Data driven approaches to estimating R_0 and R_t

Epid 814 – Fall 2021

Recap: R_0 and R_t

- R_0 is the average number of new infections generated by a single infectious individual in a fully susceptible population (typically assumes “baseline” behaviors patterns etc.)

- R_t is the time varying version of R_0—how many infections are generated by each infectious individual as population immunity builds, contact patterns change, etc.

Other equations related to R_0

- There are several common uses of R_0—however, it's important to note that **there are a lot of assumptions underlying the below equations**, and these may not be valid depending on behavior, population density, which model you're using, etc.
- Herd immunity threshold $HIT = 1 - \frac{1}{\mathcal{R}_0}$

  (and many variations of this accounting for eligible population, vaccine effectiveness/efficacy, etc.)
- Epidemic final size (R_∞) equation: $R_\infty = 1 - e^{-\mathcal{R}_0 R_\infty}$ [1]

[1] Miller J. A note on the derivation of epidemic final sizes. https://www.ncbi.nlm.nih.gov/labs/pmc/articles/PMC3506030/

How to estimate R_0 & R_t?

- We often use transmission models
  - Determine the formula for R0 (and Rt) from the next generation matrix
  - Fit the model to incidence data
  - Calculate R0 and Rt from the results (i.e. the R0 equation but without assuming S = N, and potentially with time varying parameters)

- Also use the final size approach, if: a) the epidemic is over so you can see the final size and b) you're using a model for which the final size equation is valid

- But there are also more real-time, data driven methods

Estimating R_0 & R_t : a simple approach

- Suppose we have incidence data
- R0 and Rt tell us how many cases the current generation (e.g. 1 initial case) should generate
    - **But when do those cases happen?**
- The time from infection of person A to infection of person B is called the *generation time*
- The time from onset for person A to onset for person B is called the *serial interval* (or sometimes, test date for person A to test date for person B)

Estimating R_0 & R_t: a simple approach

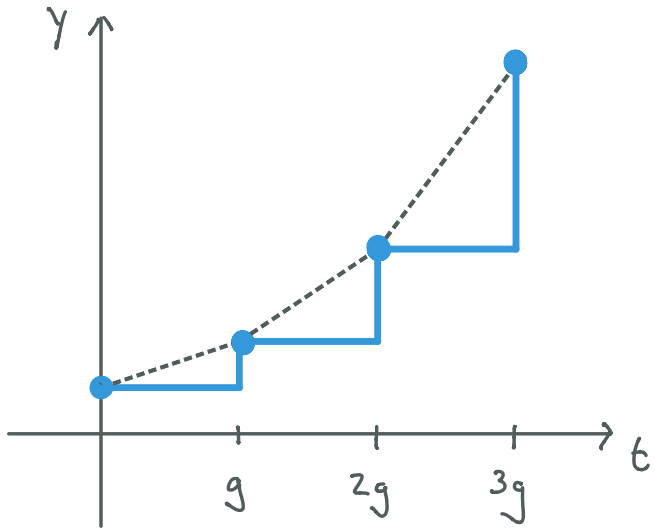
- Suppose we had a single, exact generation time g, e.g. say g = 5 days

- If R_0 or R_t = 3, then each case will generate 3 new cases exactly g = 5 days later

$$y_t = y_{t-g}\mathcal{R}_t$$

where y_t is the number of cases at time *t*.

- Note this is just the discrete equation for exponential growth! (or decay, if Rt<1)



Estimating R_0 & R_t: a simple approach

- With this simple model, we can estimate Rt as:

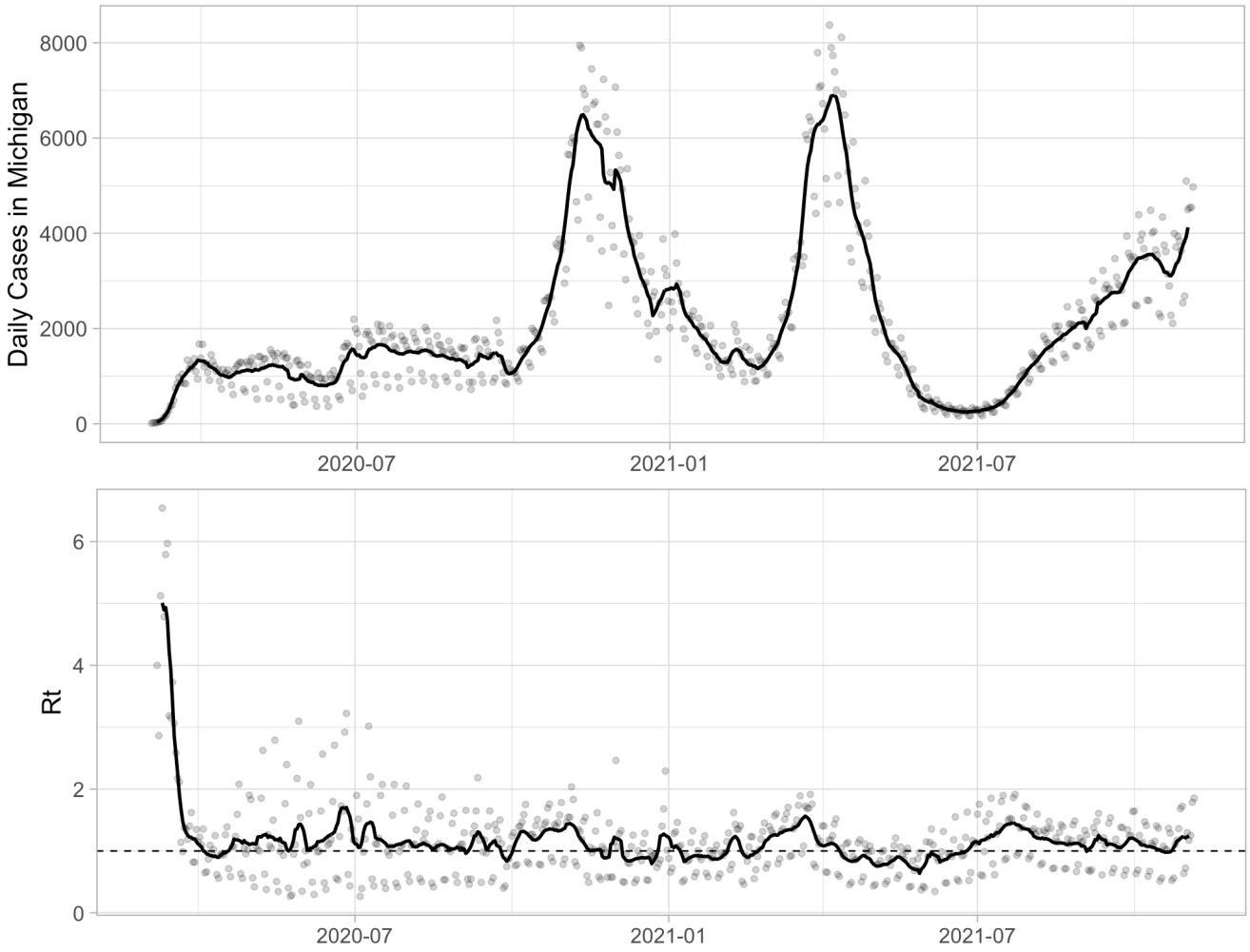
$$\mathcal{R}_t = y_t / y_{t-g}$$

- In other words, we just divide current cases by those from one generation time ago!

- Example: COVID-19

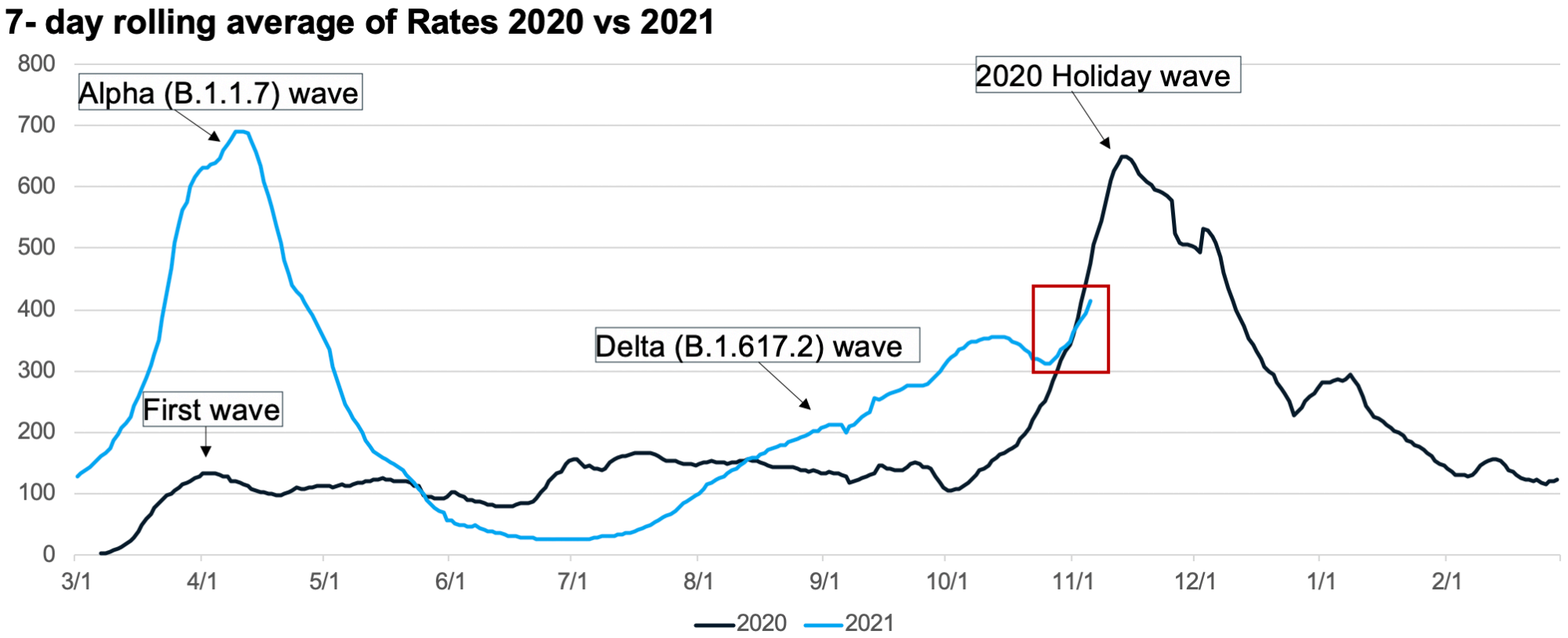
Simple Rt in Michigan

- Assumes a fixed serial interval of 5 days (recent estimates suggest somewhere from 5-6 days [1,2])



[1] 5.3 days https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7448781/, [2] 5.68 days https://wwwnc.cdc.gov/eid/article/27/5/20-4663_article

Rt for Michigan is likely to go up…



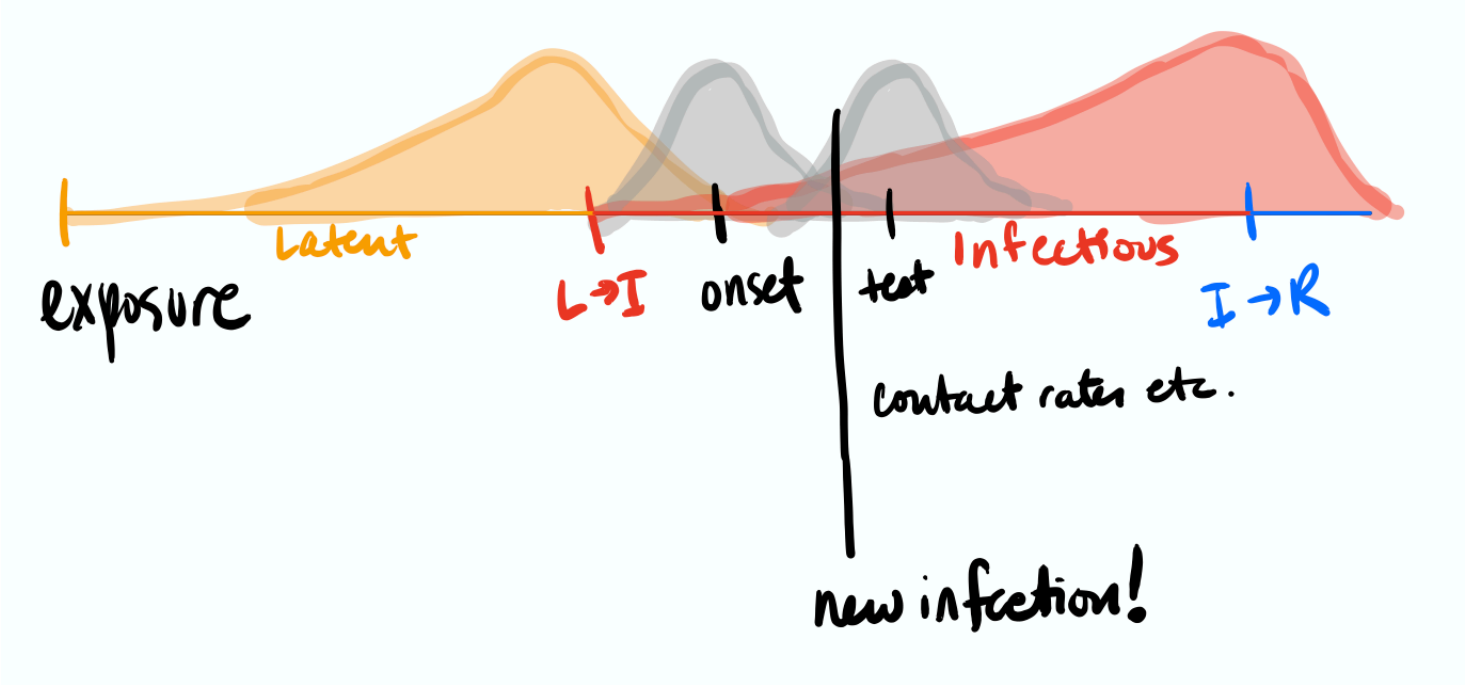
How to make this more realistic?

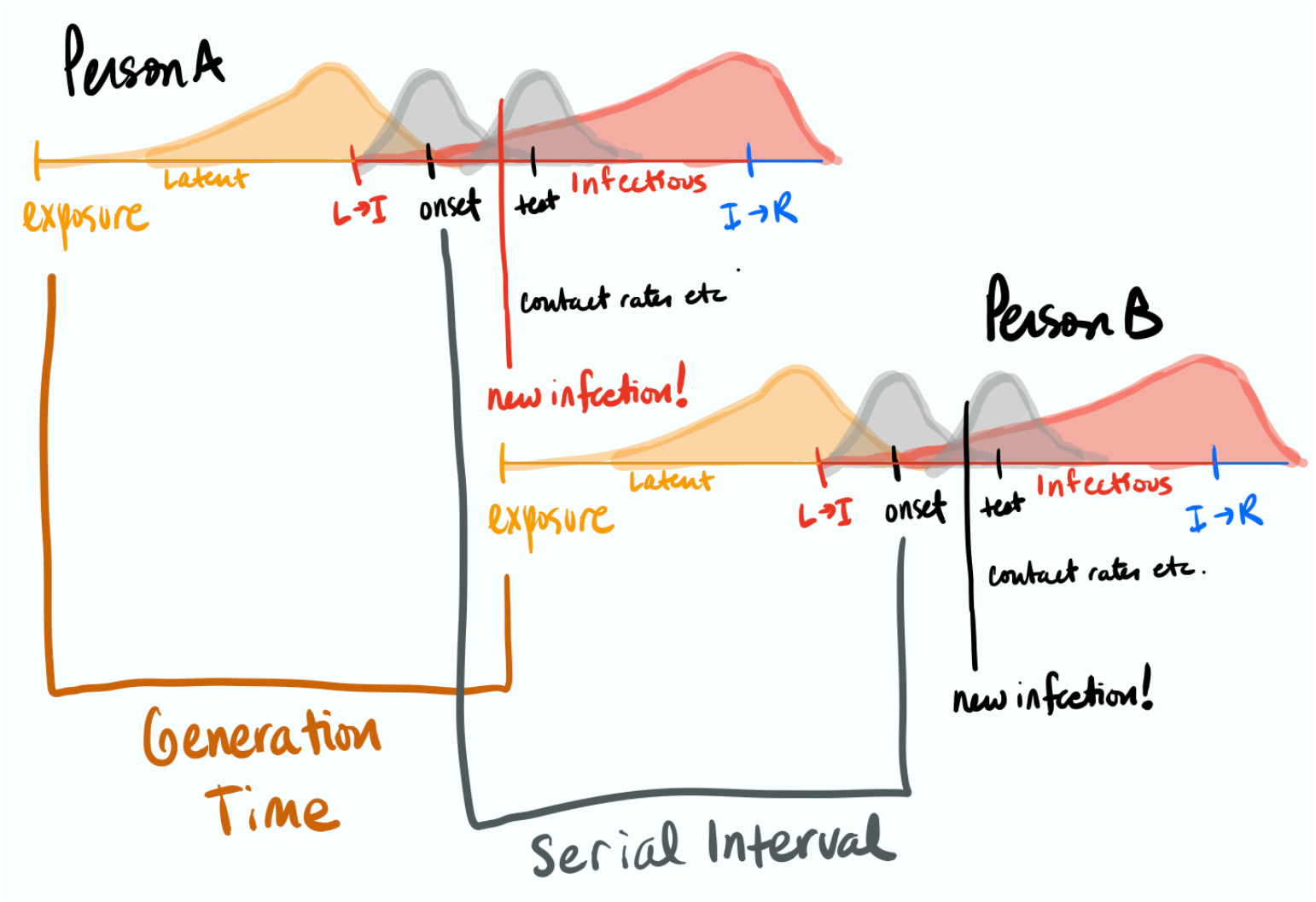
- Even with the simple model
  - You might want to interpolate so generation time can be non-integer valued
  - May not want to assume each data point is exactly correct—e.g. fit a sliding window of constant Rt with some measurement model, or estimate a spline

- But perhaps more importantly:
  - We often can't measure the generation time! Much easier to measure the serial interval
  - **Both serial interval and generation time are actually distributions! (and likely time varying ones at that…)**
  - Serial interval may not exactly equal the generation time! Often similar mean but the distribution can be different

Generation time and serial interval

- Both of these are actually distributions made up of a combination of processes—each with their own distributions!




Person A
Latent
exposure
L→I
onset
test
Infectious
I→R
contact rates etc
new infection!
Person B
Latent
exposure
L→I
onset
test
Infectious
I→R
contact rates etc.
new infection!
Generation Time
Serial Interval

Generation time and serial interval

- And this can get even more complex when you factor in time varying testing/ascertainment, asymptomatic fraction, etc.

- Often can measure the serial interval from contact tracing data, but usually cannot exactly measure the generation time
- Often generation time and serial interval are similar—but can be skewed differently or different variance, etc.

- Many of the data driven models out there are built to model these distributions and account for these different biases & delays
- Note that these biases can both bias the value of Rt but also the timing—can cause delays etc.

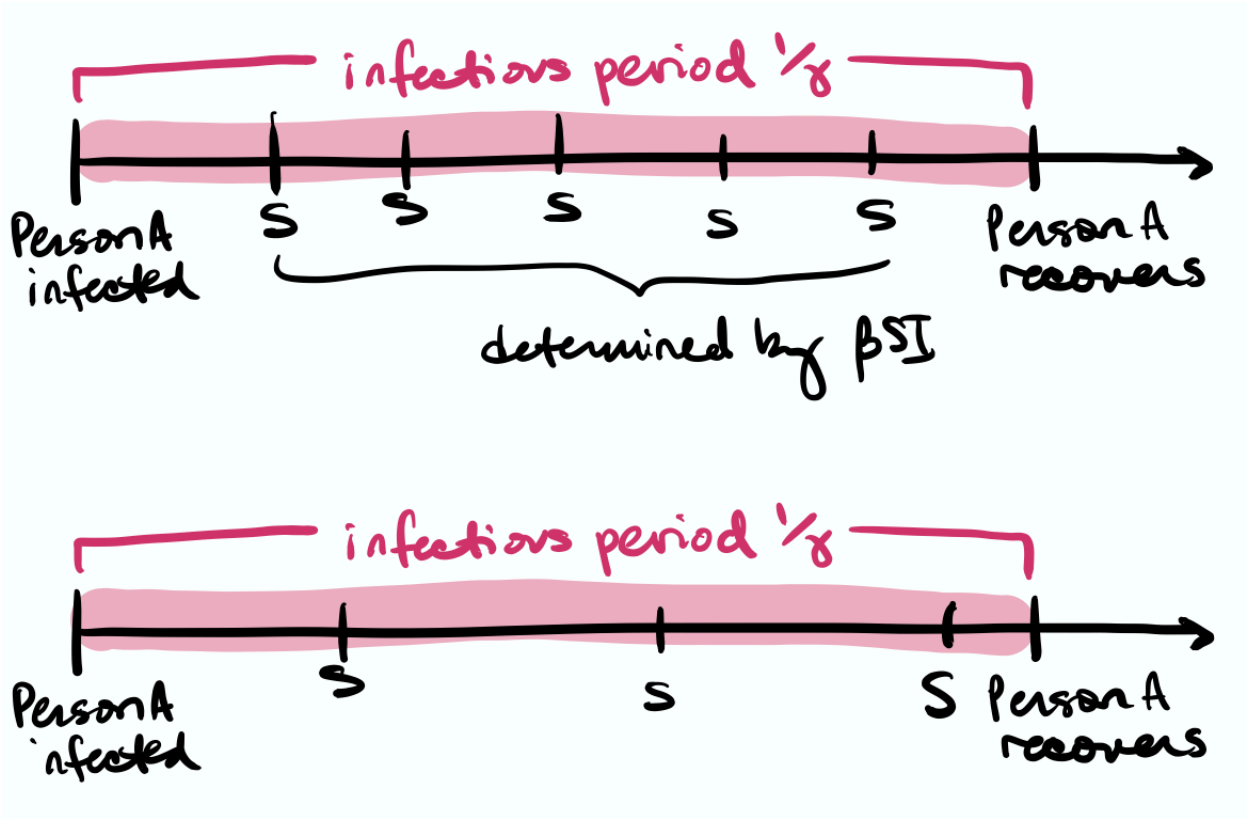
Slightly more realistic example

- Suppose we have a fixed distribution for the generation time or serial interval (and suppose the two are basically equal)

$$y_t = \sum_{i=1}^{n} p(g_i) y_{t-i} \mathcal{R}_t(t-i)$$

- And we can do the same basic approach for estimation (e.g. assuming a spline or a fitted sliding constant window, etc.)

Generation times: intrinsic, backward, & forward



Resources

- COVID Rt dashboards (some of many): https://covidestim.org, https://epiforecasts.io/covid/posts/national/united-states/
- EpiEstim package and paper
- Comparison of methods: Practical considerations for measuring the effective reproductive number, Rt
- Another post with a comparison of methods: https://staff.math.su.se/hoehle/blog/2020/04/15/effectiveR0.html
- Tutorial from rt.live (from back when rt.live was still a thing): https://github.com/rtcovidlive/covid-model/blob/master/tutorial.ipynb
- Menéndez, 2021. A poor-man's approach to the effective reproduction number: the COVID-19 case
- A nice overview: Reproduction number (R) and growth rate (r) of the COVID-19 epidemic in the UK: methods of estimation, data sources, causes of heterogeneity, and use as a guide in policy formulation