Forecasting infectious-disease outbreaks

The role of generation intervals

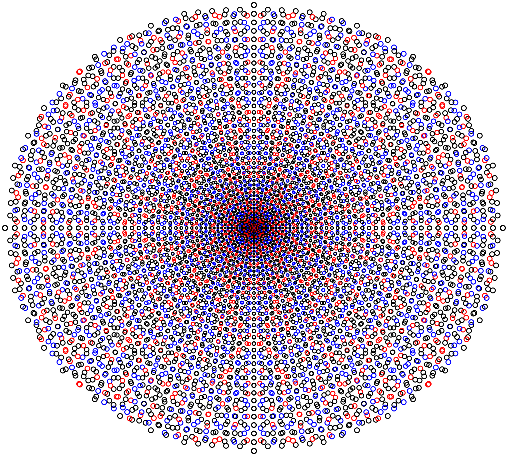
Jonathan Dushoff, McMaster University

U. Chicago, Oct 2018

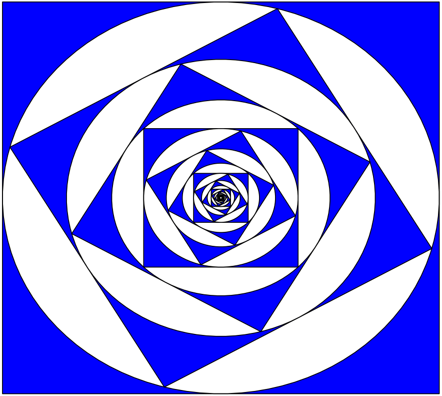
Outline

Problem (present)

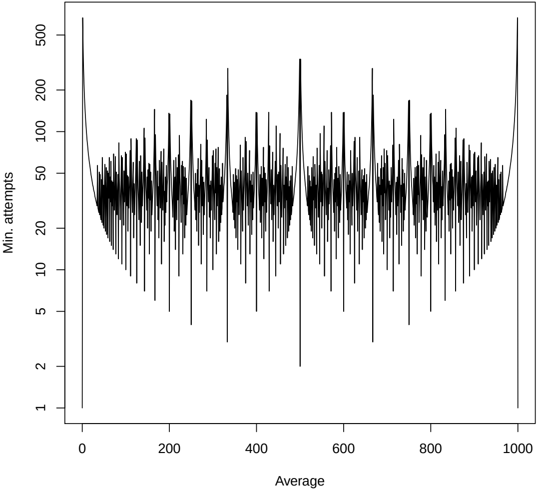
- I am fundamentally a math person . . .



What is the pattern of Pythagorean triples of integers
$a^2 + b^2 = c^2$?



Divide a square and a circle each into two complementary subsets that are pairwise similar



How many at-bats does it take to get a given batting average?

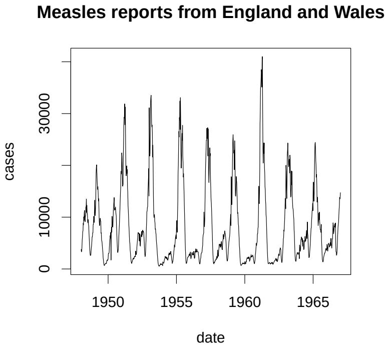
Problem

- I am fundamentally a math person
  - but I want to do work that is relevant to people

Solution

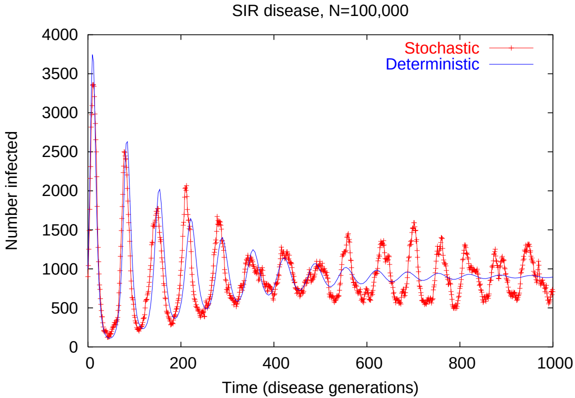
- Dynamical modeling is fun and useful

Dynamical modeling connects scales



- Start with rules about how things change in short time steps
  - Usually based on *individuals*
- Calculate results over longer time periods
  - Usually about *populations*

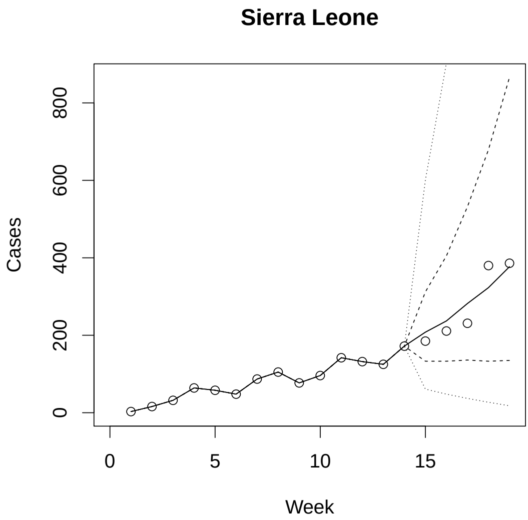
Fun and useful!



New problem

- There is (or was a gulf) between dynamical and statistical modeling
  - Dynamics are needed to incorporate mechanism
  - Statistics are needed to incorporate uncertainty

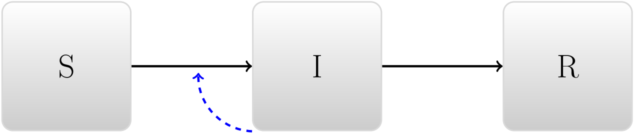
Ebola forecasting



Outline

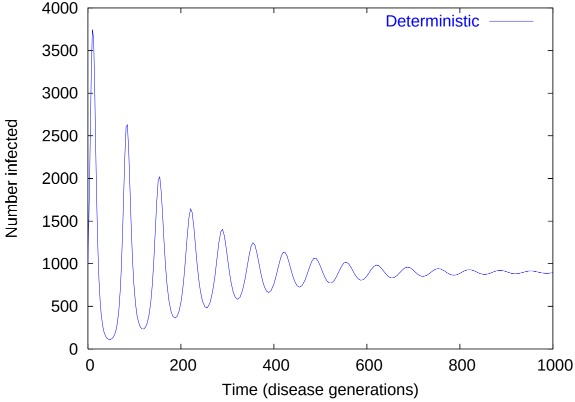
Compartmental models

Divide people into categories:

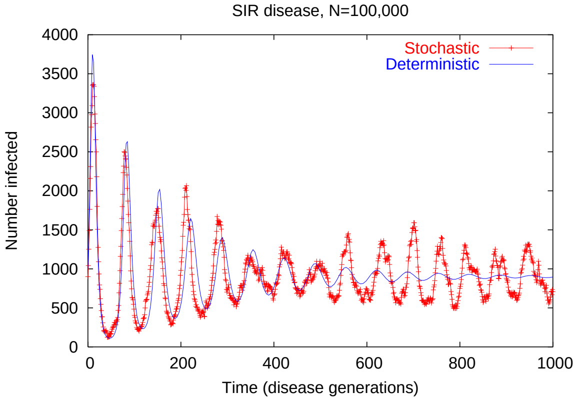


- Susceptible $\to$ Infectious $\to$ Recovered
- Individuals recover independently
- Individuals are infected by infectious people

Differential equation implementation



Individual-based implementation



Lessons

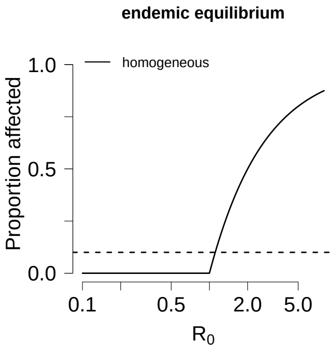
- Tendency to oscillate
- Thresholds
- Exponential growth

Big $\mathcal{R}$

- $\mathcal{R}$ is the number of people who would be infected by an infectious individual *in a fully susceptible population.*
- $\mathcal{R} = \beta/\gamma = \beta D = (cp)D$
  - c: Contact Rate
  - p: Probability of transmission (infectivity)
  - D: Average duration of infection
- A disease can invade a population if and only if $\mathcal{R} > 1$.
- Often focus on initial period (may also say $\mathcal{R}_0$)

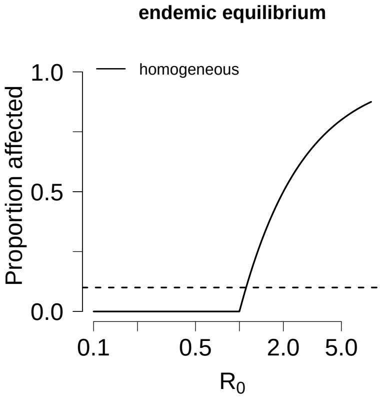
Big $\mathcal{R}$

Homogeneous endemic curve



- Threshold value
- Sharp response to changes in factors underlying transmission
- Works – sometimes
  - Sometimes predicts unrealistic sensitivity

Yellow fever in Panama



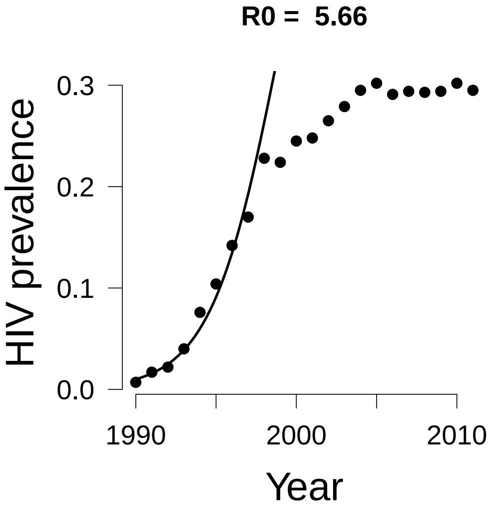
Exponential growth

- Diseases have a tendency to grow exponentially at first
  - I infect three people, they each infect 3 people . . .
  - How fast does disease grow?
  - How quickly do we need to respond?

little r

- We measure epidemic *speed* using little r:
  - *Units*: $[1/\text{time}]$
  - Disease increases like e^{rt}
- Time scale is $C = 1/r$
  - Ebola, $C \approx 1\text{month}$
  - HIV in SSA, $C \approx 18\text{month}$
- Often focus on initial period (may also say r_0)

little r



Limitations

- Many conclusions from this framework make strong assumptions:
  - **Spatial homogeneity:** everywhere is the same
  - **Individual homogeneity:** everyone is the same
    - and everyone is everywhere
  - **Temporal homogeneity:**
    - It doesn't matter how long I've been infected, I'm either infected or not

Outline

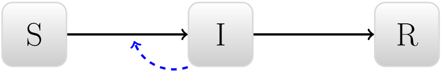
The $r\mathcal{R}$ relationship

- We're very interested in the relationship between little r and $\mathcal{R}$.
- We might have good estimates of r only
  - e.g., West African Ebola outbreak, HIV in Africa
- Or we might have good estimates of $\mathcal{R}$ only
  - Measles, influenza

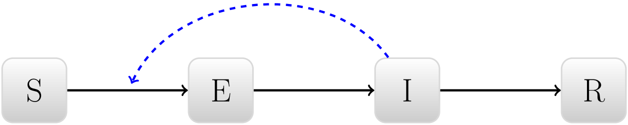
Example: Post-death transmission and safe burial

- How much Ebola spread occurs before vs. after death
- Highly context dependent
  - Funeral practices, disease knowledge
- *Weitz and Dushoff Scientific Reports 5:8751.*

Standard disease model

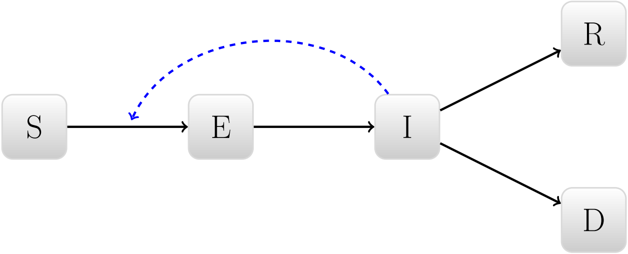


Add a latent period

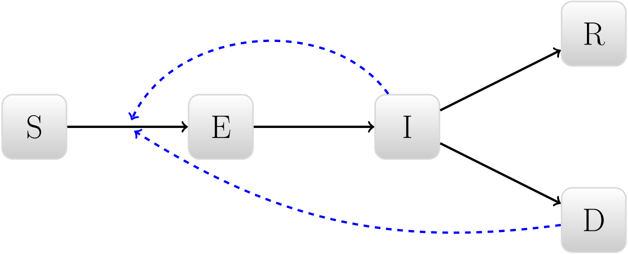


- (i.e., a lag between infection and infectiousness)

Add post-death transmission



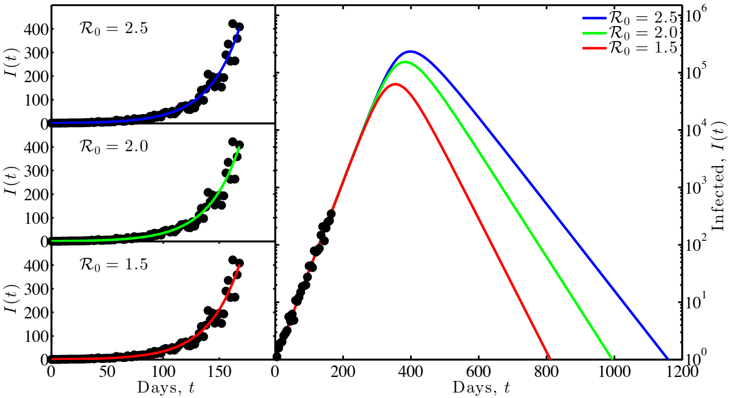
Add post-death transmission



What happens if we account for burial transmission?

- We've made the disease transmitting process slower, so obviously Ebola is *less* dangerous than we thought
- We've added another source of transmission, so obviously Ebola is *more* dangerous than we thought
- What we learn depends on what we know!

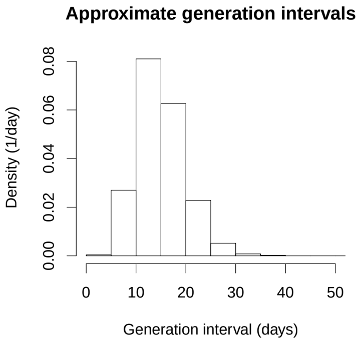
What do we know?



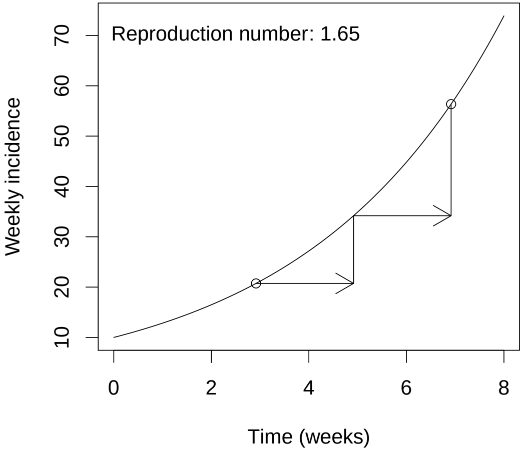
Outline

Generation intervals

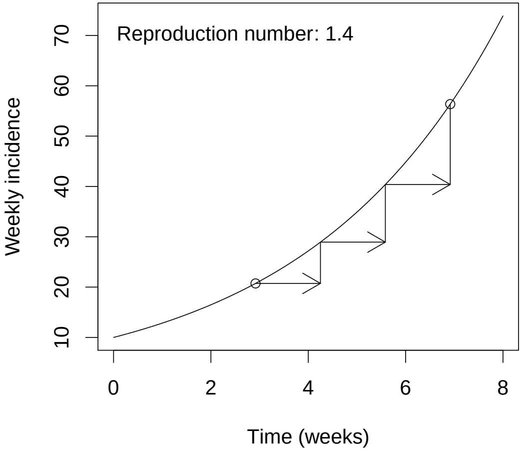
- The generation distribution measures the time between generations of the disease
  - Interval between "index" infection and resulting infection
- Generation intervals provide the link between $\mathcal{R}$ and r



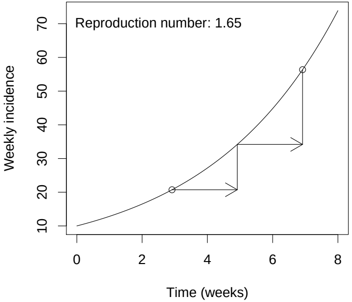
Generations and $\mathcal{R}$

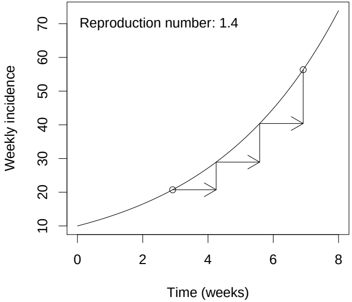


Generations and $\mathcal{R}$



Generations and $\mathcal{R}$





Conditional effect of generation time

- *Given* the reproductive number $\mathcal{R}$
  - faster generation time G means higher r
  - More danger
- *Given* r
  - faster generation time G means *smaller* $\mathcal{R}$
  - Less danger

Linking framework

- Epidemic speed (r) is a *product*:
  - (something to do with) generation speed
  - $\times$ (something to do with) epidemic strength
- Epidemic strength is therefore (approximately) a *quotient*
  - Epidemic speed
  - $\div$ (something to do with) generation speed

Filtered means

- There is a sensible way to define an "effective" generation time
- Preserve the exponential growth equation

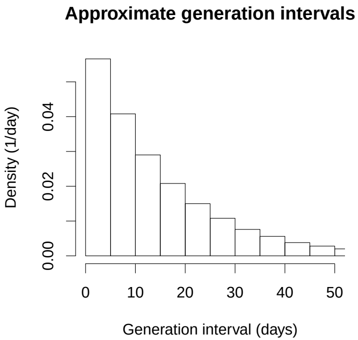
$$\mathcal{R} = \exp(r\hat{G})$$

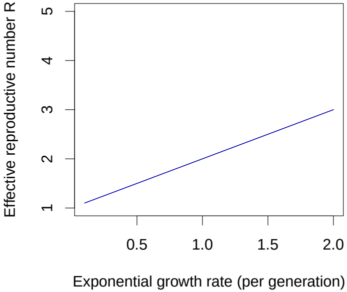
- $\hat{G}$ is a "filtered mean" of the distribution g:
  -
$$\exp(-r\hat{G}) = \langle \exp(-r\tau) \rangle_g .$$

,

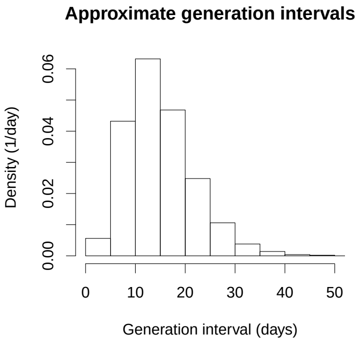
- This is cool, but not easy to interpret (our estimates about the generation time change when r changes)

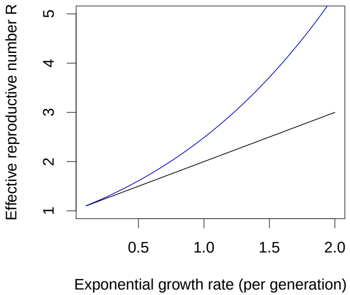
Approximations



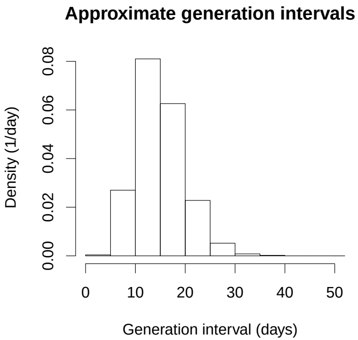


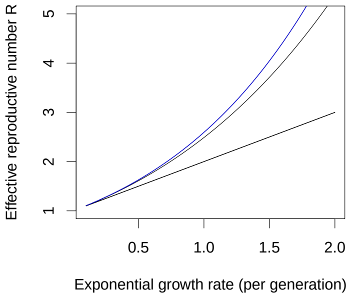
Moment approximation



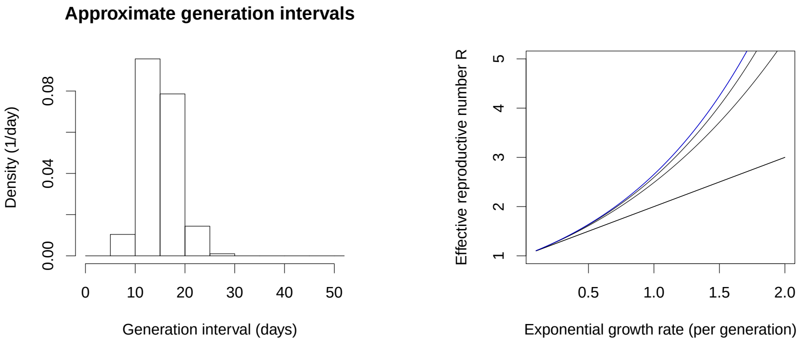


Moment approximation





Moment approximation



Compound-interest interpretation

- Define $\mathcal{R} \approx (1 + r\kappa\bar{G})^{1/\kappa} \equiv X(r\bar{G}; 1/\kappa)$
- X is the compound-interest approximation to the exponential
  - Linear when $\kappa = 1$ (i.e., when g is exponential)
  - Approaches exponential as $\kappa \to 0$
- Key quantity is $r\bar{G}$: the relative length of the generation interval compared to the characteristic time scale of spread

Qualitative response

- For a given value of $\bar{G}$, smaller values of κ mean:
  - less variation in generation interval
  - less compounding of growth
  - greater $\mathcal{R}$ required for a given r

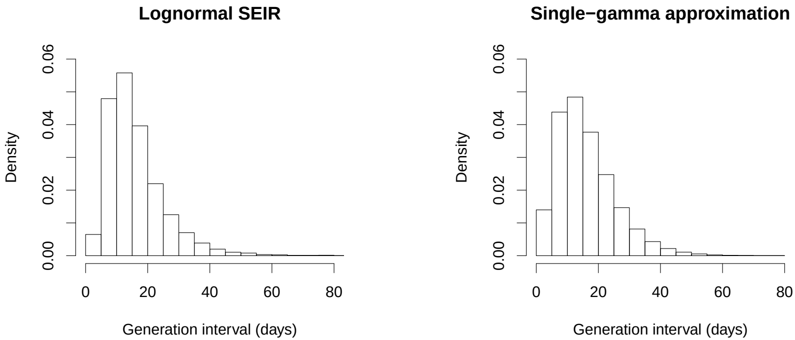
Intuition

- Longer generation times mean less speed
  - $\implies$ more strength, when speed is fixed
- What about more variation?
  - More action (both before and after the mean time)
  - But what happens early is more important in a growing system
- More variation means more speed
  - $\implies$ less strength, when speed is fixed

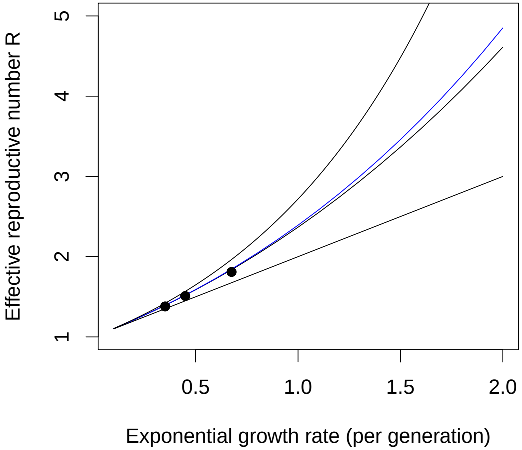
How well do approximations work

- Simulate realistic generation intervals for various diseases
- Compare approximate $r\mathcal{R}$ relationship with known exact relationship
  - Known because we are testing ourselves with simulated data

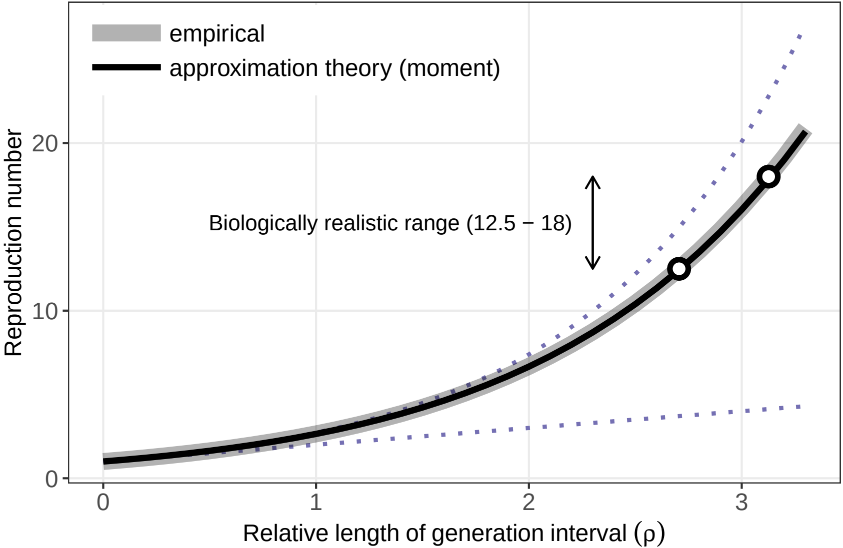
Ebola distribution



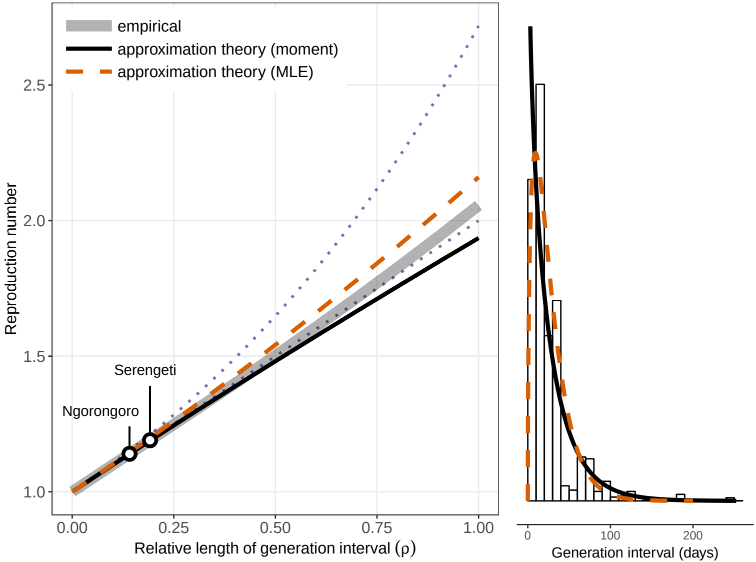
Ebola curve


Effective reproductive number R
1
2
3
4
5
0.5
1.0
1.5
2.0
Exponential growth rate (per generation)

Measles curve



Rabies curve



Generation intervals

'I cannot imagine how they will spend half of it.'

- Sort of the poor relations of disease-modeling world
- Ad hoc methods
- Error often not propagated

Summary

- Generation intervals are the missing link between r and $\mathcal{R}$
- We need better methods for estimating them, and propagating uncertainty to other parts of the model
- Filtered means may help with intuition
- For many practical applications:
  - Estimating the mean generation interval is not enough
  - But estimating the mean and CV may be enough
  - A good basis for understanding and propagating uncertainty

Outline

Generations through time

- Generation intervals can be estimated by:
  - Observing patients:
    - How long does it take to become infectious?
    - How long does it take to recover?
    - What is the time profile of infectiousness/activity?
  - Contact tracing
    - Who (probably) infected whom?
    - When did each become infected?
    - — or ill (serial interval)?

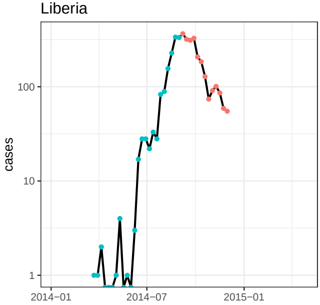
Which is the real interval?

- Contact-tracing intervals look systematically different, depending on when you observe them.
- Observed in:
  - Real data, detailed simulations, simple model
- Also differ from intrinsic (infector centered) estimates

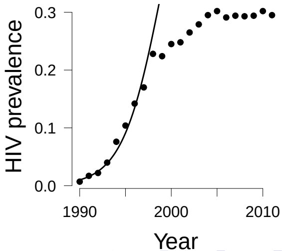
Types of interval

- Define:
  - *Intrinsic interval*: How infectious is a patient at time τ after infection?
  - *Forward interval*: When will the people infected today infect others?
  - *Backward interval*: When did the people who infected people today themselves become infected?
  - *Censored interval*: What do all the intervals observed up until a particular time look like?
    - Like backward intervals, if it's early in the epidemic

Growing epidemics



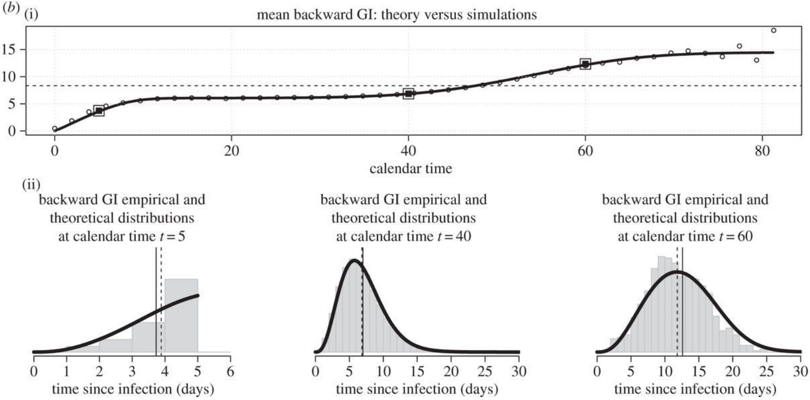
- Generation intervals look *shorter* at the beginning of an epidemic
    - A disproportionate number of people are infectious right now
    - They haven't finished all of their transmitting
    - We are biased towards observing faster events



What changes backward intervals?

- Who is likely to infect me depends on:
  - How infectious they are (intrinsic GI)
  - How many of them there are (changes in disease incidence)

Backward intervals

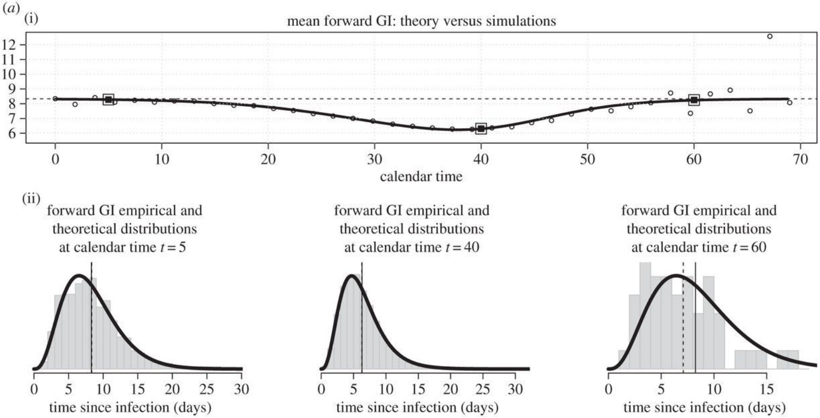


Champredon and Dushoff, 2015. DOI:10.1098/rspb.2015.2026

What changes forward intervals?

- Who I am likely to infect depends on:
  - How infectious I am (intrinsic GI)
  - How many of them there are (changes in numbers of susceptibles)

Forward intervals



Champredon and Dushoff, 2015. DOI:10.1098/rspb.2015.2026

Conclusion

- Backward intervals change if the number of infectious individuals is changing as you look back
- Forward intervals change if the number of *susceptible* individuals is changing as you look forward
- Lack of care in defining generation intervals can lead to bias
  - In particular, censored intervals look too short, lead to underestimates of $\mathcal{R}$.

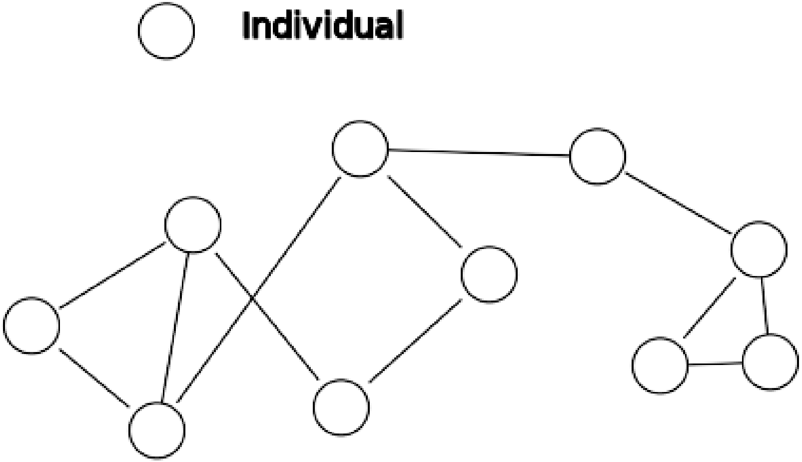
Outline

Other kinds of generation interval

- Once you think carefully about generation intervals, they're everywhere
- Spatial heterogeneity
- Individual heterogeneity

Generations in space

- How do local interactions affect realized generation intervals?

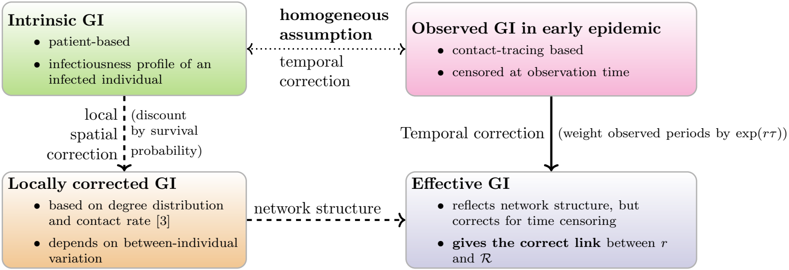


Surprising results

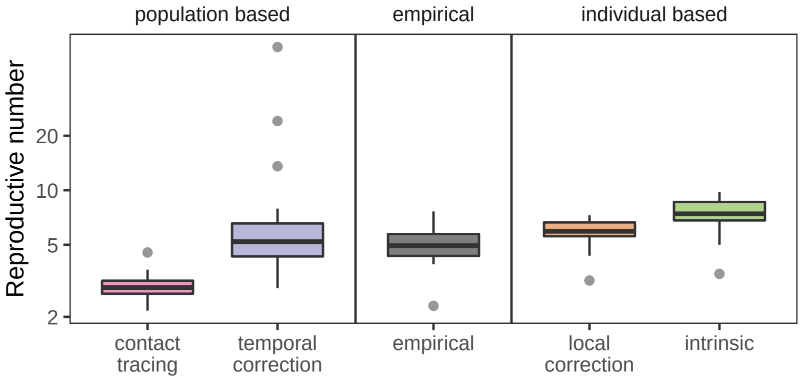
- $\mathcal{R}$ on networks generally *smaller* than values estimated using r.
  - *Trapman et al., 2016. JRS Interface DOI:10.1098/rsif.2016.0288*

- Because people don't question the intrinsic generation interval
  - Local interactions
  - $\implies$ wasted contacts
  - $\implies$ shorter generation intervals
  - $\implies$ smaller estimates of $\mathcal{R}$.

Observed and estimated intervals

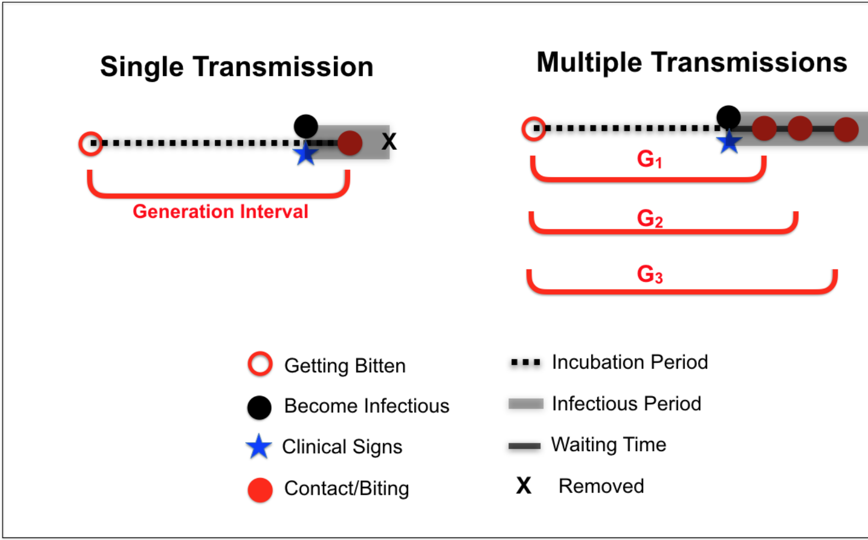


Outbreak estimation



Serial intervals

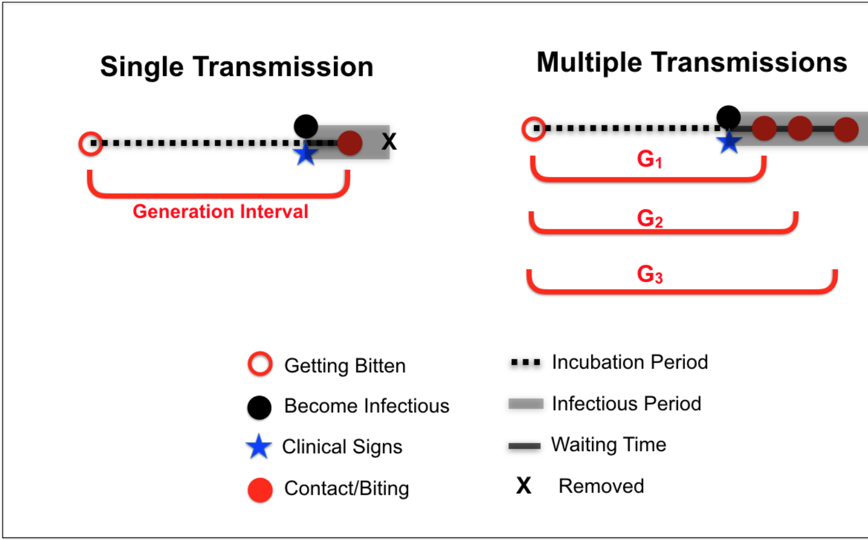
- Do serial intervals and generation intervals have the same distribution?
- It seems that they should: they describe generations of the same process
- In fact, they don't
  - Serial intervals can even be negative!
  - You might report to the clinic with flu before me, even though I infected you

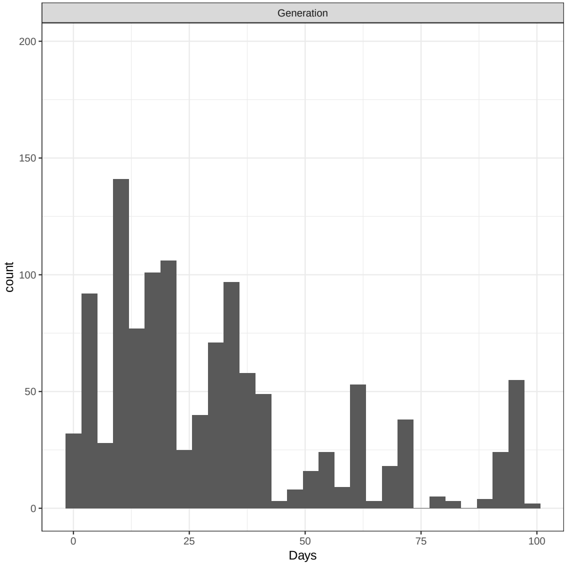


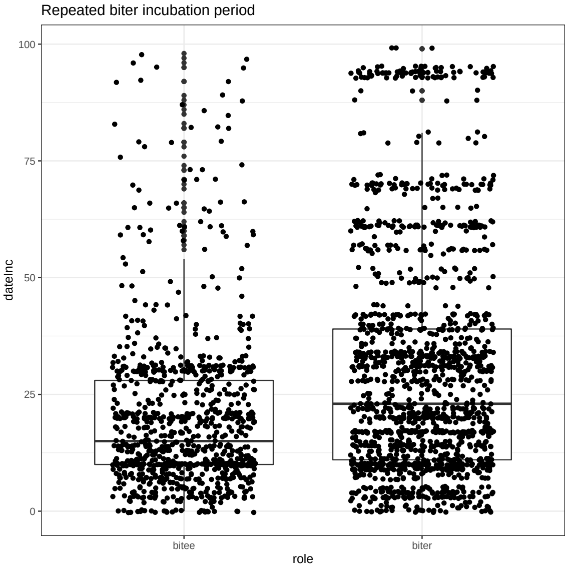
Single Transmission
Multiple Transmissions
X
Generation Interval
G1
G2
G3
Getting Bitten
Become Infectious
Clinical Signs
Contact/Biting
Incubation Period
Infectious Period
Waiting Time
X
Removed

Rabies

- If symptoms always start *before* infectiousness happens, then serial interval should equal generation interval:
  - incubation time + extra latent time + waiting time
  - extra latent time + waiting time + incubation time


Single Transmission
Multiple Transmissions
X
Generation Interval
G_1
G_2
G_3
Getting Bitten
Become Infectious
Clinical Signs
Contact/Biting
Incubation Period
Infectious Period
Waiting Time
X Removed


Generation
count
200
150
100
50
0
0
25
50
75
100
Days


Repeated biter incubation period
100
75
50
25
0
dateInc
bitee
biter
role

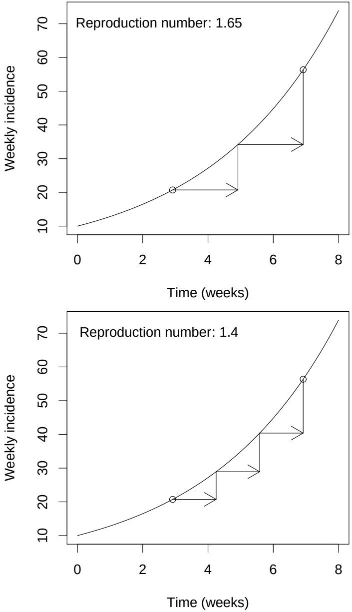
Outline

Can treatment stop the HIV epidemic?

- Modern treatments are well tolerated and highly effective
- Virus is undetectable, and transmission is negligible
- Can active testing and treatment stop the epidemic?

Are HIV generations fast or slow?

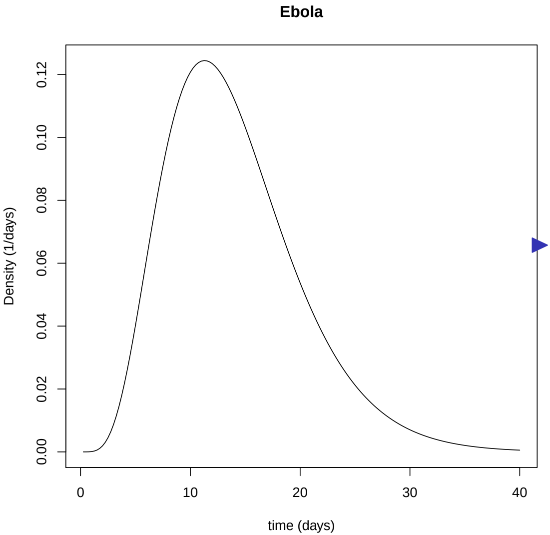
- Fast generations mean:
  - Testing and treating will help less
  - *but* lower epidemic strength



Eaton and Hallett

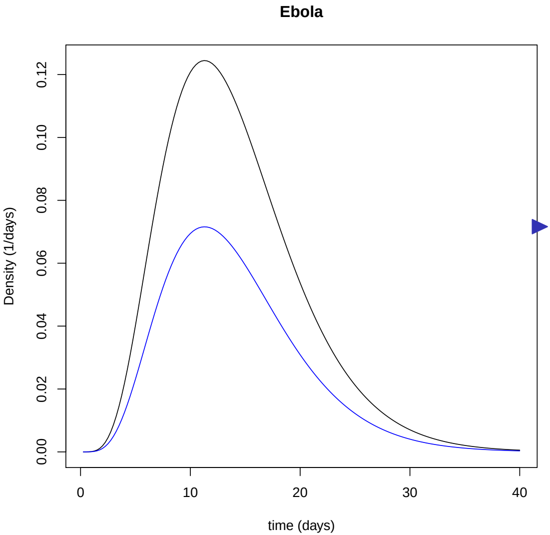
- Predicted effectiveness of test and treat intervention minimally sensitive to proportion of early transmission
- Fast transmission:
  - low proportion prevented, but low $\mathcal{R}$ estimate
- Slow transmission:
  - high proportion prevented, but high $\mathcal{R}$ estimate
- *Eaton JW, Hallett TB. Proc Natl Acad Sci U S A. 2014 Nov 11;111(45):16202-7.*

Epidemic strength (present)



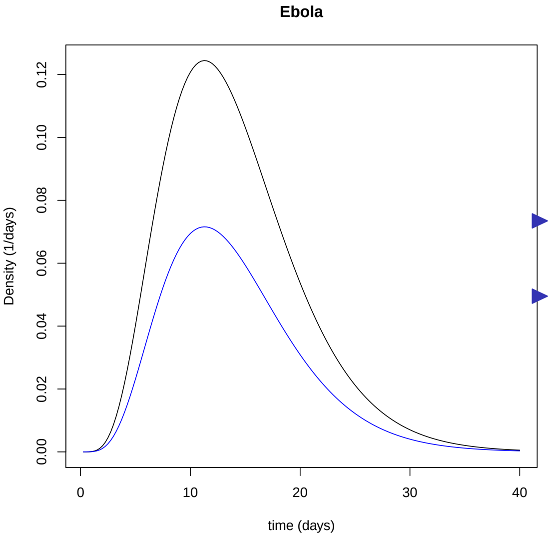
- $\mathcal{R}$, the epidemic strength, is the area under the curve.

Strength of intervention



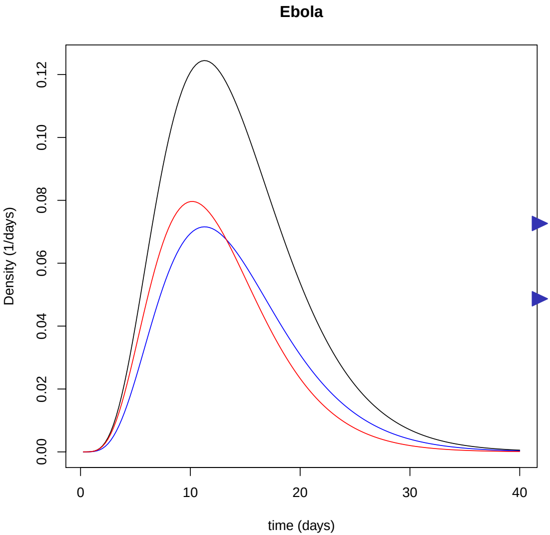
- . . . by what factor do I need to reduce this curve to eliminate the epidemic?

Different interventions (present)



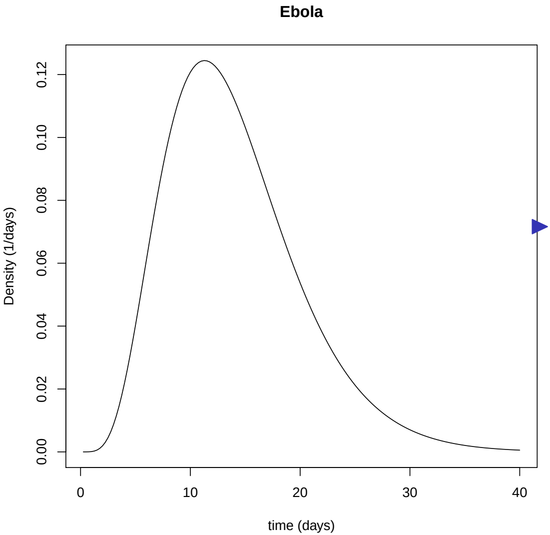
- idealized vaccination
- removes a fixed proportion of people

Different interventions (present)



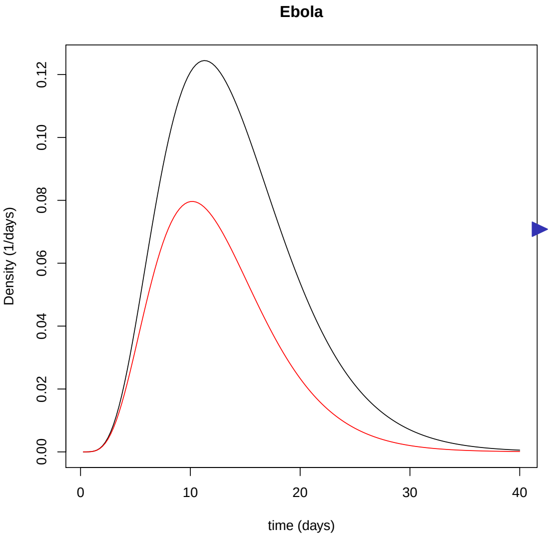
- idealized quarantine
- removes people at a fixed rate

Epidemic speed



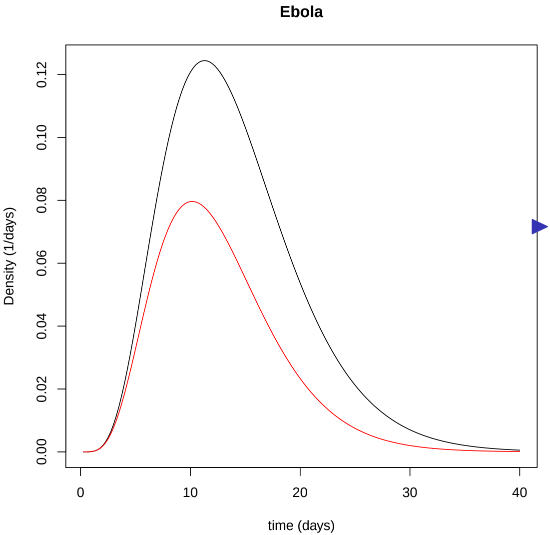
- r, the epidemic speed, is the "discount" rate required to balance the tendency to grow

Epidemic speed



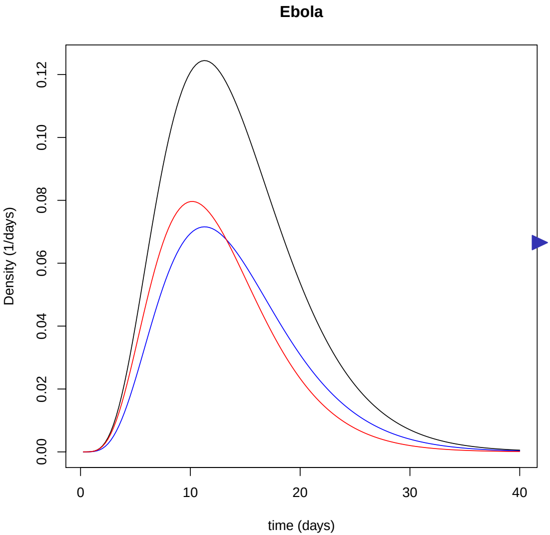
- $k(\tau) = \exp(r\tau)b(\tau)$, where $b(\tau)$ is the initial backward generation interval

Speed of intervention



- . . . how *quickly* do I need to reduce this curve to eliminate the epidemic?

Different interventions (present)



- Sometimes it's easier to estimate strength, sometimes speed

The strength paradigm

- $\mathcal{R} > 1$ is a threshold
- If we can reduce transmission by a constant *factor* of $\theta > \mathcal{R}$, disease can be controlled
- In general, we can define θ as a (harmonic) mean of the reduction factor over the course of an infection
  - weighted by the *intrinsic* generation interval
- Epidemic is controlled if $\theta > \mathcal{R}$

The speed paradigm

- $r > 0$ is a threshold
- If we can reduce transmission at a constant *hazard rate* of $\phi > r$, disease can be controlled
- In general, we can define ϕ as a (very weird) mean of the reduction factor over the course of an infection
  - weighted by the *backward* generation interval
- Epidemic is controlled if $\phi > r$

Measuring the intervention

HIV

- The importance of transmission speed to HIV control is easier to understand using the speed paradigm
  - We know the speed of invasion
    - $\approx 0.7/\mathrm{yr}$
    - Characteristic scale $\approx 1.4\mathrm{yr}$
  - And can hypothesize the speed of intervention
    - Or aim to go fast enough

Paradigms are complementary

- HIV
  - Information and current intervention are both “speed-like”
- Measles
  - Information (long-term) is strength-like
  - Intervention (vaccine) also strength-like
- Ebola vaccination
  - Information is speed-like
  - Pre-emptive vaccination is strength-like


MAMIYA - SEKOR C
1 : 1.9
f = 80 mm
No.88324

Thanks

- Department
- Collaborators
- Funders: NSERC, CIHR

Linking framework

- Epidemic speed (r) is a *product*:
  - (something to do with) generation speed $\times$
  - (something to do with) epidemic strength
- In particular:
  - $r \approx (1/\bar{G}) \times \ell(\mathcal{R}; \kappa_g)$
  - ℓ is the inverse of X