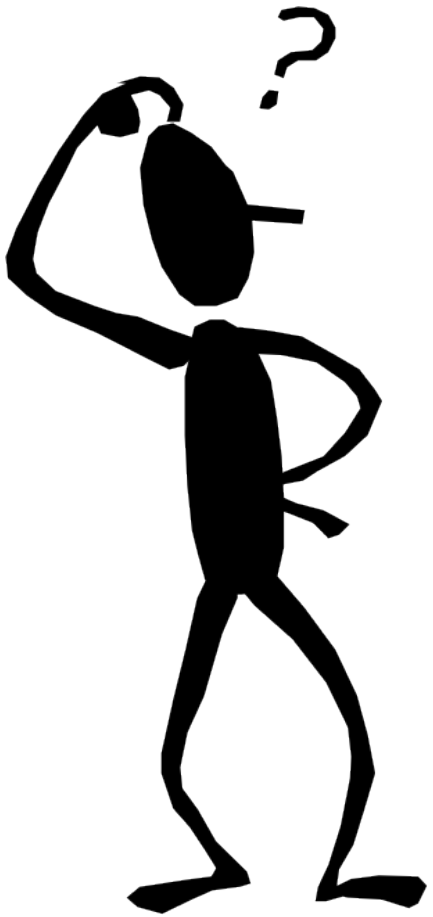
20th century measles epidemics in England and Wales



20th century measles epidemics in England and Wales



What are we STILL missing?



Demographic Stochasticity

- Differential equations describe the expected behaviour in the limit that the population size goes to infinity
- How do dynamics differ in finite populations?
- Re-cast the SEIR model as a stochastic process (Continuous time Markov process)
- Proving anything about stochastic epidemic models is difficult, but we can easily simulate them and learn a lot
- Standard algorithm for creating realizations of a stochastic epidemic model attributed to Daniel T. Gillespie

Gillespie 1976, *J. Comp. Phys.* **22**, 403–434

  - Rather than rates of change of compartment sizes consider event rates for transitions between disease states
  - Finite number of individuals
  - Assume event rates depend only on current state of population

Gillespie Algorithm

- Let a_1, a_2, ..., be the rates at which the various processes occur, *e.g.*,
  - a_1 = birth rate,
  - a_2 = rate of going from susceptible to exposed,
  - a_3 = the rate of going from infectious to removed (recovering),
  - *etc.*
- Let a_0 be the overall event rate, *i.e.*, $a_0 = \sum_i a_i$ (so average time between events $= 1/a_0$).
- Assume time spent in any state is exponentially distributed (transitions between states are "Poisson processes")
- $\therefore$ Probability next event occurs in $(t, t + dt)$ is $a_0 e^{-a_0 t} dt$
- Let $u = 1 - e^{-a_0 t}$. Then $u \in [0, 1]$ and $du = a_0 e^{-a_0 t} dt$ $\implies$ u is uniformly distributed in $[0, 1]$.
- $\therefore$ Get time t to next event by sampling u from uniform distribution in $[0, 1]$ and setting $t = \frac{1}{a_0} \ln \frac{1}{1-u}$.

Gillespie Algorithm continued

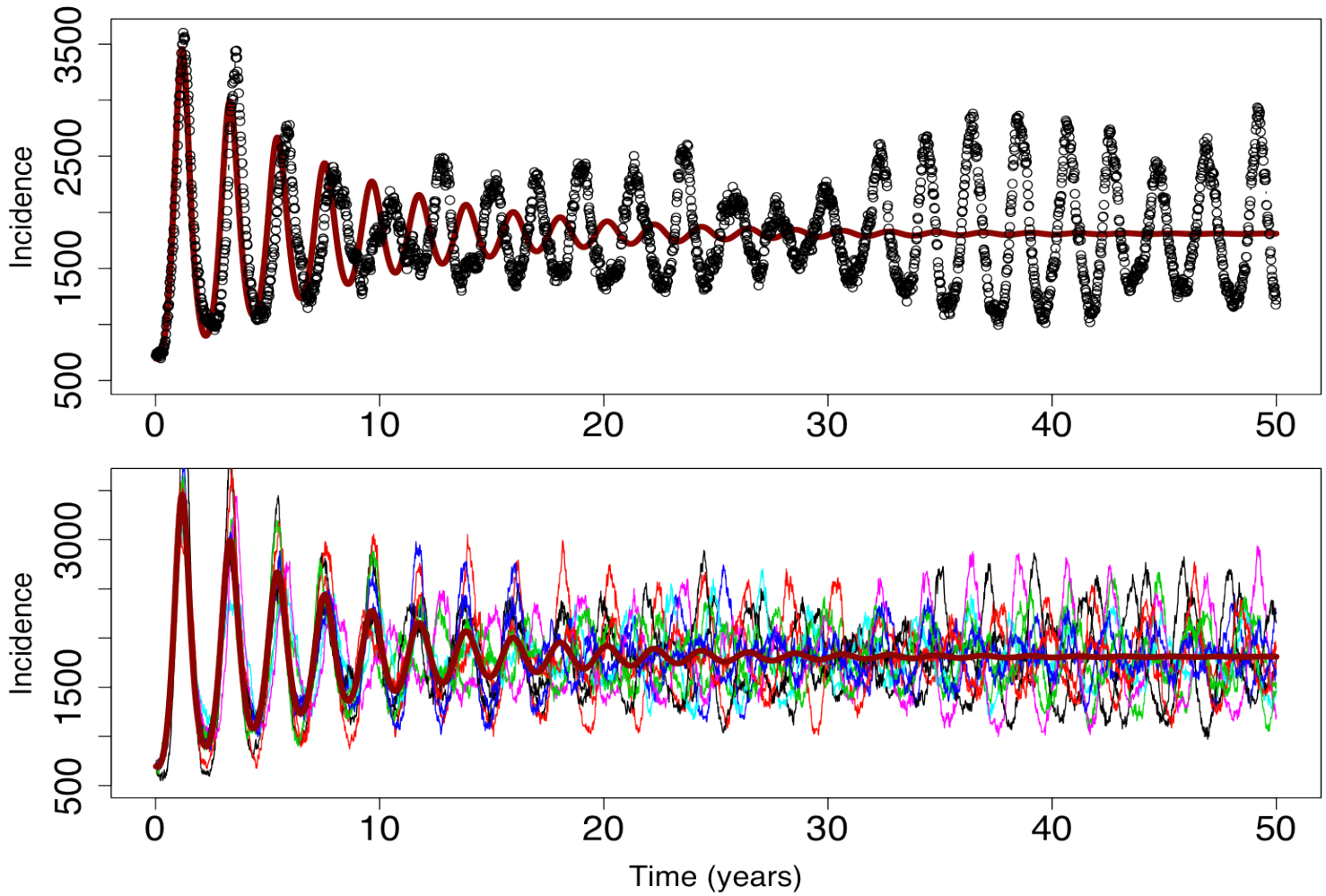
- We now know the time t of the next event, but we must still determine what type of event occurs at time t.
- Probability of event of type i is $\dfrac{a_i}{a_0}$
- $\therefore$ Can easily determine type of event by sampling a point from a uniform distribution on $[0, a_0]$:
  - Event is type i if the uniform deviate lies in the ith interval in the following list:

$$[0, a_1),\ [a_1, a_1 + a_2),\ \ldots,\ [a_1 + \cdots + a_{i-1}, a_1 + \cdots + a_i),\ \ldots$$

- How do realizations of this process differ from the solution of the deterministic (differential equation) model?

Gillespie Simulations: Results for Measles Parameters

$\mathcal{R}_0 = 17$, $T_{\rm lat} = 8$ days, $T_{\rm inf} = 5$ days, $\nu = \mu = 0.02$/year, $N = 5,000,000$



Earn 2009, *IAS/Park City Mathematics Series* **14**, 151–186

Effects of Demographic Stochasticity

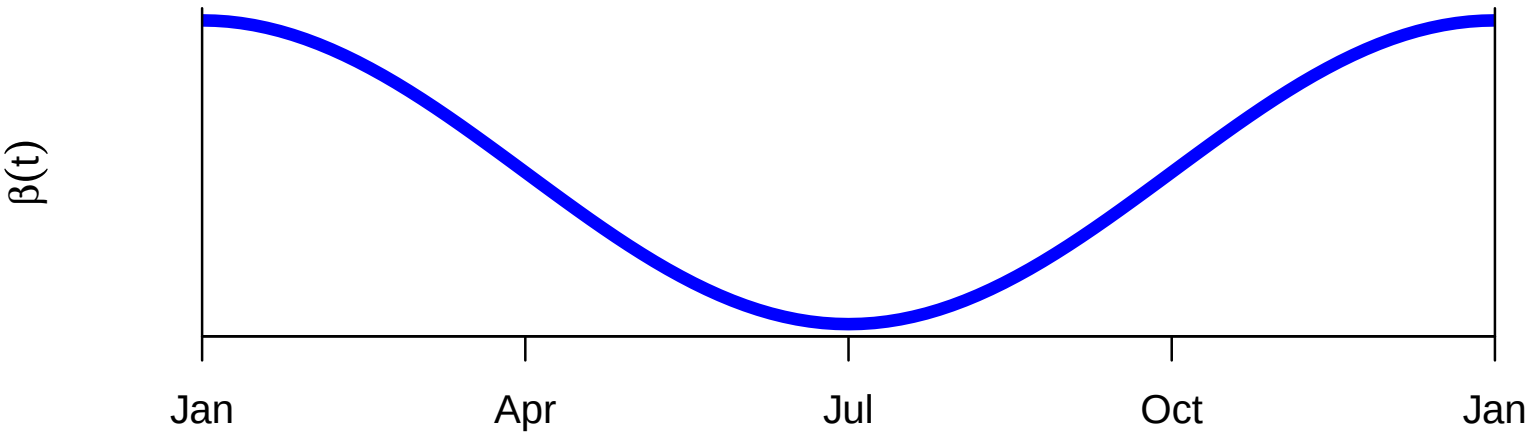
- Sustains transient behaviour (oscillations do not damp out) (Bartlett 1950's)
- Explains undamped oscillations at a single period
- But, unable to explain changes in interepidemic period, or irregularity (common in childhood diseases, *e.g.*, measles, whooping cough, rubella, . . .)
- What other mechanisms might be important?

Transmission rate variation

- Transmission rate β is not constant:
  high during school terms, low in summer
- For simplicity, model as a sine wave:

$$\beta(t) = \langle\beta\rangle \left(1 + \alpha \cos 2\pi t\right)$$

  - $\langle\beta\rangle$ = mean transmission rate
  - α = amplitude of seasonal variation in contact rate



Is this change significant?

- We now have a forced nonlinear system
- Forcing frequency can resonate with the natural timescales of the disease (*e.g.*, damping period)
- Very rich dynamical system...
  (analogy: forced pendulum)

Sinusoidal SEIR Model: Numerical Results

- Stable cycles of various lengths (annual, biennial, 3-year, . . .)
- Multiple co-existing stable cycles
- Chaotic dynamics
- Lots of work on this model in 1980s and 1990s

Smith HL, 1983, *J. Math. Biol.* **17**, 163–177

Schwartz IB, Smith HL, 1983, *J. Math. Biol.* **18**, 233–253

Aron JL, Schwartz IB, 1984, *J. theor. Biol.* **110**, 665-679

Olsen LF, Schaffer WM, 1990, *Science* **249**, 499–504

. . .

Sinusoidal SEIR Model: Rigorous Results

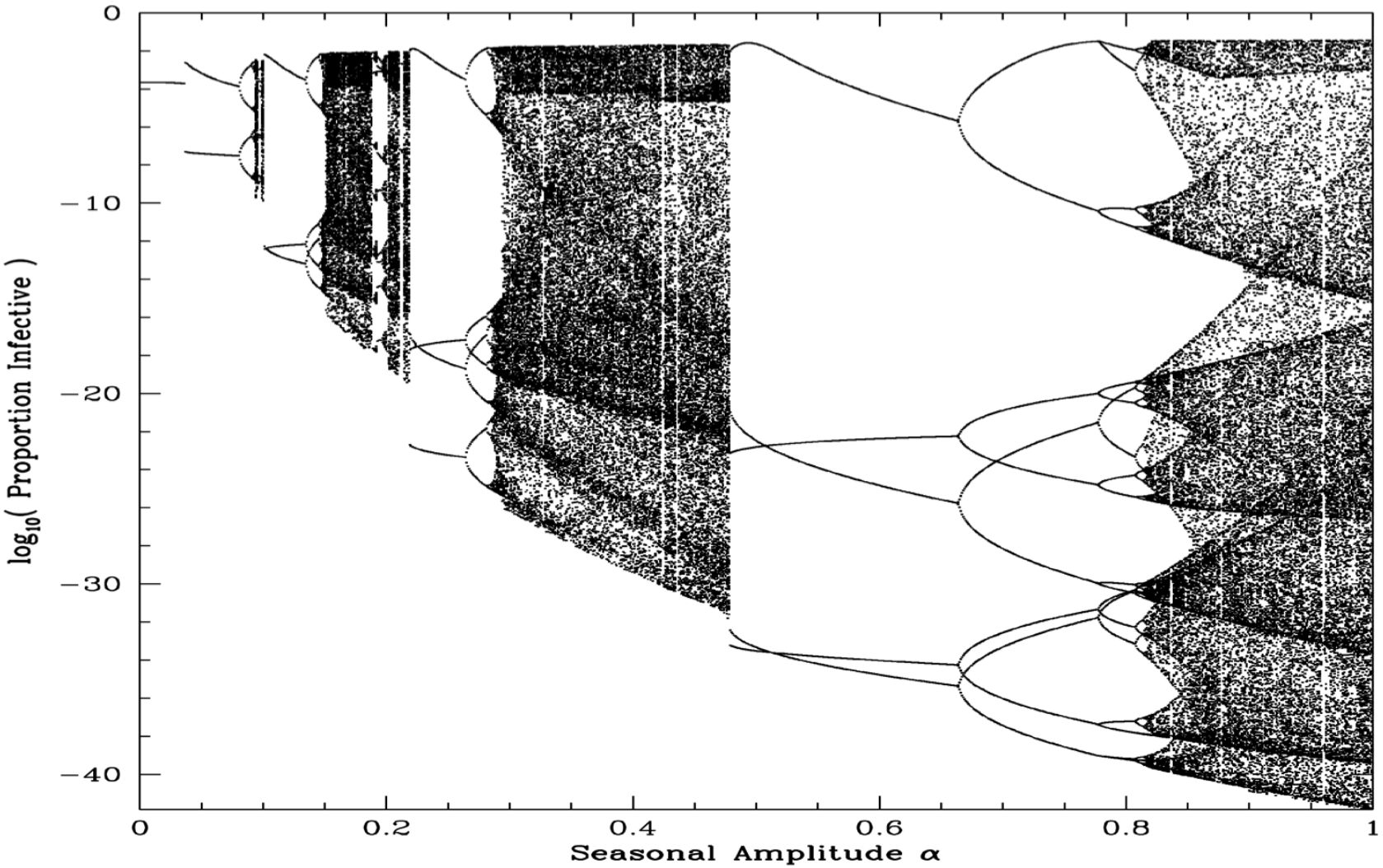
- There exist parameter values such that infinitely many stable cycles co-exist

Schwartz IB, Smith HL, 1983, *J. Math. Biol.* **18**, 233–253

- There exist chaotic repellors (in a modified SEIR model)

Glendinning P, Perry LP, 1997, *J. Math. Biol.* **35**, 359–373

Measles Bifurcation Diagram (Sinusoidal SEIR model)



Earn (2009) *IAS/Park City Mathematics Series* **14**, 151–186

Effects of transmission rate forcing

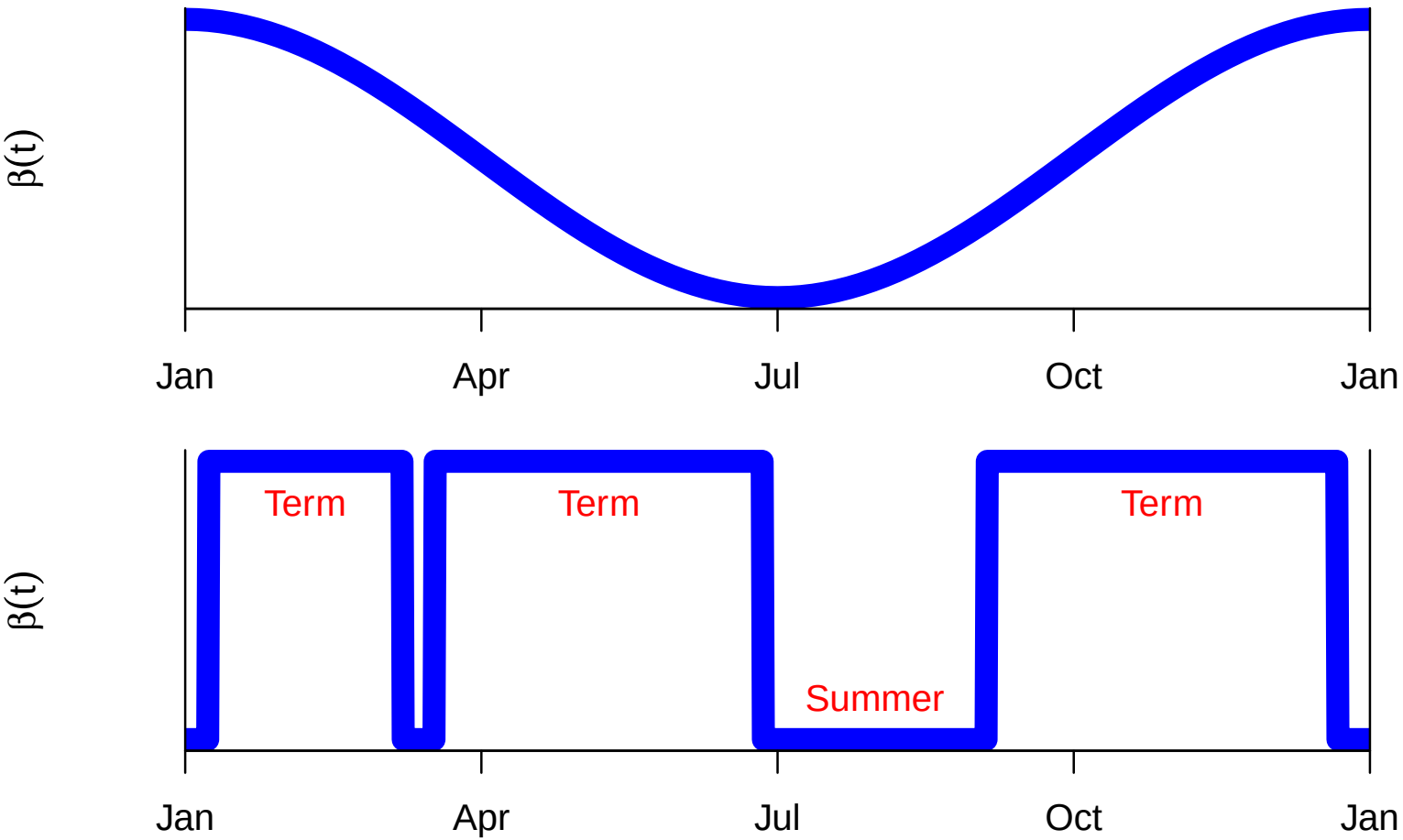
SEIR model with sinusoidal forcing:

- Produces recurrent undamped epidemics of all frequencies observed in measles time series.
- Produces chaos, which can explain irregular behaviour and transitions from one type of cycle to another
  - If correct, this implies these transitions are *unpredictable*.

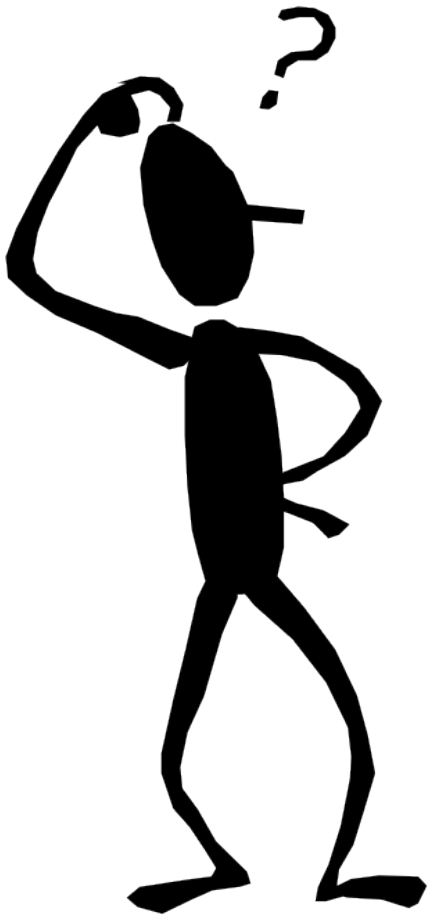
Note: transmission rate (β) might be time-dependent for other reasons, *e.g.*, weather, social distancing, ...

- Functional form of $\beta(t)$ will affect detailed patterns of epidemics

Sinusoidal forcing vs Term-time forcing



What else might affect transmission dynamics?



Is Age Structure Important?

- Real system is not homogeneously mixed
- Contact structure is age-dependent

- Schenzle (1984) argued for creating a Realistically Age-Structured (RAS) SEIR model
  - 21 age classes (0–1, 1–2, . . . , 19–20, > 20)
  - SEIR compartments for each age class
  - Different contact rates between all these age classes

$$\beta(t) \quad \longrightarrow \quad \begin{pmatrix} \beta_{1,1}(t) & \beta_{1,2}(t) & \cdots & \beta_{1,21}(t) \\ \beta_{2,1}(t) & \beta_{2,2}(t) & \cdots & \beta_{2,21}(t) \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{21,1}(t) & \beta_{21,2}(t) & \cdots & \beta_{21,21}(t) \end{pmatrix}$$

Schenzle D (1984) *IMA Journal of Mathematics Applied in Medicine and Biology* **1**, 169–191

- Lots of work on RAS models since Schenzle (1984)

Is Age Structure Important?

- When is this additional structure important?
- *If you have an age-structured question then you need an age-structured model.*
  - *e.g.*, Who should be vaccinated first?
- But not clear the 84 ODEs in Schenzle's model are necessary.
- Fewer age classes $\Longrightarrow$ fewer parameters to estimate

How do we estimate $\mathcal{R}_0$?

- For the basic SIR model, we can just estimate the initial growth rate $(\beta - \gamma)$ and the mean infectious period $(1/\gamma)$, and compute $\mathcal{R}_0 = \beta/\gamma$.
- For the SEIR model with vital dynamics, we also need estimates of
  - mean latent period $(1/\sigma)$
  - birth rate (μ)
- What if our model is much more complicated? (*e.g.*, 84 ODEs!)
- How do we figure this out more generally?

What next?

$\mathcal{R}_0$ or  ?

McMaster University

Mathematics and Statistics

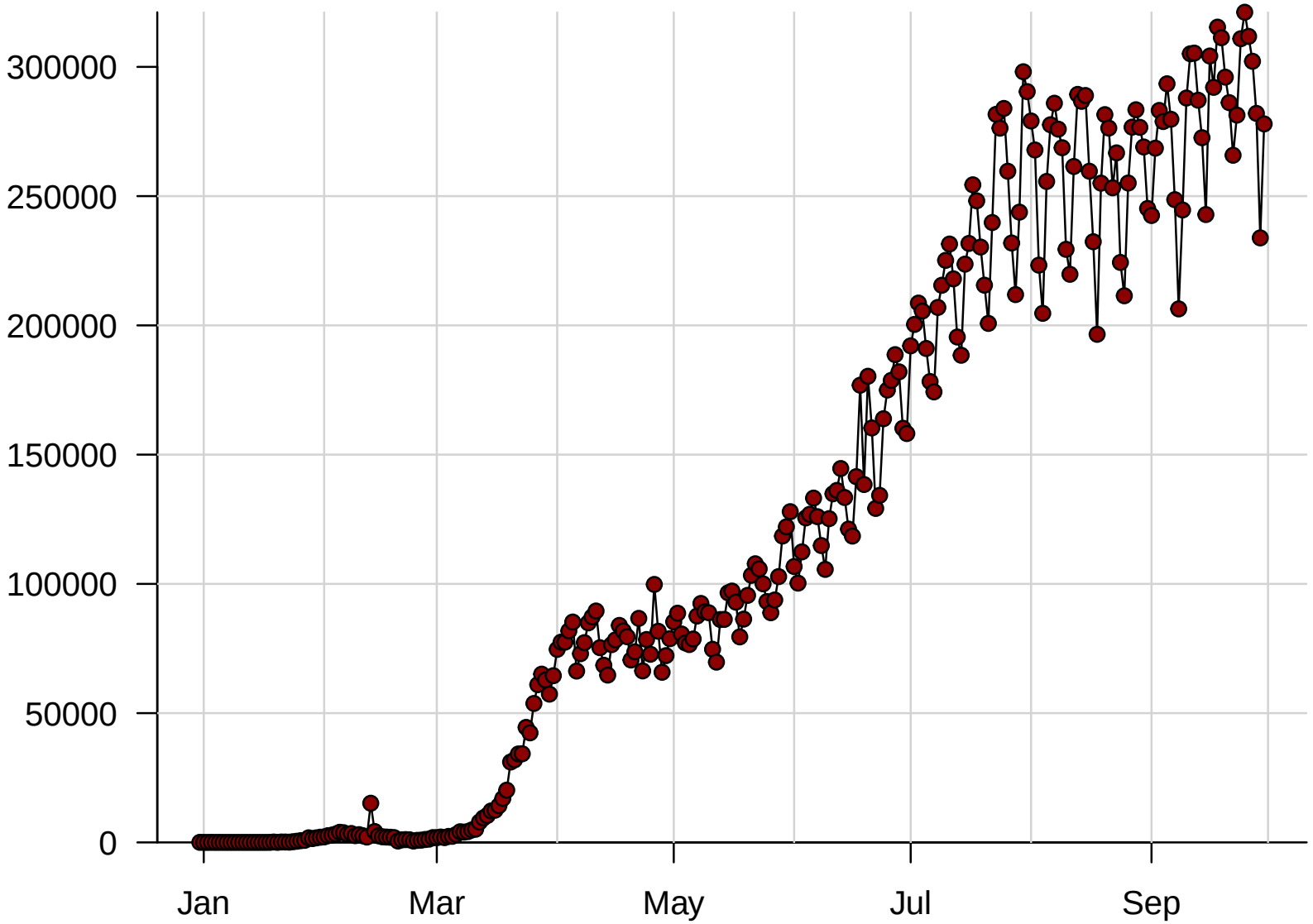
$$\int_M d\omega = \int_{\partial M} \omega$$

Mathematics 747 / 5GT3
Topics in Mathematical Biology

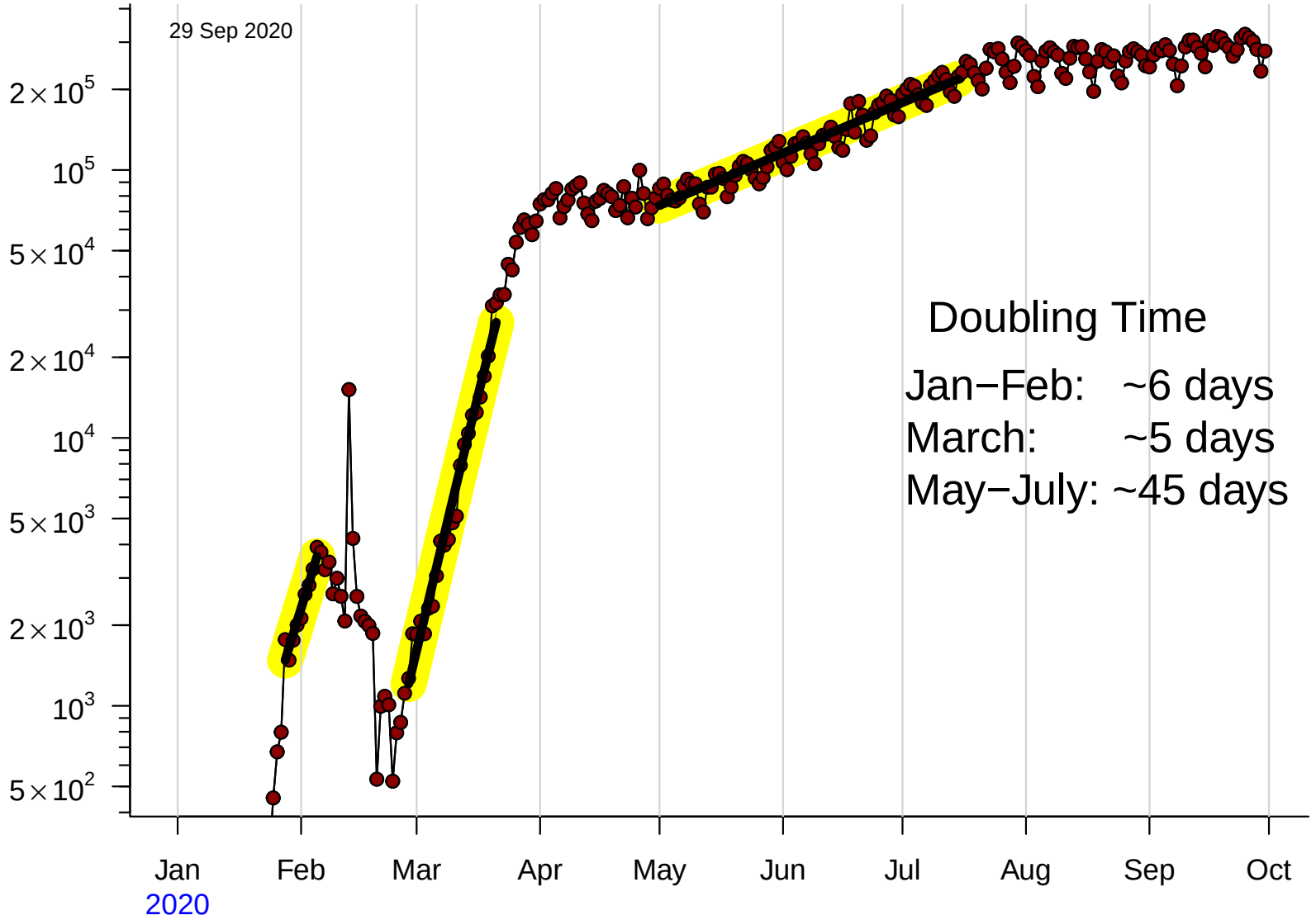
Instructor: David Earn

Lecture 3
Epidemic Modelling Intro 3; $\mathcal{R}_0$
Thursday 1 October 2020

Daily SARS-CoV-2 in 2020 (Worldwide)



Daily SARS-CoV-2 (Worldwide) exponential growth fits



$\mathcal{R}_0$

$\mathcal{R}_0$: biological definition

The ***basic reproduction number*** $\mathcal{R}_0$ is:

the expected number of secondary cases produced, in a completely susceptible population, by a typical infective individual

e.g., Anderson and May (1991) "Infectious Diseases of Humans"

$\mathcal{R}_0$: more mathematical definition

The ***basic reproduction number*** $\mathcal{R}_0$ is:

the number of new infections produced by a typical infective individual in a population at a disease free equilibrium (DFE)

van den Driessche and Watmough (2002) *Mathematical Biosciences* **180**, 29–48

$\mathcal{R}_0$: most mathematical definition

The ***basic reproduction number*** $\mathcal{R}_0$ is:

the spectral radius of the next generation operator at a disease free equilibrium (DFE)

Diekmann, Heesterbeek & Metz (1990) *J. Math. Biol.* **28**, 365–382

Definitions from matrix analysis

Definition (Spectrum of a matrix)

Let M be an $n \times n$ real (or complex) matrix. The ***spectrum of*** M is

$$\sigma(M) = \{\lambda : Mv = \lambda v \text{ for some non-zero } v \in \mathbb{C}^n\},$$

i.e., $\sigma(M)$ is the set of eigenvalues of M.

Definition (Spectral radius of a matrix)

Let M be an $n \times n$ real (or complex) matrix. The ***spectral radius of*** M is

$$\rho(M) = \max\{|\lambda| : \lambda \in \sigma(M)\},$$

i.e., $\rho(M)$ is the maximum modulus of the eigenvalues of M.

Computing $\mathcal{R}_0$

- In very simple models, $\mathcal{R}_0$ is the product of the transmission rate and the mean time in the infectious class. *e.g.*, In the SIR model with vital dynamics,

$$\mathcal{R}_0 = \beta \cdot \frac{1}{\gamma + \mu} .$$

- When there are multiple infected classes, it is more complicated to compute $\mathcal{R}_0$.
- In the SEIR model, we found (based on a biological argument) that

$$\mathcal{R}_0 = \beta \cdot \frac{\sigma}{\sigma + \mu} \cdot \frac{1}{\gamma + \mu} .$$

- Mathematically, the spectral radius of the next generation operator at the DFE is exactly this quantity. With this definition, it is also true that the disease persists if $\mathcal{R}_0 > 1$ and goes extinct if $\mathcal{R}_0 < 1$.

SEIR model (with vital dynamics)

$$\frac{dS}{dt} = \mu N - \frac{\beta SI}{N} - \mu S$$

$$\frac{dE}{dt} = \frac{\beta SI}{N} - \sigma E - \mu E$$

$$\frac{dI}{dt} = \sigma E - \gamma I - \mu I$$

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

- Birth and death rate (μ)
- Transmission rate (β)
- Mean latent period ($1/\sigma$)
- Mean infectious period ($1/\gamma$)

Next generation matrix for the SEIR model

- Consider flows in and out of the infected compartments, and highlight flows that correspond to new infections:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} E \\ I \end{pmatrix} = \begin{pmatrix} \beta SI - \sigma E - \mu E \\ \sigma E - \gamma I - \mu I \end{pmatrix}$$

- $\mathcal{F} =$ inflow of new infecteds to infected compartments $= \begin{pmatrix} \beta SI \\ 0 \end{pmatrix}$
- $\mathcal{V} =$ outflow from infected compartments minus inflow of *non-new* infecteds $= \begin{pmatrix} \sigma E + \mu E \\ -\sigma E + \gamma I + \mu I \end{pmatrix}$
- Let $F =$ linearization of $\mathcal{F}$ at DFE
- Let $V =$ linearization of $\mathcal{V}$ at DFE
- Then the ***next generation matrix*** is FV^{-1}
- Analogous to $\beta\gamma^{-1}$ in simple case.

Interpretation of FV^{-1} as next generation matrix

Almost verbatim from p. 33 of van den Driessche and Watmough (2002) *Mathematical Biosciences* **180**, 29–48

- To interpret the entries of FV^{-1} and develop a meaningful definition of $\mathcal{R}_0$, consider the fate of an infected individual introduced into compartment k of a disease free population.
- The (j,k) entry of V^{-1} is the average length of time this individual spends in compartment j during its lifetime, assuming that the population remains near the DFE and barring reinfection.
- The (i,j) entry of F is the rate at which infected individuals in compartment j produce new infections in compartment i.
- Hence, the (i,k) entry of the product FV^{-1} is the expected number of new infections in compartment i produced by the infected individual originally introduced into compartment k.
- Following Diekmann et al. (1990), we call FV^{-1} the next generation matrix for the model and set

$$\mathcal{R}_0 = \rho(FV^{-1}),$$

where $\rho(A)$ denotes the spectral radius of a matrix A.

$\mathcal{R}_0$ via FV^{-1} for the SEIR model

$$\mathcal{F} = \begin{pmatrix} \beta SI \\ 0 \end{pmatrix} \qquad \mathcal{V} = \begin{pmatrix} \sigma E + \mu E \\ -\sigma E + \gamma I + \mu I \end{pmatrix}$$

$$F = \begin{pmatrix} 0 & \beta \\ 0 & 0 \end{pmatrix} \qquad V = \begin{pmatrix} (\sigma+\mu) & 0 \\ -\sigma & (\gamma+\mu) \end{pmatrix}$$

$$V^{-1} = \begin{pmatrix} \frac{1}{\sigma+\mu} & 0 \\ \frac{\sigma}{(\sigma+\mu)(\gamma+\mu)} & \frac{1}{\gamma+\mu} \end{pmatrix} \implies FV^{-1} = \begin{pmatrix} \frac{\beta\sigma}{(\sigma+\mu)(\gamma+\mu)} & \frac{\beta}{\gamma+\mu} \\ 0 & 0 \end{pmatrix}$$

$$\mathcal{R}_0 = \rho(FV^{-1}) = \beta\sigma/(\sigma+\mu)(\gamma+\mu)$$

- Note wrt previous slide that the $(2,1)$ entry of V^{-1} is the average time an individual who enters the E compartment spends in the I compartment: only a proportion $\sigma/(\sigma+\mu)$ of such individuals make it to the I compartment, where the average time spent—by individuals who get there—is $1/(\gamma+\mu)$.

Computing $\mathcal{R}_0$ for other compartmental ODE models

- The method applied in the previous slides to obtain $\mathcal{R}_0$ for the SEIR model works more generally for a very large class of "reasonable" infectious disease ODE models. "Reasonable" means:
  1. The vector field can be written $\mathcal{F} - \mathcal{V}$, where $\mathcal{F} \geq 0$ corresponds to *new infections* and $\mathcal{V}$ can be written $\mathcal{V} = \mathcal{V}^+ - \mathcal{V}^-$, where $\mathcal{V}^+ \geq 0$ corresponds to *outflow* and $\mathcal{V}^- \geq 0$ corresponds to *inflow of infectives that are not new*.
  2. The biologically relevant part of the state space is forward-invariant. In particular, if a compartment is empty, then there can be no transfer of individuals out of the compartment by death, infection, nor any other means.
  3. The DFE is stable in the absence of new infection (if there is more than one DFE, $\mathcal{R}_0$ may depend on which one we focus on).
  4. The population size N is constant (or the model is expressed in terms of proportions in each compartment).

Computing $\mathcal{R}_0$ for other compartmental ODE models

Theorem (van den Driessche and Watmough (2002))

If the vector field associated with an ODE infectious disease model satisfies the conditions specified on the previous slide, then

1. *$\mathcal{R}_0$ can be computed as $\rho(FV^{-1})$.*

If, moreover, zero is a simple eigenvalue of the Jacobian matrix of the vector field at the disease-free equilibrium (DFE) when $\mathcal{R}_0 = 1$, then

2. *if $\mathcal{R}_0 < 1$ then the DFE is locally asymptotically stable (LAS), whereas if $\mathcal{R}_0 > 1$ then there is a LAS endemic equilibrium (EE).*

Note: For the SIR model, the eigenvalues of the Jacobian at $(S, I) = (1, 0)$ are $-\mu$ and $\beta - (\gamma + \mu)$, which are both zero if $\mu = 0$ and $\mathcal{R}_0 = 1$.

$\mathcal{R}_0$ calculation: summary

- The biological method of deriving $\mathcal{R}_0$ is generally more informative in terms of what is going on. But it can be challenging to apply to complex models.
- The formal approach, *i.e.*, $\mathcal{R}_0 = \rho(FV^{-1})$, works in almost any situation you will encounter, even very complicated models with many compartments.
- If possible, it is best to use both methods to find an expression for $\mathcal{R}_0$, and make sure they agree.
- A completely different challenge is to estimate $\mathcal{R}_0$ for a real epidemic from data...

Estimating $\mathcal{R}_0$ based on the SEIR model

- If the SEIR model captures the natural history of some disease well, how can you *estimate* $\mathcal{R}_0 = \dfrac{\beta\sigma}{(\sigma+\mu)(\gamma+\mu)}$?
  - Mean latent period $1/\sigma$
  - Mean infectious period $1/\gamma$
  - Birth rate μ
  - Estimate β via initial growth rate r:
    - For the simplest SIR model, $r = \beta - \gamma$ so $\beta = r + \gamma$.
    - More generally, r is the largest positive (or least negative) real part of the eigenvalues of $F - V$.
    - For SEIR model we find:

$$r = \frac{1}{2}\left(\sqrt{4\beta\sigma + (\gamma-\sigma)^2} - (\gamma+\sigma+2\mu)\right)$$

    - Solving this for β we obtain: $\beta = \dfrac{(r+\sigma+\mu)(r+\gamma+\mu)}{\sigma}$

Estimating $\mathcal{R}_0$ directly from epidemic data

So far, our approach to estimating $\mathcal{R}_0$ has been:

- specify an epidemiological model, *e.g.*, SIR, SEIR, *etc.*
- estimate the initial exponential growth rate r
- estimate other model parameters via stage duration distributions (latent period, infectious period, ...)
  - can estimate these by studying course of infection in many individuals
- use expression for $\mathcal{R}_0$ in terms of other parameters

Can we avoid committing to a specific epidemic model?

- Yes, using contact tracing data (if available!)

From r to $\mathcal{R}_0$ via generation interval (GI) distribution

- The ***generation interval*** (GI) is the difference between the time when an individual is infected by an infector and the time when this infector was infected.

  Champredon and Dushoff 2015, *Proc. R. Soc. B* **282**:20152026

- The distribution of the GI (denoted g) depends on the natural history of infection (*e.g.*, latent period distribution, infectious period distribution, ...).
- There is a very general relationship between the initial growth rate r, the GI distribution g, and the basic reproduction number $\mathcal{R}_0$.

  Wallinga and Lipsitch 2007, *Proc. R. Soc. B* **274**:599–604

From r to $\mathcal{R}_0$ via generation interval (GI) distribution

During initial growth phase, incidence $i(t) \sim e^{rt}$, so the renewal equation implies

$$\frac{1}{\mathcal{R}_0} = \int_0^\infty e^{-rs} g(s)\,\mathrm{d}s$$

Wallinga and Lipsitch 2007, *Proc. R. Soc. B* **274**:599–604

- $1/\mathcal{R}_0$ is Laplace transform of GI distribution $g(t)$
- If we can estimate r and $g(t)$ then we can estimate $\mathcal{R}_0$
- Estimating the GI distribution $g(t)$ is tricky

Champredon and Dushoff 2015, *Proc. R. Soc. B* **282**:20152026

Park *et al* 2020, *medRxiv* 10.1101/2020.06.04.20122713v1