

# 1 Epidemic Modelling Intro



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](mailto:earn@math.mcmaster.ca)

# Software

- **ASAP**, install the software discussed on the [Software page](#) on the [course web site](#):

- L<sup>A</sup>T<sub>E</sub>X



- R



- RStudio



- Emacs



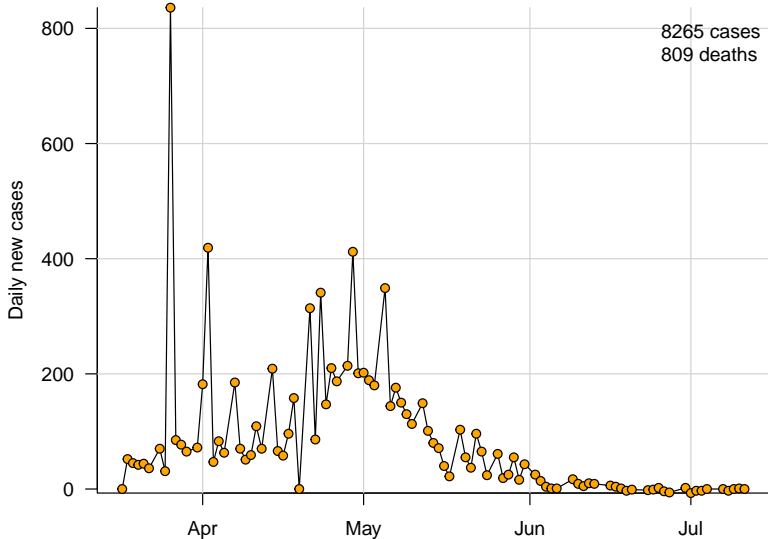
- *Note*: the [Software page](#) also contains some info about spell-checking and counting words in L<sup>A</sup>T<sub>E</sub>X documents.

# Attendance

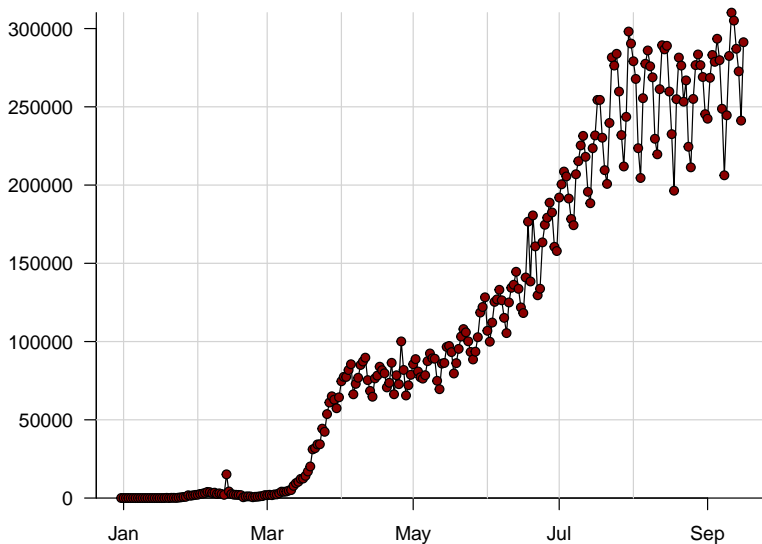
Who is here?

# Epidemic Modelling

# Daily SARS-CoV-1 in 2003 (Worldwide)

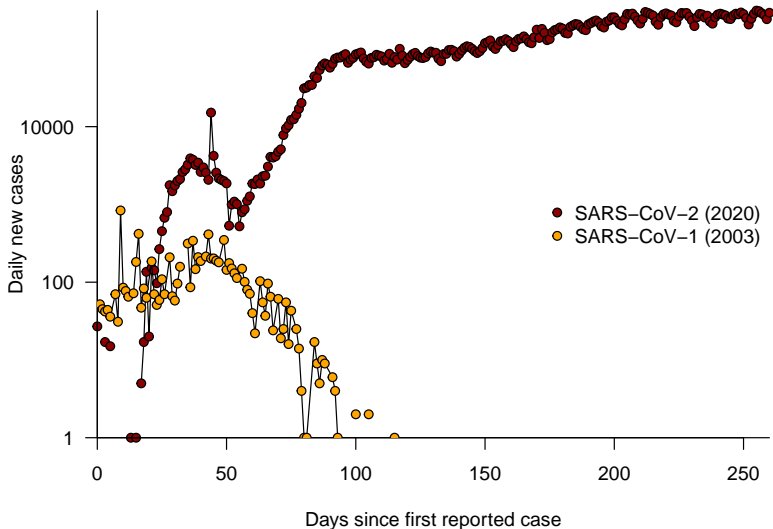


# Daily SARS-CoV-2 in 2020 (Worldwide)

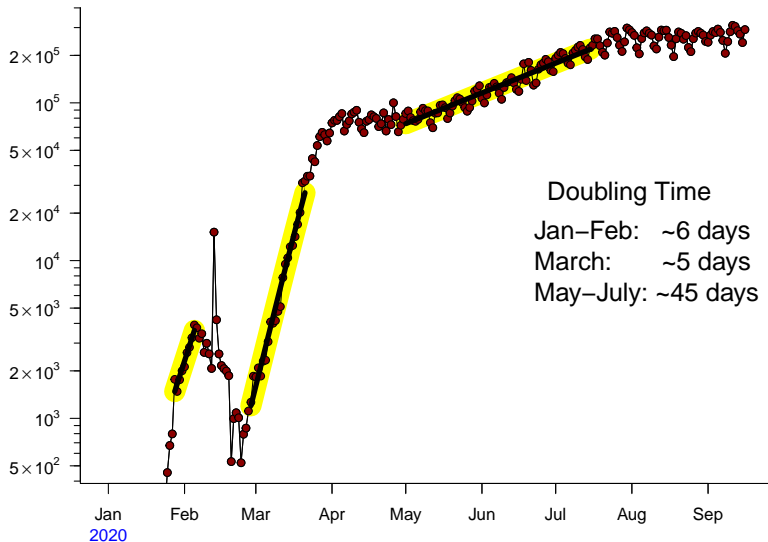




# Daily SARS-CoV-1 vs SARS-CoV-2 (Worldwide)

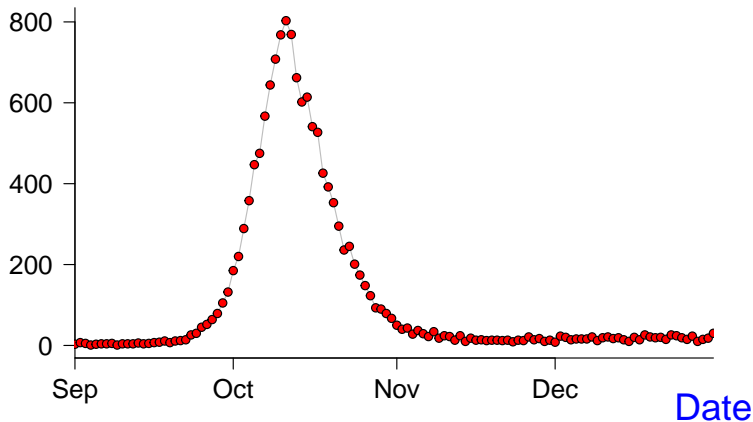


# Daily SARS-CoV-2 (Worldwide) exponential growth fits



# Pneumonia & Influenza Mortality, Philadelphia, 1918

## P&I Deaths



# 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  $\Delta t \rightarrow 0$ , we have

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

# Beware: implicit assumptions that should be explicit

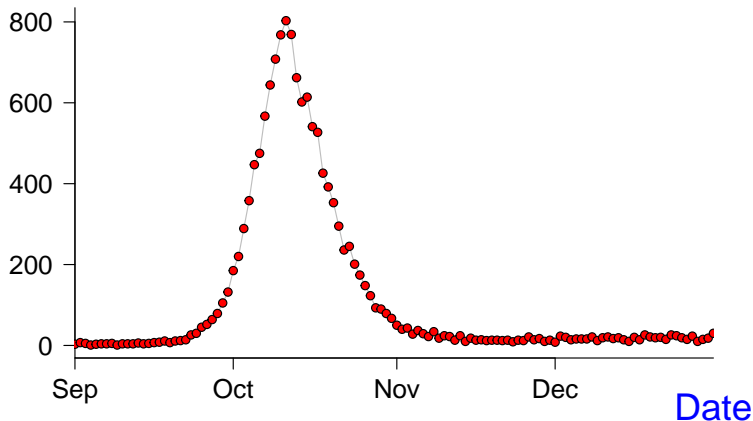
- Ignored discrete nature of individuals when taking limit.
- *Ignored finite infectious periods!*
  - Sometimes it isn't obvious that we've made some assumptions until after we see what the model predicts.

# 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.

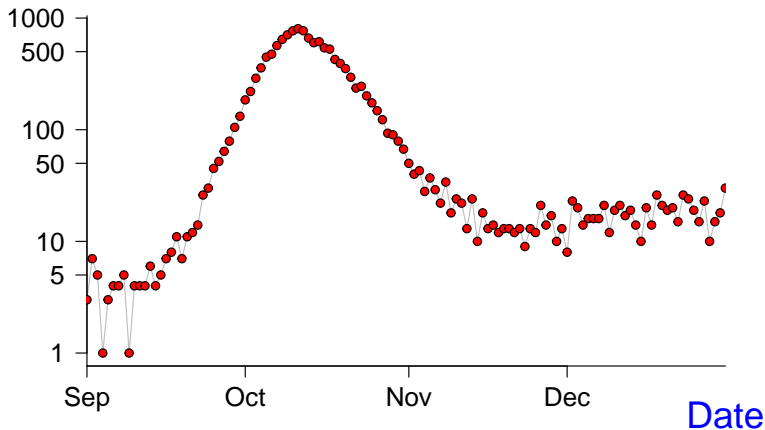
# Original data: P&I Mortality, Philadelphia, 1918

## P&I Deaths



# Logarithmic scale: P&I Mortality, Philadelphia, 1918

## P&I Deaths



# Parameter estimation

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

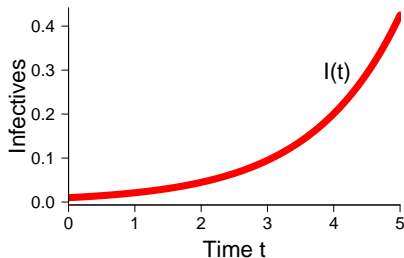
- Fit a straight line through the part of the **logarithmic mortality curve** that looks straight.
- The slope of the line is  $B$ .
- The “intercept” is  $\log I_0$ .
  - “Intercept” in quotes because we need to define  $t = 0$  as the time when exponential growth begins.
- **Note:** Parameter estimation 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  $B = \text{constant}$ .  
Is  $B$  really constant?
- $B$  depends on how many **susceptibles** there are.
- $B = \beta S(t)$
- $\beta$  = average number of contacts between **susceptibles** and **infectives** that lead to a new **infective**  
*per unit time*  
*per infective*  
*per susceptible*
- $\beta$  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:

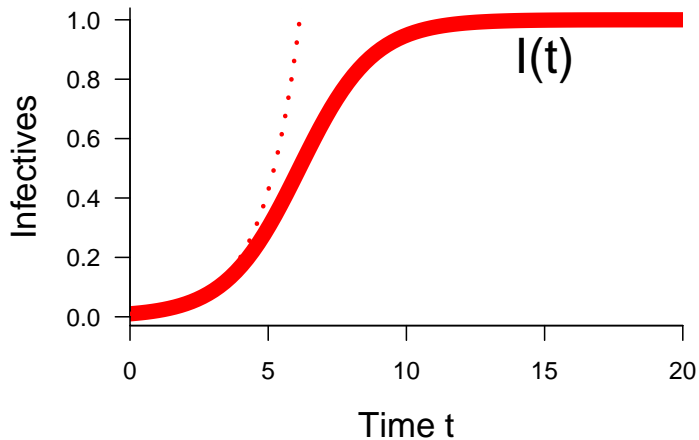
$$\frac{dS}{dt} = -\beta S(t)I(t)$$

- But  $S(t) = N - I(t) \implies I(t)$  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

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

- 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  $\beta$ :
  - 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  $\gamma$  = rate of removal from the **infective** class (via recovery or death)

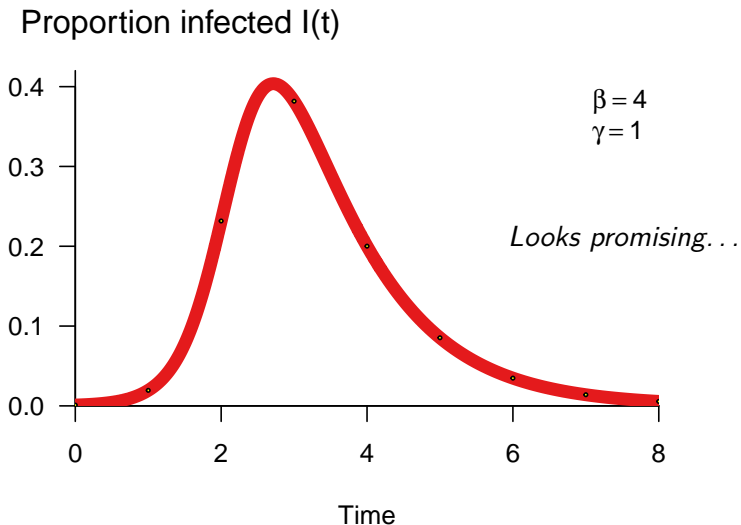
$$\frac{dS}{dt} = -\beta SI$$

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

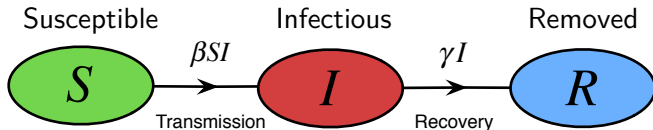
$$\frac{dR}{dt} = \gamma I$$

- 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



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

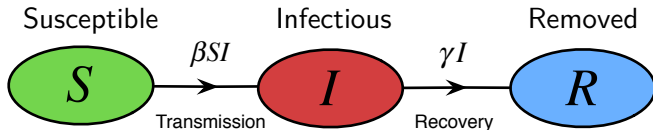
$$\frac{dI}{dt} = \beta SI - \gamma I$$

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

## ■ Parameters:

- Transmission rate  $\beta$
- Recovery rate  $\gamma$   
(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  $I \ll 1$ ):

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

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



# 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

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

- 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  $p(t) = C'(t) = \frac{d}{dt}(1 - e^{-\gamma t}) = \gamma e^{-\gamma t}$
- 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 $\mathcal{R}_0$

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

- $\mathcal{R}_0$  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?
  - $\frac{dI}{dt} = \beta SI - \gamma I = (\mathcal{R}_0 S - 1)\gamma I$
  - $\therefore \mathcal{R}_0 \leq 1 \implies \frac{dI}{dt} \leq 0$  for all  $(S, I) \in [0, 1]^2 \implies$  no growth

# The SIR model: Analysis (biological well-posedness)

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

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

- 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  
 $\implies S(t) + I(t) \leq S(0) + I(0) \leq 1$ .

**Be careful:  
Is this a sensible  
biological  
model?**

# 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:

$$\blacksquare DF_{(S,I)} = \begin{pmatrix} -\beta I & -\beta S \\ \beta I & \beta S - \gamma \end{pmatrix}$$

$$\blacksquare DF_{(S_0,0)} = \begin{pmatrix} 0 & -\beta S_0 \\ 0 & \beta S_0 - \gamma \end{pmatrix}$$

- All equilibria are *non-hyperbolic*.

## ■ Periodic orbits:

- $(S + I)' = -\gamma I$   
 $\implies$  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

## Nullclines:

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

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

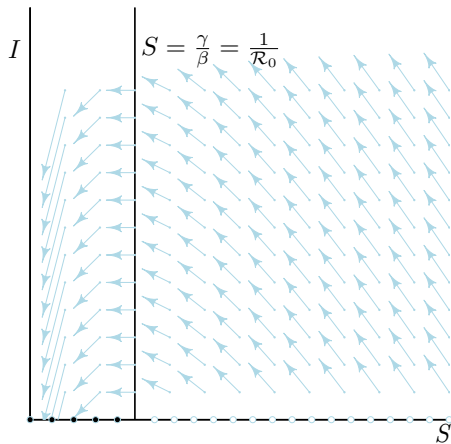
- $S' = 0 \implies S = 0$  or  $I = 0$ 
  - $S$  nullclines: both coordinate axes
  
- $I' = 0 \implies I = 0$  or  $S = \gamma/\beta$ 
  - $I$  nullclines:  $S$  axis and vertical line at  $S = 1/\mathcal{R}_0$
  - Is the  $I$  nullcline at  $S = 1/\mathcal{R}_0$  always relevant?
    - If, and only 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

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

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

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





# The SIR model: Analysis

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

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

- *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  $S$ :

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

## Phase Portrait:

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

- *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:**

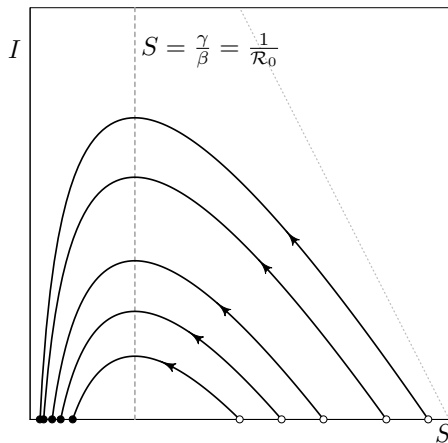
$$\frac{dS}{dt} = -\beta SI$$

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

**Solution Curves in  
Phase Plane:**

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

**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:

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

## Final Size of Epidemic:

- As  $t \rightarrow \infty$  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  $I_0 \rightarrow 0$ , 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.,

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

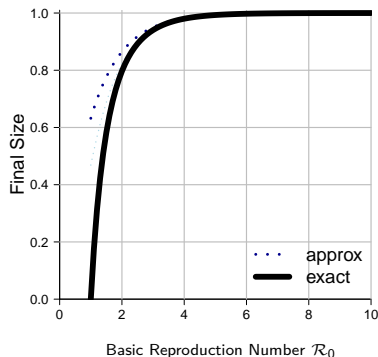
- This is a famous formula, derived by Kermack and McKendrick in 1927.

# 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?
- $\sim 60\text{--}80\%$

# 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  $\mathcal{R}_0$ :

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

- 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  $\beta/\gamma$  because it has a natural interpretation.
- But we are still left with  $\gamma$  as a second parameter.
- Can we simplify the model somehow?
- $\gamma$  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.

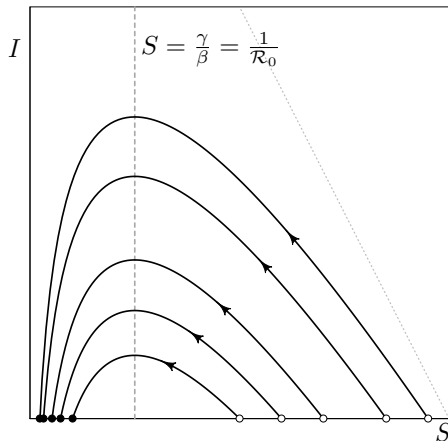
# The SIR model: Results so far

## 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

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

- $\therefore$  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:

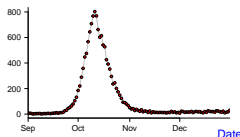
- $\mathcal{R}_0$  is extremely important to estimate in practice!
- Epidemic occurs if and only if  $\mathcal{R}_0 > 1$ .
- Single epidemic, then disease disappears.
- Can prevent epidemic by vaccinating (or otherwise removing) a proportion  $1 - \frac{1}{\mathcal{R}_0}$  from the transmission process.

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?

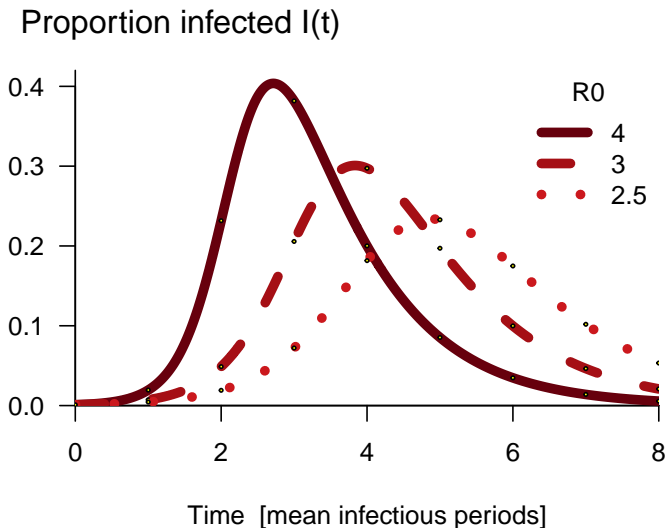
## What about 1918 flu in Philadelphia?

P&I Deaths



- 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 $\mathcal{R}_0$



CPU time: 0.067S, Vector field evaluations: 1944, Ratio: 29014.9

# 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,

$$\frac{dS}{dt} = -i(t), \quad (1a)$$

$$i(t) = \mathcal{R}_0 S(t) \int_0^\infty i(t-s) g(s) ds, \quad (1b)$$

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](#).