DEPARTMENT FOR CONTINUING EDUCATION
UNIVERSITY OF OXFORD

Infectious Disease Modelling: Applied Methods in R

Dr Emma L Davis

Week 2

Week 2: Analysing data in R

- Terminology
- Key epidemiological metrics
  - Growth rate
  - Doubling time
  - The basic reproduction number

Key concepts and terminology

Prevalence: The proportion (or percentage) of a population who have a specific characteristic at a specific point in time.

E.g.

- UK prevalence of flu = Proportion of people in the UK who currently have the flu
- Global prevalence of obesity = Proportion of people in the world who are currently obese
- Prevalence of post-partum depression = Proportion of post-partum women who currently have depression

$$Prevalence = \frac{\#\ of\ people\ in\ sample\ with\ characteristic}{total\ \#\ people\ in\ the\ sample}$$

Key concepts and terminology

Prevalence: The proportion (or percentage) of a population who have a specific characteristic at a specific point in time.

E.g. Prevalence of severe (hospitalized) COVID-19 in the UK:

- Number of people currently hospitalized is: 5,658 (31st Aug 2022, UK Coronavirus Dashboard)
- Number of people in the UK: 68.66 million (5th Sep 2022, Worldometer)

$$Prevalence = \frac{5658}{68660000} = 0.0000824$$

$$\times 100$$

or 0.00824% of people

Key concepts and terminology

Prevalence: The proportion (or percentage) of a population who have a specific characteristic at a specific point in time.

Why might it be difficult to calculate the prevalence of all COVID-19 cases in the UK?

How many people are currently infected?

- When did people test positive?
- When were people infected?
- How long are people infected for?
- How many cases are never detected?

Which of these can we measure?

DEPARTMENT FOR CONTINUING EDUCATION
UNIVERSITY OF OXFORD

Key concepts and terminology

Incidence: The number of new instances of a specific characteristic or event that occur in a population over a specified time period

E.g.

- The annual incidence of knife crime in New York = The number of knife-related crimes in New York each year
- A lifetime incidence of 1 in 10,000 people = 1 in 10,000 people will experience this in their lifetimes
- Weekly incidence of COVID-19 in the UK = number of new COVID-19 cases in a week

$$Incidence = \frac{(\#\ of\ new\ cases\ or\ events\ in\ a\ particular\ time\ frame)\times(desired\ time\ frame)}{(total\ \#\ people\ in\ the\ population)\times(observed\ time\ frame)}$$

Key concepts and terminology

Incidence: The number of new instances of a specific characteristic or event that occur in a population over a specified time period

E.g. Daily incidence of COVID-19 hospitalizations in the UK

- Number of people admitted to hospital with COVID-19: 602 (28th Aug 2022, UK Coronavirus Dashboard)
- Number of people in the UK: 68.66 million (5th Sep 2022, Worldometer)
- Time frame of observations: 1 day
- Time frame of desired incidence: 1 day

$$Incidence = \frac{602 \times 1}{68660000 \times 1} = 0.00000877$$

$\times 100{,}000$

or 0.877 per 100,000 population

Key concepts and terminology

Incidence: The number of new instances of a specific characteristic or event that occur in a population over a specified time period

E.g. What about daily incidence of any COVID-19 cases in the UK?

- Number of new **detected** COVID-19 cases: 2,876 (26th Aug 2022, UK Coronavirus Dashboard)
- Number of people in the UK: 68.66 million (5th Sep 2022, Worldometer)
- Time frame of observations: 1 day
- Time frame of desired incidence: 1 day

$$Incidence = \frac{2876 \times 1}{68660000 \times 1} = 0.0000419$$

$\times 100{,}000$

or 4.19 per 100,000 population

Key concepts and terminology

Incidence: The number of new instances of a specific characteristic or event that occur in a population over a specified time period

E.g. What about **weekly** incidence of any COVID-19 cases in the UK?

- Number of new detected COVID-19 cases: 2,876 (26th Aug 2022, UK Coronavirus Dashboard)
- Number of people in the UK: 68.66 million (5th Sep 2022, Worldometer)
- Time frame of observations: 1 day
- Time frame of desired incidence: 7 days

$$Incidence = \frac{2876 \times 7}{68660000 \times 1} = 0.000303$$

$\times 100{,}000$

or 30.3 per 100,000 population

DEPARTMENT FOR CONTINUING EDUCATION
UNIVERSITY OF OXFORD

Key concepts and terminology

Incidence: The number of new instances of a specific characteristic or event that occur in a population over a specified time period

With more data we can get a more detailed estimation of weekly incidence:

Day	20th Aug	21st Aug	22nd Aug	23rd Aug	24th Aug	25th Aug	26th Aug
# cases	2821	3522	4493	3987	3805	3170	2876

- Number of people in the UK: 68.66 million (5th Sep 2022, Worldometer)
- Time frame of observations: 7 days
- Time frame of desired incidence: 7 days

$$Incidence = \frac{(2821 + 3522 + 4493 + 3987 + 3805 + 3170 + 2876)\times 7}{68660000 \times 7} = 0.000359$$

$\times 100{,}000$

or 35.9 per 100,000 population

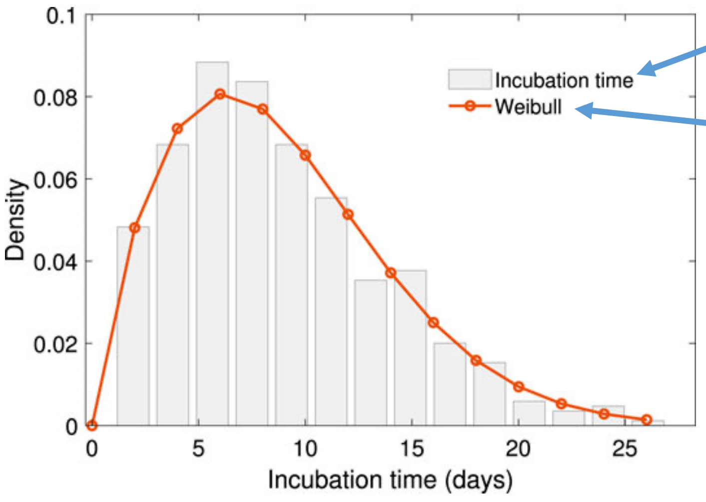
Key concepts and terminology

Incubation period: The time from exposure to a disease (resulting in infection) to the presentation of symptoms

This can be reported in different ways

E.g. COVID-19

- As a range: 2-14 days (Centre for Disease Control, USA)
- As an average (“typical”): 6.9 days (Xin et al. CID 2022)
- As a distribution:

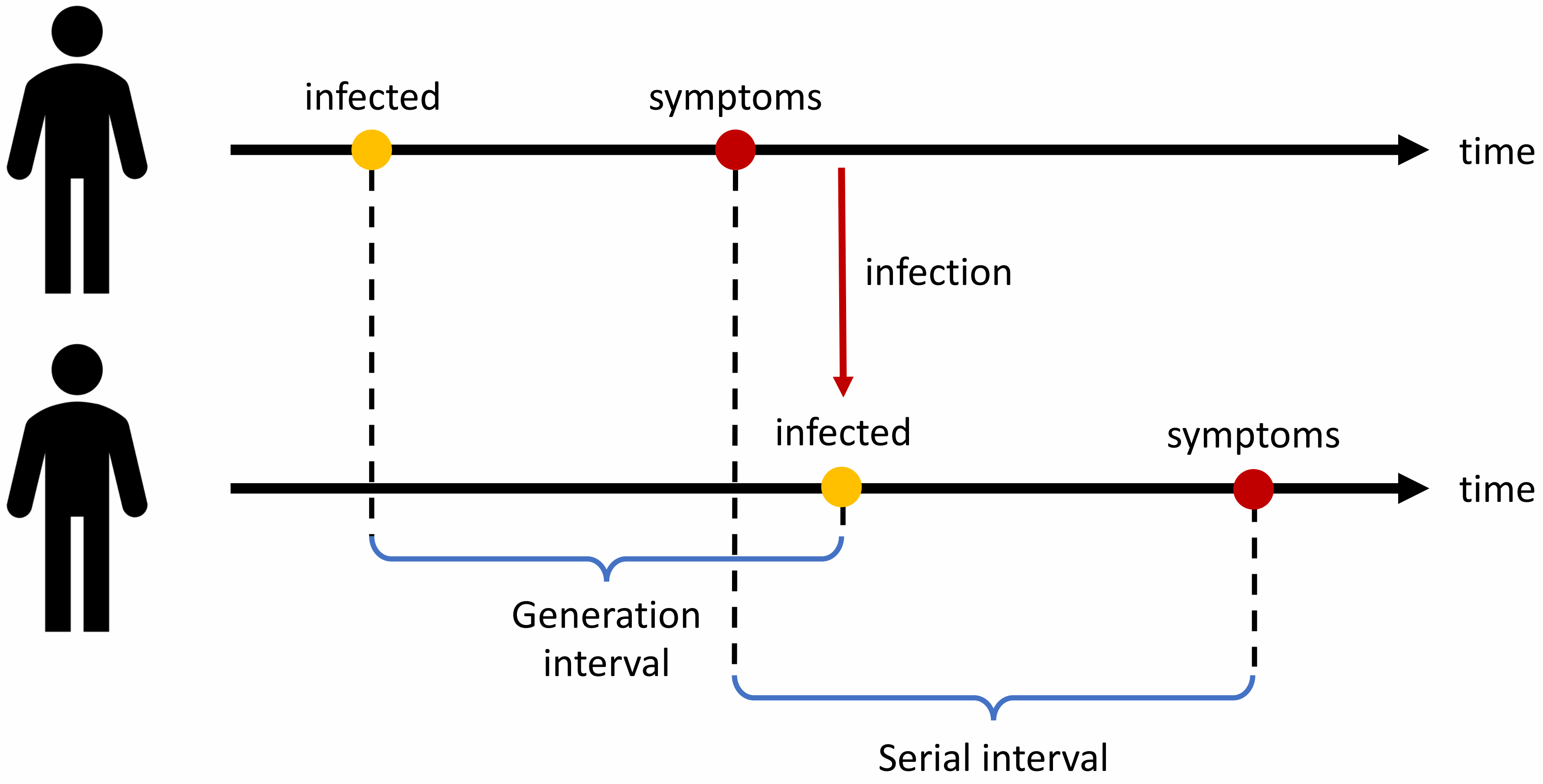


Data: height of bar represents proportion of people with that incubation period

Model (statistical): a distribution, in this case Weibull, that looks similar to the data

Ranges and averages can then be calculated from the data OR the model

Generation interval and Serial interval



Key concepts and terminology

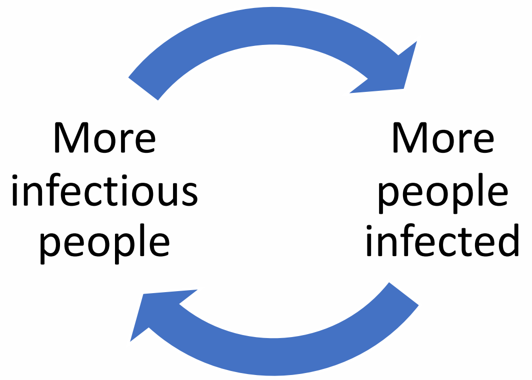
Other key terminology:

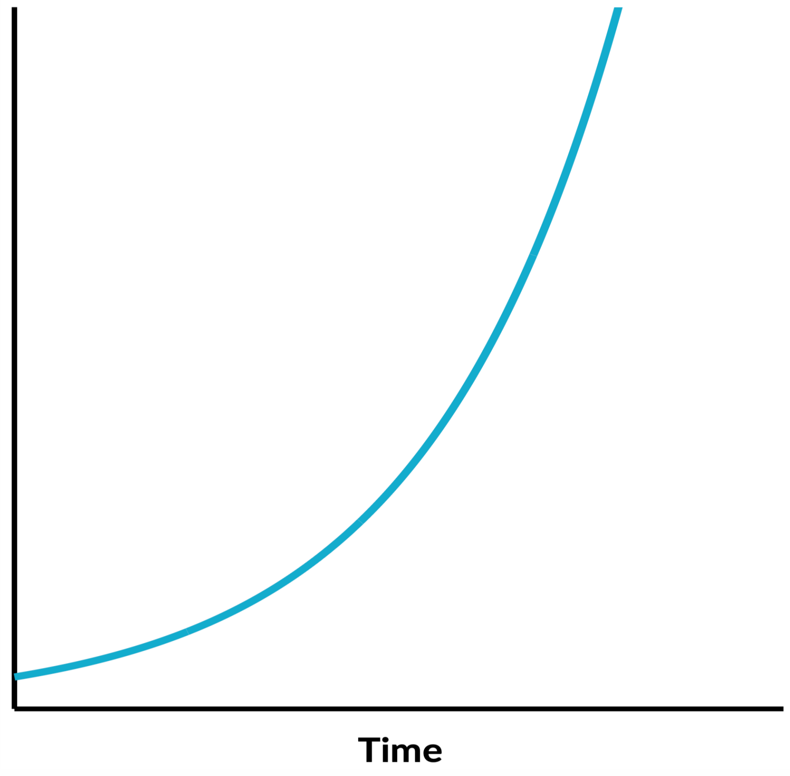
- Case fatality rate: The proportion of cases that result in death
- Epidemic: A number of cases of an infectious disease in excess of the expected frequency in a population
- Host: Person or living species capable of being infected
- Risk: The probability that an event, outcome, disease or condition will develop in a specified time period
- Sensitivity: The probability that an individual who has a disease will get a positive test result
- Specificity: The probability than an individual who does not have a disease will get a negative test result
- Surveillance: Collection of data on a particular disease
- Susceptible: An individual who is at risk of contracting a particular disease

Growth rate

Early on in an epidemic the number of cases grows **exponentially**

Exponential growth is where the rate of change is proportional to the current quantity





DEPARTMENT FOR CONTINUING EDUCATION
UNIVERSITY OF OXFORD

Growth rate

If the number of cases, $I(n)$ = number of infectious people on day n, is growing with a daily growth rate of λ then

$$I(n+1) = \lambda I(n)$$

Where λ is the **daily growth rate**

Growth rate

If $\lambda > 0$: Number of infectious people is increasing

If $\lambda < 0$: Number of infectious people is decreasing

If $\lambda = 0$: Number of infectious people is constant

Growth rate

When λ is small (i.e. close to zero, either positive or negative) the daily increase / decrease in cases is approximately:

$$(100 \times \lambda)\%$$

e.g.

If $\lambda = 0.01$ then cases will *increase* by approximately 1% each day

If $\lambda = -0.01$ then cases will *decrease* by approximately 1% each day

DEPARTMENT FOR CONTINUING EDUCATION
UNIVERSITY OF OXFORD

Growth rate

For larger values of λ (or for a more exact estimate) the daily increase / decrease in cases is given by:

$$100\times(e^{\lambda}-1)\%$$

The *exponential* function

e.g.

If $\lambda = 0.5$ then cases will increase by $100(e^{0.5}-1)\% = 64.9\%$ each day

DEPARTMENT FOR CONTINUING EDUCATION
UNIVERSITY OF OXFORD

Doubling time

The *doubling time* is the time taken for the number of cases to double in size

E.g. if the doubling time is 10 days, and we start with 1 case:

Day	Number of cases
0	1
10	2
20	4
30	8
40	16
50	32

Doubling time

If case numbers are decreasing, then we talk about the *half-life* instead of the doubling time: the time taken for the number of cases to half in size

E.g. if the doubling time is 10 days, and we start with 4,000 cases:

Day	Number of cases
0	4000
10	2000
20	1000
30	500
40	250

Calculating the growth rate and doubling time from data

We can use the incidence package in R to estimate the growth rate and the doubling time from incidence data

Step one: install and load the package

Install in the console (you only need to do this once):

```
install.packages('incidence')
```

In your markdown file 'setup' code chunk (you need to run this each session):

```
library(incidence)
```

Calculating the growth rate and doubling time from data

We can use the incidence package in R to estimate the growth rate and the doubling time from incidence data

Step two: manipulate the data

The incidence package uses a special type of object called an incidence object

Calculating the growth rate and doubling time from data

We can use the incidence package in R to estimate the growth rate and the doubling time from incidence data

Step three: fit a log-linear regression model

This type of model assumes exponential growth (or decline) and can give us mean estimates, plus 95% confidence intervals, of:

- Growth rate
- Doubling time

The basic reproduction number

The basic reproduction number is defined as:

"The expected number of new cases directly generated by one case in a population where all individuals are susceptible to infection"

We call this value R_0

This is a *similar* concept to the "R number" which you may have seen discussed in the media during the COVID-19 pandemic

The basic reproduction number

The **effective** reproduction number is defined as:

"The expected number of new cases directly generated by one case in the current population"

We call this value R_e (or R_t)

This is more likely the "R number" which was being referred to.

The basic reproduction number

The **basic** reproduction number (R_0) describes what might happen in the case of **new introduction** of an infection into a population

The **effective** reproduction number (R_e) describes what might happen next given the **current** situation (including any interventions)

Early-on in an outbreak: $R_e \approx R_0$

Later-on in an outbreak: $R_e = \frac{S}{N} R_0$

The basic reproduction number

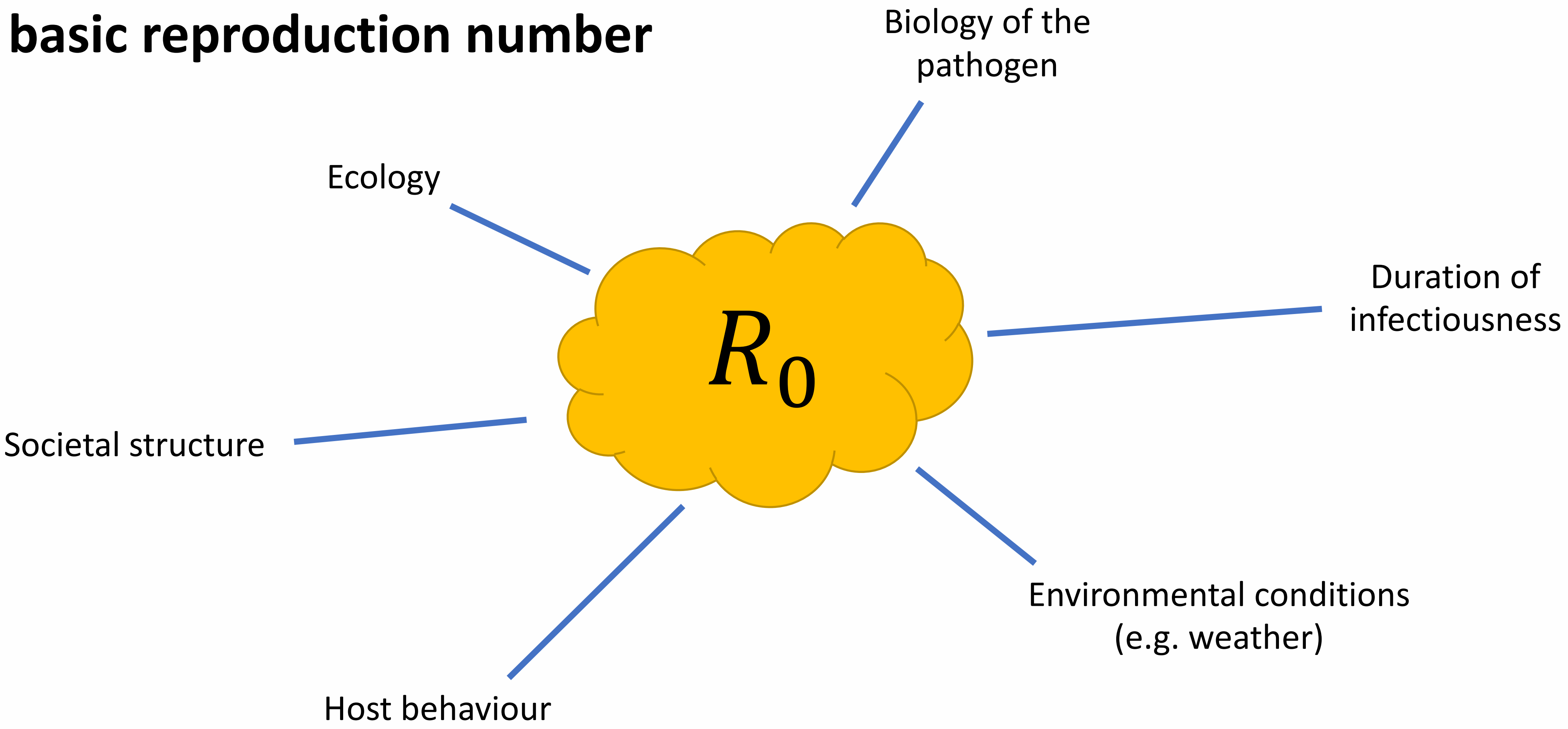
We can estimate the **basic** reproduction number (R_0) using data from the **start** of an outbreak

- Outbreak in a fully susceptible population (before any immunity is gained)
- Before any major interventions are introduced

We can estimate the **effective** reproduction number (R_e) using the **most recent** available data from an outbreak

- Restrict to only the most recent data (ignore "older" data)
- Want to pick a window in which interventions are **constant**
- Longer time window → more data, so more confident, but less "real-time" estimate
- Shorter time window → less data, so less confident, but more "real-time" estimate

The basic reproduction number



DEPARTMENT FOR CONTINUING EDUCATION
UNIVERSITY OF OXFORD

The basic reproduction number

General properties:

- Outbreaks if $R_0 > 1$
- Extinction if $R_0 < 1$

Remember R_0 is an *average*

- The larger R_0 $\rightarrow$ the harder an outbreak is to control
- R_0 does not describe how *quickly* an outbreak will spread through a population, just how *likely* it is to happen

If we are seeing an outbreak, then this means $R_0 > 1$, but we still want to know how hard it will be to control

DEPARTMENT FOR CONTINUING EDUCATION
UNIVERSITY OF OXFORD

The basic reproduction number

General properties:

- Outbreaks if $R_0 > 1$
- Extinction if $R_0 < 1$

Remember R_0 is an *average*

- The larger $R_0 \rightarrow$ the harder an outbreak is to control
- R_0 does not describe how *quickly* an outbreak will spread through a population, just how *likely* it is to happen

If cases are decreasing, then it's likely $R_e < 1$

The basic reproduction number

Examples

Disease	Measles	Chickenpox	Mumps	Rubella	Polio	Smallpox	HIV/AIDs	COVID-19 (early strain)
R_0	12-18	10-12	10-12	6-7	5-7	3.5-6	2-5	2.4-3.4

DEPARTMENT FOR CONTINUING EDUCATION
UNIVERSITY OF OXFORD

The basic reproduction number

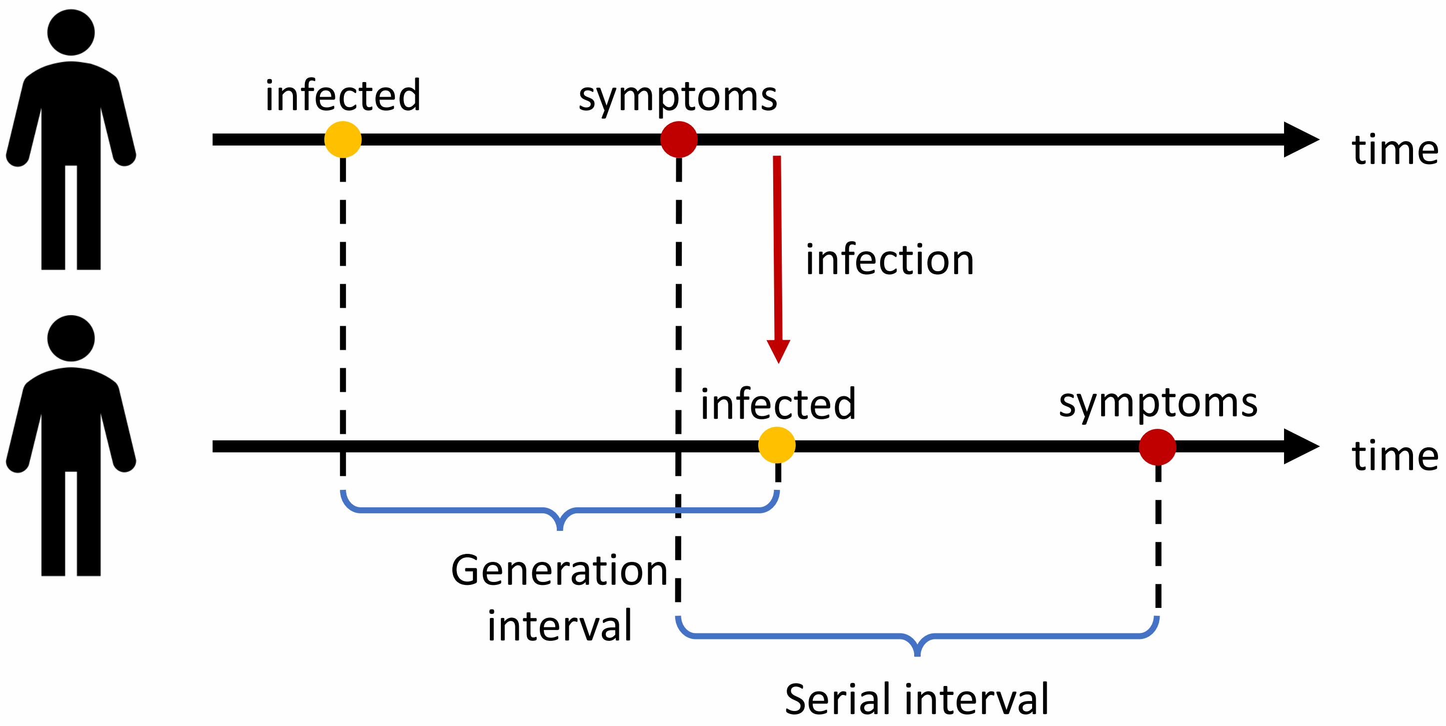
Examples

Disease	SARS	Diphtheria	Common cold	Monkeypox	Ebola (2014 outbreak)	Influenza	Nipah virus	MERS
R_0	2-4	1.7-4.3	2-3	1.5-2.7	1.4-1.8	1.2-1.4	0.5	0.3-0.8

DEPARTMENT FOR CONTINUING EDUCATION
UNIVERSITY OF OXFORD

The basic reproduction number

It is only possible to estimate the basic reproduction number from data if we have a reasonable understanding of the distribution of the **serial interval**



The basic reproduction number

It is only possible to estimate the basic reproduction number from data if we have a reasonable understanding of the distribution of the **serial interval**

This is because incidence is a combination of

- How many new infections are caused by each person (R_e)
- How quickly these new infections happen (serial interval)

Our main data source is generally incidence data

The basic reproduction number

It is only possible to estimate the basic reproduction number from data if we have a reasonable understanding of the distribution of the **serial interval**

R_0 + Serial interval = Data

Our main data source is generally incidence data

Calculating the reproduction number from data

We can use the EpiEstim package in R to estimate the instantaneous reproduction number from data, if we have information on the serial interval

For COVID-19 [1]:

- Serial interval mean = 5.59
- Serial interval standard deviation = 4.15

[1] Challen, Robert, et al. "Estimates of regional infectivity of COVID-19 in the United Kingdom following imposition of social distancing measures." *Philosophical Transactions of the Royal Society B* 376.1829 (2021): 20200280.

DEPARTMENT FOR CONTINUING EDUCATION
UNIVERSITY OF OXFORD

Summary

- The generation interval is the time between infection events
- The serial interval is the time between symptom onsets

In consecutive cases

- The daily growth rate, λ, describes how quickly an epidemic is growing
  - The mean number of new cases on a given day is equal to λ times the number of new cases on the previous day
  - $\lambda > 0 =$ growing incidence
  - $\lambda < 0 =$ declining incidence
- The doubling time is the time it takes for the daily incidence to double
- The halving time is the time it takes for the daily incidence to half
- We can use the incidence package to estimate the growth rate and doubling time from data

Summary

- The basic reproduction number, R_0, is the average number of new cases directly caused by one infectious person in an entirely susceptible population
  - This is constant over time unless conditions change
- The effective reproduction number, R_e, is the average number of new cases directly caused by one infectious person at any given point during the epidemic
  - This will change over time across the epidemic depending on number of susceptible people and interventions
- Incidence comes from a combination of R_e and the serial interval
- If we know the serial interval, we can use the EpiEstim package to estimate R_e from incidence data
  - $R_e > 1 =$ epidemic is growing
  - $R_e < 1 =$ epidemic is declining

Weekly reading

https://intro2r.com

Chapter 2: Some R basics

Generation time:

https://plus.maths.org/content/why-generation-time-covid-19-important

Growth rate:

https://plus.maths.org/content/epidemic-growth-rate