

# Stochasticity during early epidemic dynamics

PROBABILITY OF ESTABLISHMENT, INITIAL GROWTH  
RATE AND THE OUTBREAK SIZE AT FIRST DETECTION

---

Pete(r) Czuppon

Institute for Evolution and Biodiversity  
University of Münster (Germany)

BIOHASARD conference  
21<sup>st</sup> of April 2021

## Joint work with...



François Blancart  
(Collège de France)



Florence Débarre  
(Sorbonne Université)



Emmanuel Schertzer  
(University of Vienna)

# COVID-19 (and its containment)

## Timeline of the project (and motivation)

- March 2020:
- ... how large are clusters at their detection?

# COVID-19 (and its containment)

## Timeline of the project (and motivation)

- March 2020:
- ... how large are clusters at their detection?
- October 2020:



The screenshot shows a news article from France 24. At the top left is the France 24 logo. In the center is a 'LIVE' indicator with a play button icon. On the right is a hamburger menu icon. Below the logo is a navigation bar with tabs for '#COVID-19', '#UK', 'FRANCE', 'AFRICA', 'CULTURE', 'SHOWS', and 'FIGHT THE FAKE'. A dark blue banner below the navigation bar reads 'Coronavirus notice • View the recommendations and information for travellers issued by the French Government →'. The main headline is 'Slovakia to test entire population for coronavirus in global first'. Below the headline are social media sharing icons for Facebook, WhatsApp, and Twitter, followed by a share icon and the text '743 shares'. At the bottom, it says 'Issued on: 31/10/2020 - 10:46 Modified: 31/10/2020 - 10:48'.

# COVID-19 (and its containment)

## Timeline of the project (and motivation)

- March 2020:
- ... how large are clusters at their detection?
- October 2020:
- ... how successful is such a one-shot mass testing effort?



The screenshot shows a news article from France 24. At the top left is the France 24 logo. In the center is a play button icon with the word "LIVE" next to it. On the right is a hamburger menu icon. Below the navigation bar is a blue banner with the text "Coronavirus notice • View the recommendations and information for travellers issued by the French Government →". Underneath is a breadcrumb trail: "Home / Europe". The main headline is "Slovakia to test entire population for coronavirus in global first". Below the headline are social media sharing icons for Facebook, WhatsApp, and Twitter, followed by a share icon and the text "743 shares". At the bottom, it says "Issued on: 31/10/2020 - 10:46 Modified: 31/10/2020 - 10:48".

# COVID-19 (and its containment)

## Timeline of the project (and motivation)

- March 2020:
- ... how large are clusters at their detection?
- October 2020:
- ... how successful is such a one-shot mass testing effort?
- December 2020:



The screenshot shows the top navigation bar of The Guardian website. It includes the text "Support the Guardian" with a subtext "Available for everyone, funded by readers" and buttons for "Contribute" and "Subscribe". There are also links for "Search jobs", "Sign in", "Search", and "International". The main navigation menu includes "News", "Opinion", "Sport", "Culture", "Lifestyle", and "More". Below the navigation, there are links for "The Guardian view", "Columnists", "Cartoons", "Opinion videos", and "Letters". The main content area features an article titled "Here's what we know about the new variant of coronavirus" by Sharon Peacock. A yellow banner above the article title states "This article is more than 3 months old". The article text begins with "My team at the Genomics UK consortium sequenced the new Sars-CoV-2 variant, but we'll need more evidence to understand". The date "Tue 22 Dec 2020 13:46 GMT" is visible at the bottom left of the article.

## Timeline of the project (and motivation)

- March 2020:
  - ... how large are clusters at their detection?
- October 2020:
  - ... how successful is such a one-shot mass testing effort?
- December 2020:
  - ... when did the variant B.1.1.7 arise?

## Timeline of the project (and motivation)

- March 2020:
  - ... how large are clusters at their detection?
- October 2020:
  - ... how successful is such a one-shot mass testing effort?
- December 2020:
  - ... when did the variant B.1.1.7 arise?
- February 2021:
  - ... can we say something about the minimal testing effort needed to detect clusters before they are too large to control easily?



# Table of contents

1. Probability of establishment
2. Mean epidemic size over time
3. Time of first detection and size of the outbreak at first detection
4. (Minimal testing frequency and mass testing)

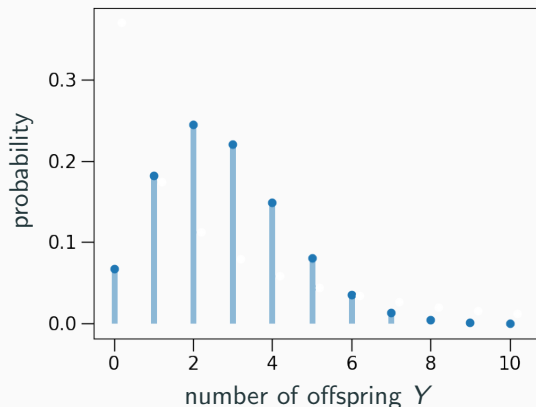
# Probability of establishment

---

# Distribution of secondary infections

= offspring distribution

- $R$  = average number of offspring = basic/effective reproduction number



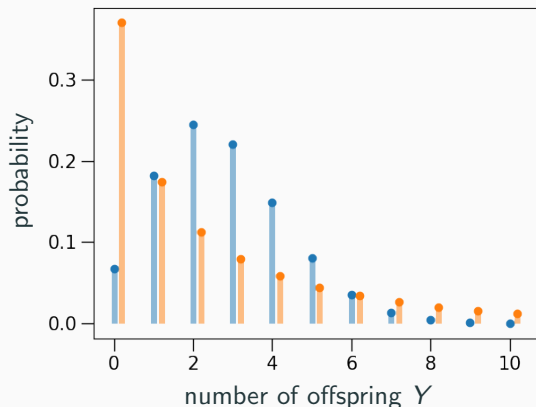
— Poisson

$$P(Y = j) = \frac{R^j}{j!} e^{-R}$$

# Distribution of secondary infections

= offspring distribution

- $R$  = average number of offspring = basic/effective reproduction number
- $\kappa$  = dispersion (measure of variance)  
⇒ super-spreading (Lloyd-Smith et al. *Nature* (2005))



— Poisson

$$P(Y = j) = \frac{R^j}{j!} e^{-R}$$

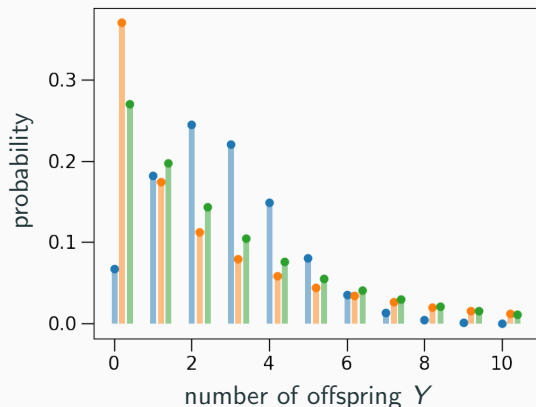
— neg. Binomial

$$P(Y = j) = \binom{j+\kappa-1}{j} \left(\frac{\kappa}{\kappa+R}\right)^\kappa \times \left(\frac{R}{\kappa+R}\right)^j$$

# Distribution of secondary infections

= offspring distribution

- $R$  = average number of offspring = basic/effective reproduction number
- $\kappa$  = dispersion (measure of variance)  
⇒ super-spreading (Lloyd-Smith et al. *Nature* (2005))



— Poisson

$$P(Y = j) = \frac{R^j}{j!} e^{-R}$$

— neg. Binomial

$$P(Y = j) = \binom{j+\kappa-1}{j} \left(\frac{\kappa}{\kappa+R}\right)^\kappa \times \left(\frac{R}{\kappa+R}\right)^j$$

— Geometric

$$P(Y = j) = \frac{1}{1+R} \left(\frac{R}{1+R}\right)^j$$

## Probability of establishment

Compute the probability of extinction  $p_{\text{ext}}$ :

$$p_{\text{ext}} = \underbrace{P(Y = 0)}_{\text{no offspring}} + \underbrace{p_{\text{ext}} P(Y = 1)}_{\text{one offspring}} + \underbrace{p_{\text{ext}}^2 P(Y = 2)}_{\text{two offspring}} + p_{\text{ext}}^3 P(Y = 3) + \dots$$

# Probability of establishment

Compute the probability of extinction  $p_{\text{ext}}$ :

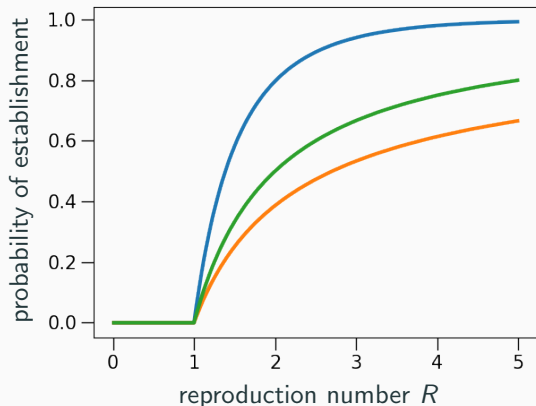
$$p_{\text{ext}} = \underbrace{P(Y = 0)}_{\text{no offspring}} + \underbrace{p_{\text{ext}} P(Y = 1)}_{\text{one offspring}} + \underbrace{p_{\text{ext}}^2 P(Y = 2)}_{\text{two offspring}} + p_{\text{ext}}^3 P(Y = 3) + \dots$$
$$= E[p_{\text{ext}}^Y] = \text{probability generating function}$$

# Probability of establishment

Compute the probability of extinction  $p_{\text{ext}}$ :

$$p_{\text{ext}} = \underbrace{P(Y = 0)}_{\text{no offspring}} + \underbrace{p_{\text{ext}} P(Y = 1)}_{\text{one offspring}} + \underbrace{p_{\text{ext}}^2 P(Y = 2)}_{\text{two offspring}} + p_{\text{ext}}^3 P(Y = 3) + \dots$$
$$= E[p_{\text{ext}}^Y] = \text{probability generating function}$$

$$p_{\text{estab}} = 1 - p_{\text{ext}}$$



Poisson ( $\kappa \rightarrow \infty$ )

Geometric ( $\kappa = 1$ )

neg. Binomial ( $\kappa = 0.57$ )

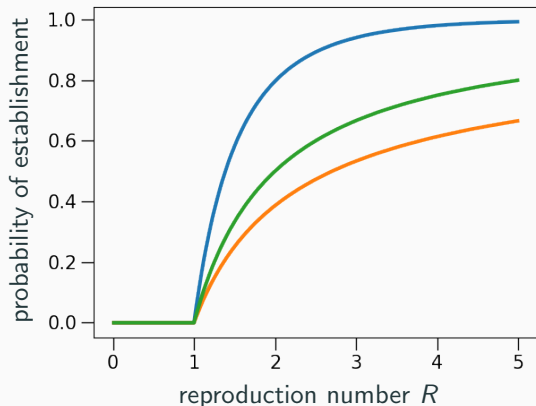


# Probability of establishment

Compute the probability of extinction  $p_{\text{ext}}$ :

$$p_{\text{ext}} = \underbrace{P(Y = 0)}_{\text{no offspring}} + \underbrace{p_{\text{ext}} P(Y = 1)}_{\text{one offspring}} + \underbrace{p_{\text{ext}}^2 P(Y = 2)}_{\text{two offspring}} + p_{\text{ext}}^3 P(Y = 3) + \dots$$
$$= E[p_{\text{ext}}^Y] = \text{probability generating function}$$

$$p_{\text{estab}} = 1 - p_{\text{ext}}$$



Poisson ( $\kappa \rightarrow \infty$ )

Geometric ( $\kappa = 1$ )

neg. Binomial ( $\kappa = 0.57$ )

$$p_{\text{estab}} = 1 - \frac{1}{R} \quad (R > 1)$$

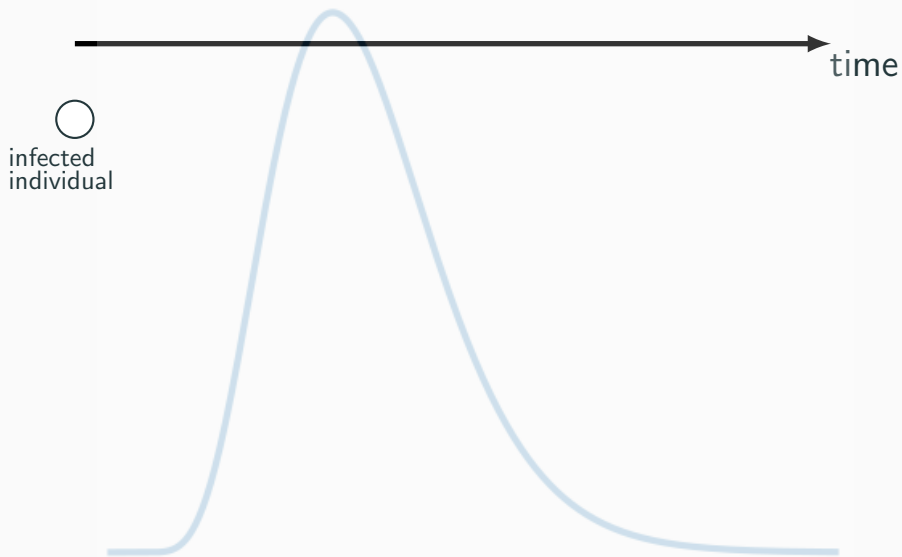
## Mean epidemic size



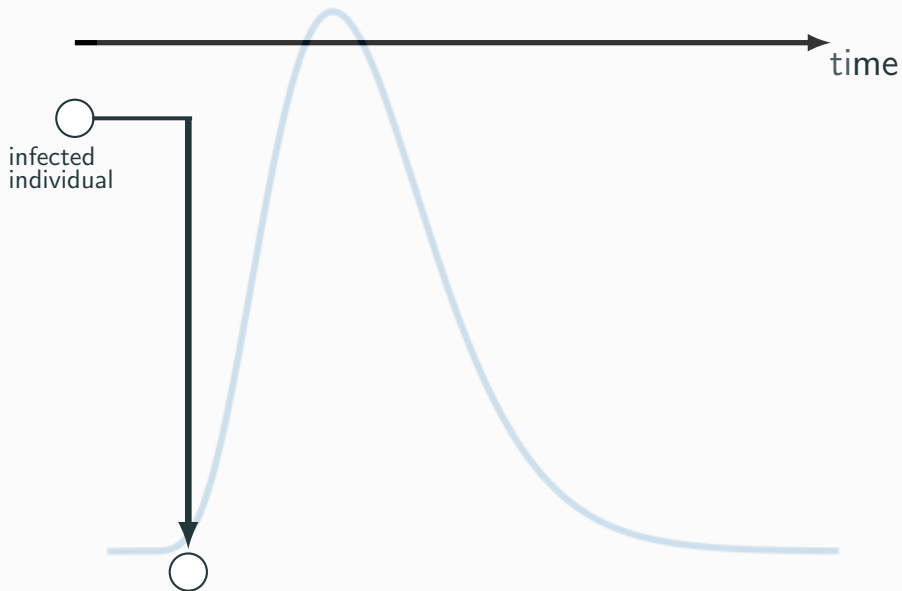
# Epidemiological model



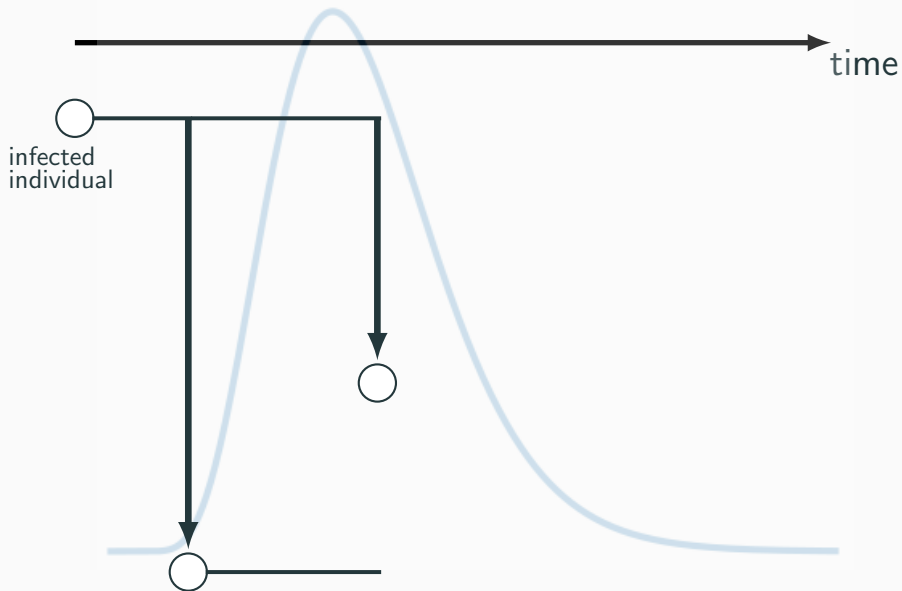
# Epidemiological model



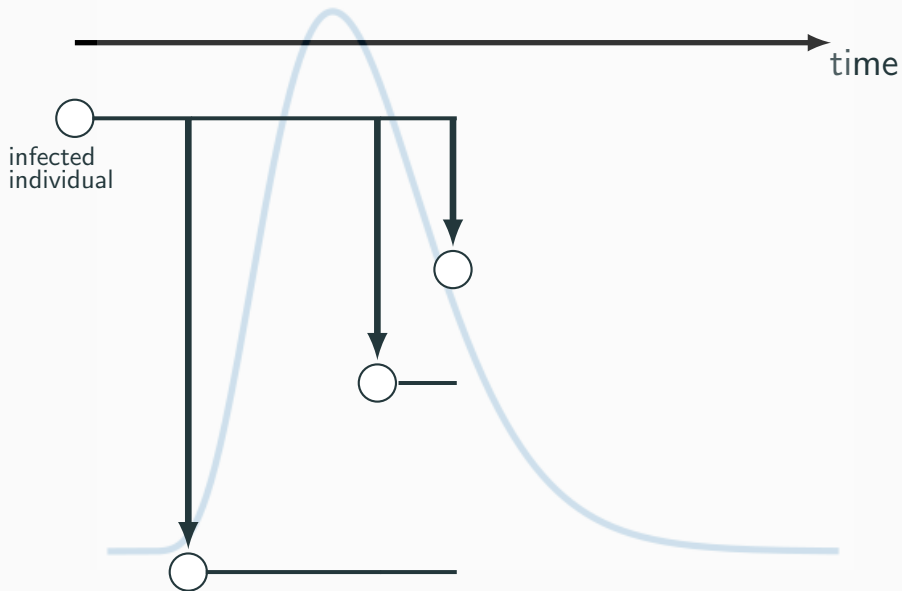
# Epidemiological model



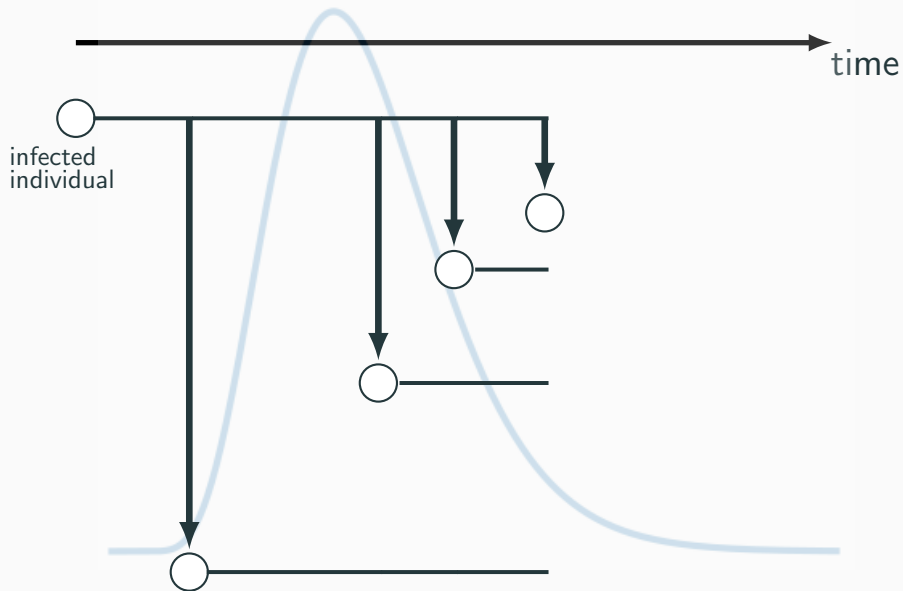
# Epidemiological model



# Epidemiological model

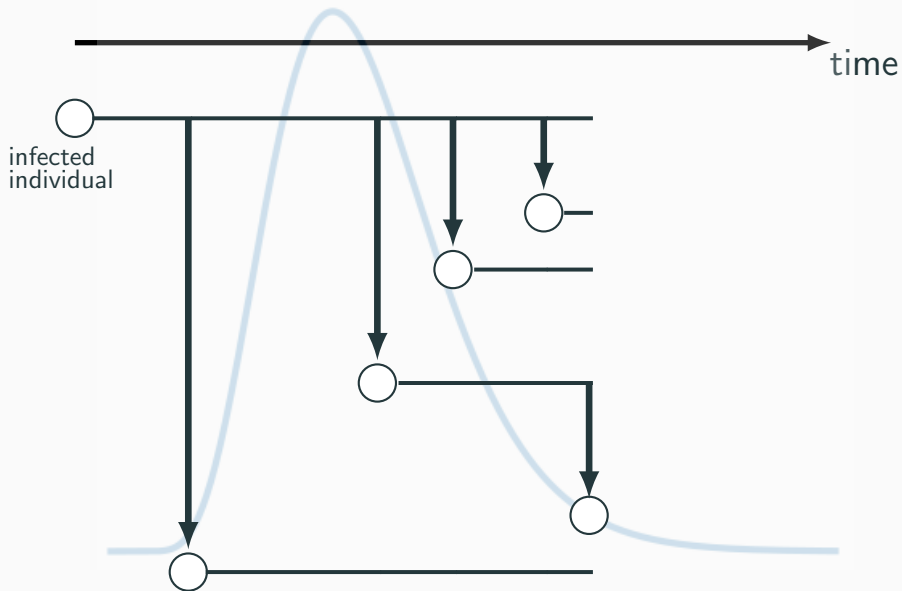


# Epidemiological model

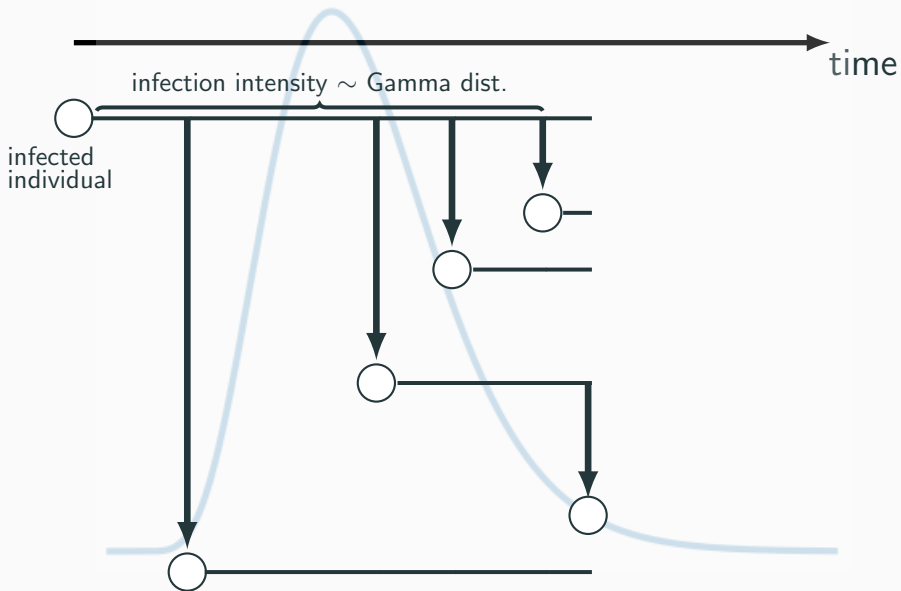




# Epidemiological model



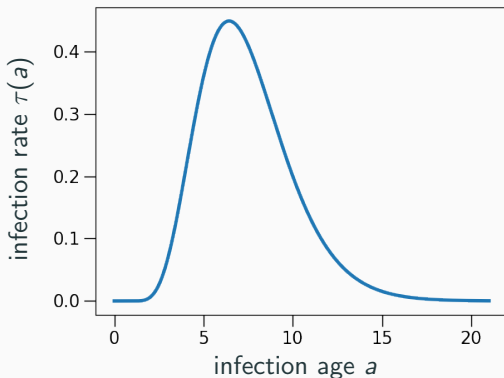
# Epidemiological model



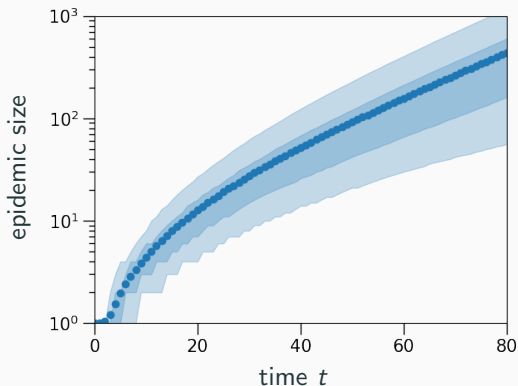
# Infection kernel

Infected individuals ...

- ... are characterized by their **infection age**  $a$
- ... transmit the disease with **rate**  $\tau(a) = R \times \mu(a)$ , where  $\mu$  is the probability density of a Gamma distribution  
→ parameters from Hinch et al. *medRxiv preprint* (2020+) – OpenABM-Covid19 from Oxford Modeling Group



## Simulations ( $R = 1.3$ , Poisson offspring distribution)



- ● = simulation mean
- dark shade = 50% inter-quantile range
- light shade = 90% inter-quantile range

## Mean epidemic size

- $I(t)$  = mean epidemic size of the individual-based process

Renewal equation:

$$I(t) = I(0) + \int_0^t I(t-a)\tau(a)da$$

# Mean epidemic size

- $I(t)$  = mean epidemic size of the individual-based process

Renewal equation:

$$I(t) = I(0) + \int_0^t I(t-a)\tau(a)da$$

How to think about this equation?

- $I(t-a)$  is the mean cluster size of a cluster that emerged at time  $a$
- ... this happens at rate  $\tau(a)$

# Mean epidemic size

- $I(t)$  = mean epidemic size of the individual-based process

Renewal equation:

$$I(t) = I(0) + \int_0^t I(t-a)\tau(a)da$$

How to think about this equation?

- $I(t-a)$  is the mean cluster size of a cluster that emerged at time  $a$
- ... this happens at rate  $\tau(a)$

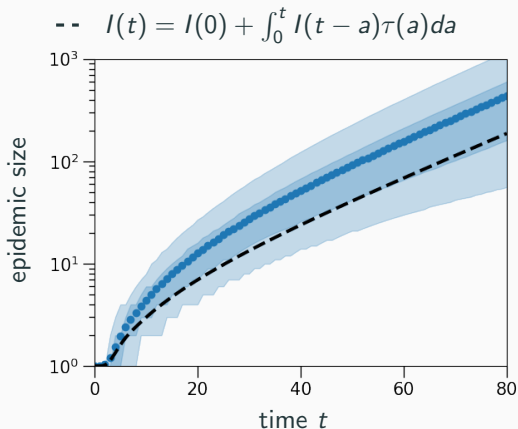
In a discrete world:

$$I(1) = I(0) + \tau(1)I(0)$$

$$I(2) = I(0) + \tau(1)I(1) + \tau(2)I(0)$$

$$I(3) = I(0) + \underbrace{\tau(1)I(2) + \tau(2)I(1) + \tau(3)I(0)}_{\text{the integral in continuous time}}$$

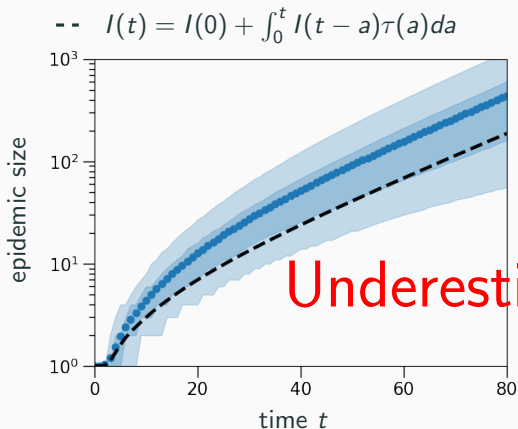
## Simulations ( $R = 1.3$ , Poisson offspring distribution)



- ● = simulation mean
- dark shade = 50% inter-quantile range
- light shade = 90% inter-quantile range



## Simulations ( $R = 1.3$ , Poisson offspring distribution)



- • = simulation mean
- dark shade = 50% inter-quantile range
- light shade = 90% inter-quantile range

## Initial growth - stochastic correction

Underestimation because we do not take into account that the epidemic establishes! (mean also reflects non-establishing epidemics)

Solution: **conditioning on non-extinction** in the infection rates

→ Method: **Doob's h-transform** (Doob (1957))

## Initial growth - stochastic correction

Underestimation because we do not take into account that the epidemic establishes! (mean also reflects non-establishing epidemics)

Solution: **conditioning on non-extinction** in the infection rates

→ Method: **Doob's h-transform** (Doob (1957))

$$\tilde{\tau}(a, t) = \tau(a) \left( \frac{1 - p_{\text{ext}}(0) \prod_{j=1}^{N(t)} p_{\text{ext}}(a_j)}{1 - \prod_{j=1}^{N(t)} p_{\text{ext}}(a_j)} \right),$$

where  $N(t)$  = the number of infected individuals in the stochastic process and

$$p_{\text{ext}}(a) = p_{\text{ext}}(0) \exp \left( (1 - p_{\text{ext}}(0)) \int_0^a \tau(s) ds \right)$$

is the extinction probability of an epidemic that started with an infected individual at age  $a$ .

## Initial growth - stochastic correction

Underestimation because we do not take into account that the epidemic establishes! (mean also reflects non-establishing epidemics)

Solution: **conditioning on non-extinction** in the infection rates

→ Method: **Doob's h-transform** (Doob (1957))

$$\tilde{\tau}(a, t) = \tau(a) \left( \frac{1 - p_{\text{ext}}(0) \prod_{j=1}^{N(t)} p_{\text{ext}}(a_j)}{1 - \prod_{j=1}^{N(t)} p_{\text{ext}}(a_j)} \right),$$

where  $N(t)$  = the number of infected individuals in the stochastic process and

$$p_{\text{ext}}(a) = p_{\text{ext}}(0) \exp \left( (1 - p_{\text{ext}}(0)) \int_0^a \tau(s) ds \right)$$

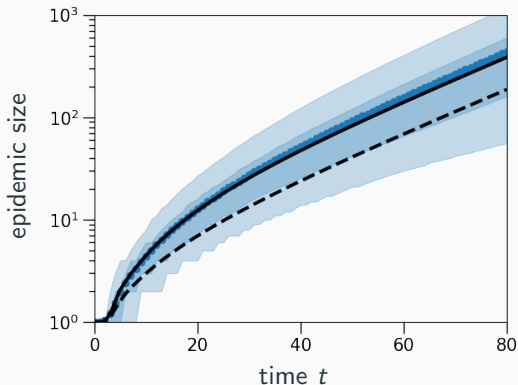
is the extinction probability of an epidemic that started with an infected individual at age  $a$ .

**Caveat:** This only works nicely in the case of a Poisson offspring distribution.

## Simulations ( $R = 1.3$ , Poisson offspring distribution)

$$\text{--- } \tilde{I}(t) = I(0) + \int_0^t \tilde{I}(t-a) \tilde{\tau}(a, t) da$$

$$\text{--- } I(t) = I(0) + \int_0^t I(t-a) \tau(a) da$$



- ● = simulation mean
- dark shade = 50% inter-quantile range
- light shade = 90% inter-quantile range

## Summary of the theory

- **Establishment probability** depends on the offspring distribution
  - ⇒ the larger the variance, the smaller the establishment probability
  - ⇒ a reasonable proxy is  $1/R_0$

- **Epidemic size** is described by a renewal equation

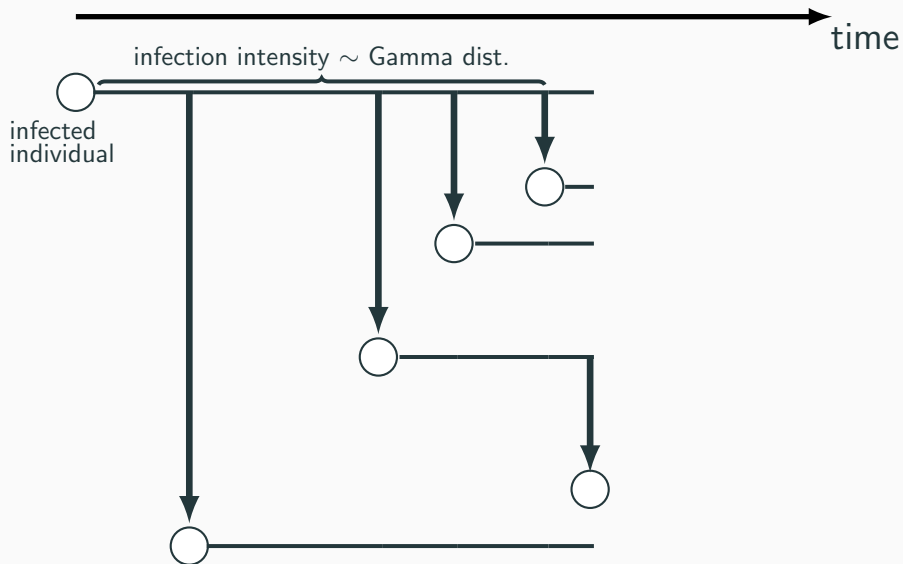
$$I(t) = I(0) + \int_0^t I(t-a)\tau(a)da$$

- To account for initial stochasticity, the infection measure  $\tau(a)$  can be corrected by **conditioning the stochastic process on establishment**

## Detection time and size

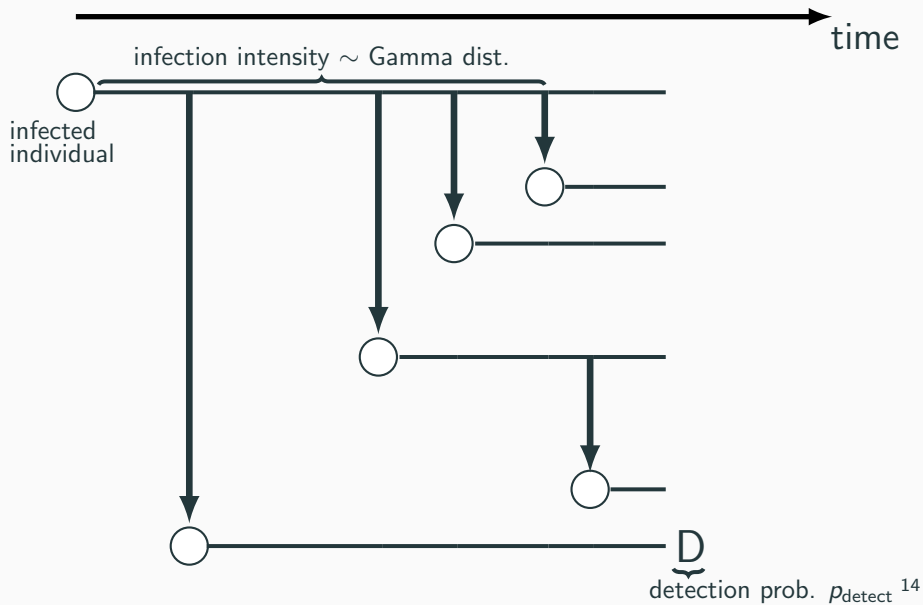
---

# Epidemiological model with detection

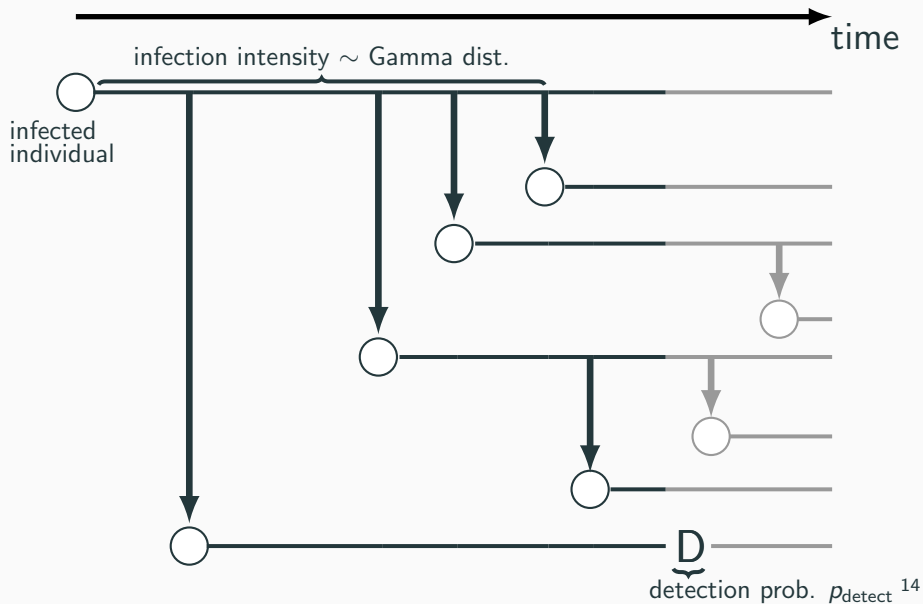




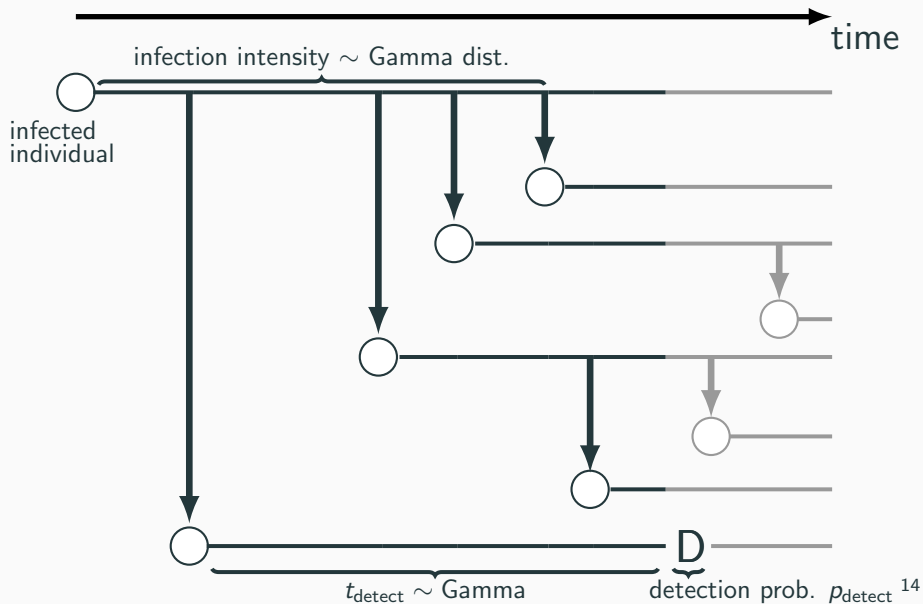
# Epidemiological model with detection



# Epidemiological model with detection



# Epidemiological model with detection



**How long does it take to  
detect the first case?**

## Time of the first detection

- Number of infected individuals until the first one will be detected =  
 $K \sim \text{Geom}(p_{\text{detect}})$

## Time of the first detection

- Number of infected individuals until the first one will be detected =  
 $K \sim \text{Geom}(p_{\text{detect}})$

Approximation procedure of the **first detection time**:

1. estimate deterministic time  $t_K$  so that  $\tilde{I}(t_K) = K$

# Time of the first detection

- Number of infected individuals until the first one will be detected =  
 $K \sim \text{Geom}(p_{\text{detect}})$

Approximation procedure of the **first detection time**:

1. estimate deterministic time  $t_K$  so that  $\tilde{I}(t_K) = K$
2. add the waiting time until detection:

$$\text{first detection} = t_K + \underbrace{t_{\text{detect}}}_{\sim \text{Gamma}}$$

# Time of the first detection

- Number of infected individuals until the first one will be detected =  
 $K \sim \text{Geom}(p_{\text{detect}})$

Approximation procedure of the **first detection time**:

1. estimate deterministic time  $t_K$  so that  $\tilde{I}(t_K) = K$
2. add the waiting time until detection:

$$\text{first detection} = t_K + \underbrace{t_{\text{detect}}}_{\sim \text{Gamma}}$$

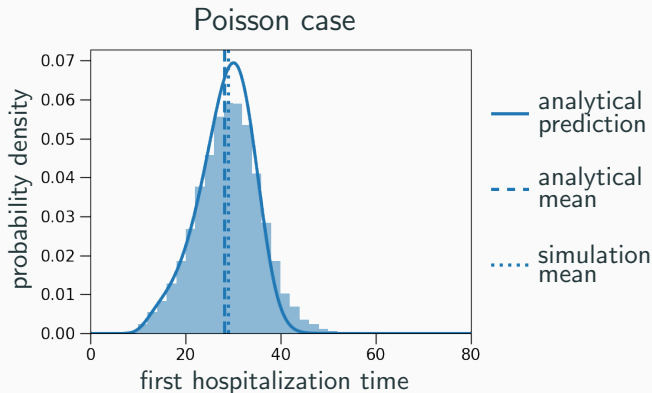
3. probability density of the first detection time (**analytical prediction**):

$$h(t) = \sum_{K=1}^{\infty} \underbrace{p_{\text{detect}} (1 - p_{\text{detect}})^{K-1}}_{\text{geometric dist.}} \underbrace{f_{\text{detect}}(t - t_K)}_{\text{Gamma dist.}}$$



## Example: First hospitalization time

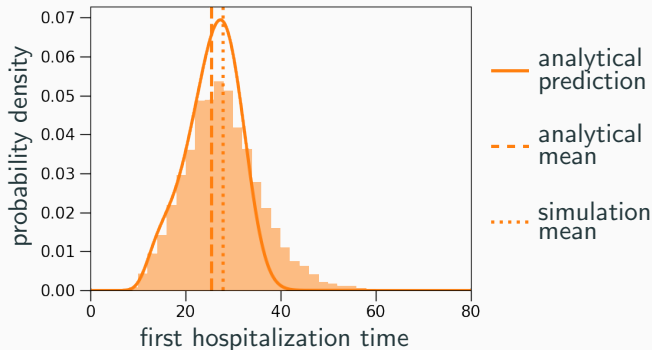
- Hospitalization probability:  $p_{\text{detect}} = 0.029$  (Salje et al. *Science* (2020))
- Time from infection to hospitalization: Gamma(31,0.463) (mean = 14.4 days) (Foutel-Rodier *arXiv preprint* (2020+))
- reproduction number:  $R = 2.7$  (March 2020 in France)



## Example: First hospitalization time

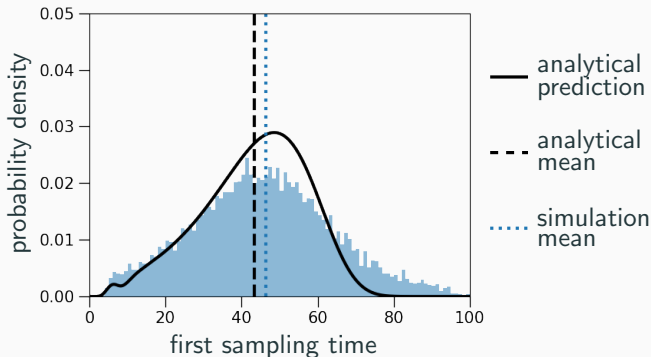
- Hospitalization probability:  $p_{\text{detect}} = 0.029$  (Salje et al. *Science* (2020))
- Time from infection to hospitalization: Gamma(31,0.463) (mean = 14.4 days) (Foutel-Rodier *arXiv preprint* (2020+))
- reproduction number:  $R = 2.7$  (March 2020 in France)

### Negative binomial case



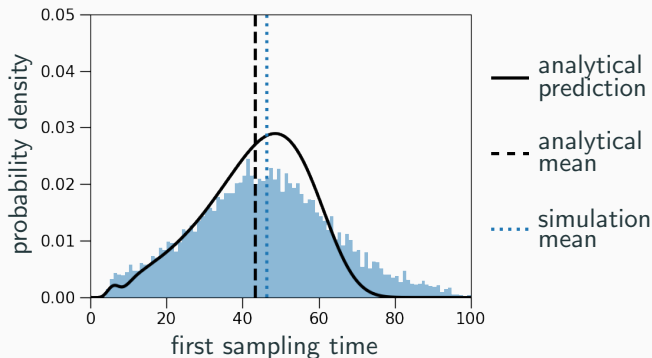
## Example: Variant B.1.1.7

- detection probability = sampling rate in the UK in September 2020 = 0.0105 (according to COVID-19 Genomics Consortium UK + underreporting rate in Colman et al. *medRxiv* (2021+))
- reproduction number of the variant = 1.5 (Volz et al. *medRxiv* (2021+))
- time from infection to sampling: Gamma(12,7/12) (mean: 7 days)



## Example: Variant B.1.1.7

- detection probability = sampling rate in the UK in September 2020 = 0.0105 (according to COVID-19 Genomics Consortium UK + underreporting rate in Colman et al. *medRxiv* (2021+))
- reproduction number of the variant = 1.5 (Volz et al. *medRxiv* (2021+))
- time from infection to sampling: Gamma(12,7/12) (mean: 7 days)



⇒ Variant B.1.1.7 emergence around the 4<sup>th</sup> of August 2020

**How large is the epidemic  
cluster at the first  
detection?**

## Size of the first detection

The probability distribution of the size of the epidemic cluster at its first detection is

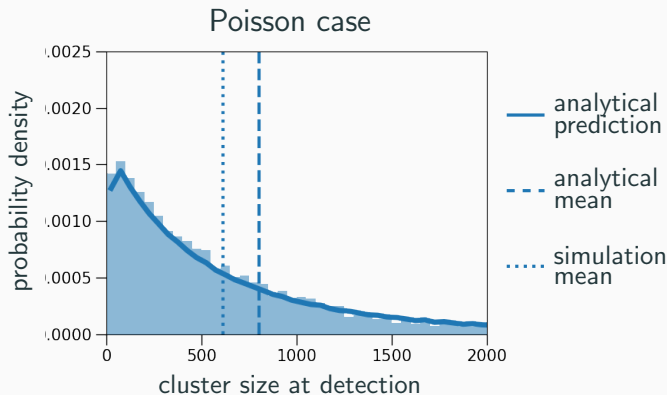
$$P(\text{size at detection} = k) = \int_0^{\infty} h(t) \mathbb{1}_{\tilde{I}(t) \in [k-1/2, k+1/2]} dt$$

## Size of the first detection

The probability distribution of the size of the epidemic cluster at its first detection is

$$P(\text{size at detection} = k) = \int_0^{\infty} h(t) \mathbb{1}_{\tilde{I}(t) \in [k-1/2, k+1/2]} dt$$

Example: **Size at first hospitalization**

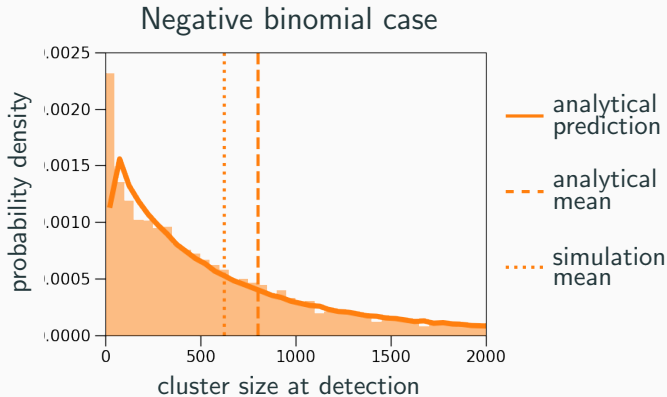


## Size of the first detection

The probability distribution of the size of the epidemic cluster at its first detection is

$$P(\text{size at detection} = k) = \int_0^{\infty} h(t) \mathbb{1}_{\tilde{I}(t) \in [k-1/2, k+1/2]} dt$$

Example: **Size at first hospitalization**





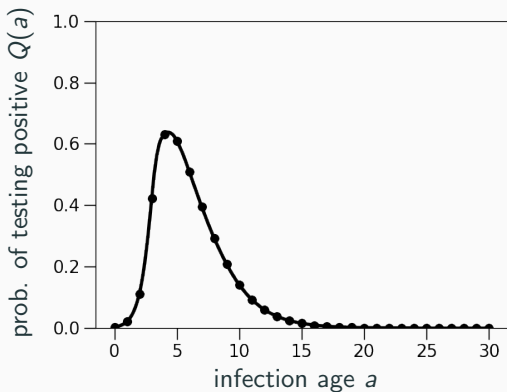
# Testing



**What is the minimal  
random testing frequency  
to detect clusters before  
they exceed a certain  
threshold size?**

## Probability of infected individuals to test positive

- Probability to test positive, e.g. by a rapid test, depends on infection age  $a$  of the tested individual (Hellewell et al. *BMC Medicine* (2021))
- Denote this probability by  $Q(a)$



# Minimal testing frequency to detect clusters of a certain size

**Assumption:** Individuals in a population (susceptibles and infecteds) are randomly tested, independently of their infection status

- $f$  = fraction of individuals that is tested per day
- Infected individual is detected with probability

$$p_{\text{detect}} = 1 - \prod_{a=1}^{\infty} \underbrace{(1 - fQ(a))}_{\text{prob. of not being detected at age } a} \approx f \sum_{a=1}^{\infty} Q(a)$$

# Minimal testing frequency to detect clusters of a certain size

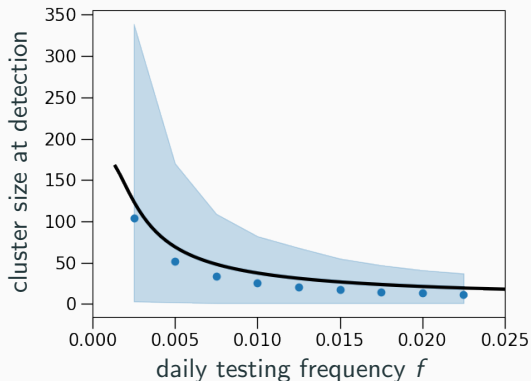
**Assumption:** Individuals in a population (susceptibles and infecteds) are randomly tested, independently of their infection status

- $f$  = fraction of individuals that is tested per day
- Infected individual is detected with probability

$$p_{\text{detect}} = 1 - \prod_{a=1}^{\infty} \underbrace{(1 - fQ(a))}_{\text{prob. of not being detected at age } a} \approx f \sum_{a=1}^{\infty} Q(a)$$

- Now apply the results from before (size of a cluster at its first detection) and 'solve' (numerically) for the minimal value of  $f$

## Minimal testing frequency and resulting cluster size ( $R = 1.1$ )



In numbers: threshold size of 30  $\Rightarrow f = 0.013$

- France:  $\sim 870,000$  tests per day (currently  $\sim 320,000$ )
- Germany:  $\sim 1,080,000$  tests per day (currently  $\sim 185,000$ )
- (according to data from ECDC)

# Conclusion

---

## Conclusion

- **Renewal equation** derived from the stochastic process conditioned on establishment is a good description of the epidemic size



# Conclusion

- **Renewal equation** derived from the stochastic process conditioned on establishment is a good description of the epidemic size
- **Probability distribution of the first detection time** depends on offspring distribution, the epidemic parameters and most importantly on the detection probability per infected individual  
⇒ Variant of concern B.1.1.7 most likely started circulating early August

# Conclusion

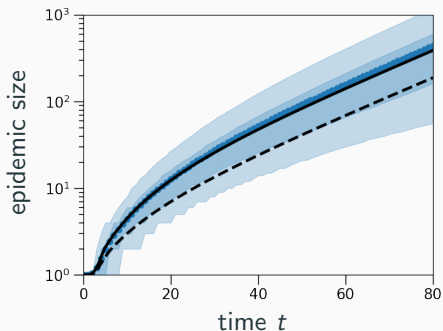
- **Renewal equation** derived from the stochastic process conditioned on establishment is a good description of the epidemic size
- **Probability distribution of the first detection time** depends on offspring distribution, the epidemic parameters and most importantly on the detection probability per infected individual  
⇒ Variant of concern B.1.1.7 most likely started circulating early August
- The **probability distribution of the size of a cluster at its first detection** is right-skewed (geometric?)

# Conclusion

- **Renewal equation** derived from the stochastic process conditioned on establishment is a good description of the epidemic size
- **Probability distribution of the first detection time** depends on offspring distribution, the epidemic parameters and most importantly on the detection probability per infected individual  
⇒ Variant of concern B.1.1.7 most likely started circulating early August
- The **probability distribution of the size of a cluster at its first detection** is right-skewed (geometric?)
- The **minimal testing frequency** to detect clusters can be approximated by combining the probability to test positive with the detection probability  
⇒ might be relevant in a Zero-Covid-scenario

## Open questions (at least to me)

- Is there a way to estimate the variance of the limit process (something like a central limit theorem for the convergence of the conditioned individual-based model to the McKendrick-von Foerster equation)
- Or alternatively: the variance of the asymptotic growth rate of the super-critical branching process?



**Questions?**

**Thank you for your  
attention!**

## Exponential growth, asymptotic growth

The **exponential growth rate** ( $r$ ) of the epidemic is computed as follows

$$1 = \int_0^{\infty} e^{-ra} \tau(a) da = R \int_0^{\infty} e^{-ra} \mu(a) da$$

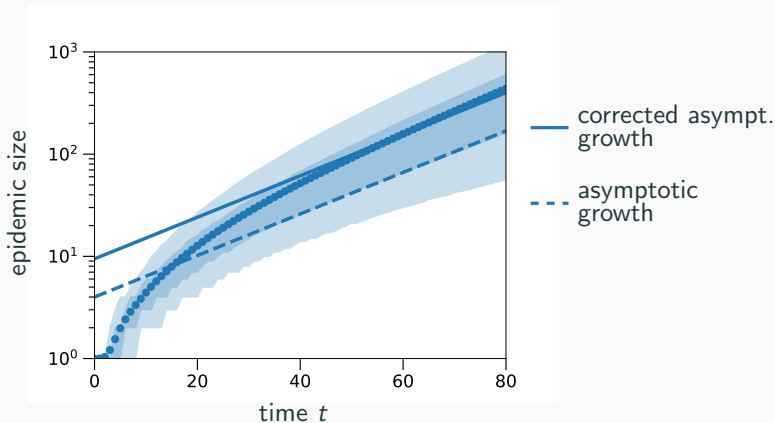
The **asymptotic growth rate** of the epidemic can be derived from the renewal equation and is

$$I(t) = I(0) \frac{e^{rt}}{r \int_0^{\infty} R e^{-ra} a \mu(a) da} \quad \text{for } t \rightarrow \infty$$

The **'corrected' asymptotic growth rate** is

$$I(t) = \frac{I(0)}{p_{\text{estab}}} \frac{e^{rt}}{r \int_0^{\infty} R e^{-ra} a \mu(a) da} \quad \text{for } t \rightarrow \infty$$

## Asymptotic growth ( $R = 1.3$ )



The exponential growth implies that the **stationary infection age distribution** is an **exponential distribution** with parameter  $r$ .

## Doob's h-transform

*Infinitesimal generator* of the original process:

$$Lf(a) = \sum_{i=1}^{|a|} \left[ \tau(a_i) \left( f((a, 0)) - f(a) \right) + \partial_i f(a) \right]$$

Probability of extinction of a cluster seeded by a single infected individual with infection age  $a$  (and Poisson offspring distribution):

$$p(a) = p(0) \exp \left( (1 - p(0)) \int_0^a \tau(s) ds \right)$$

*Harmonic function* for the infinitesimal generator:

$$h(a) := 1 - \prod_{i=1}^{|a|} p(a_i) \quad \Rightarrow \quad Lh(a) = 0$$

The infinitesimal generator of the transformed process is then given by:

$$\tilde{L}f = \frac{1}{h} L(fh)$$



# Detection rate of a single mass testing effort

How many infected individuals can be detected by a single mass testing effort, as for example performed in Slovakia in fall 2020?

## Approximation:

- Assumption: infection age distribution has reached stationarity
- Detection probability is given by  $Q(a)$
- Proportion of detection among the (cumulative) infected individuals:

$$I_{\text{detected}}(t) = qI(t) = I(t) \int_0^{\infty} Q(a) r e^{-ra} da$$

# Detection rate of a single mass testing effort

How many infected individuals can be detected by a single mass testing effort, as for example performed in Slovakia in fall 2020?

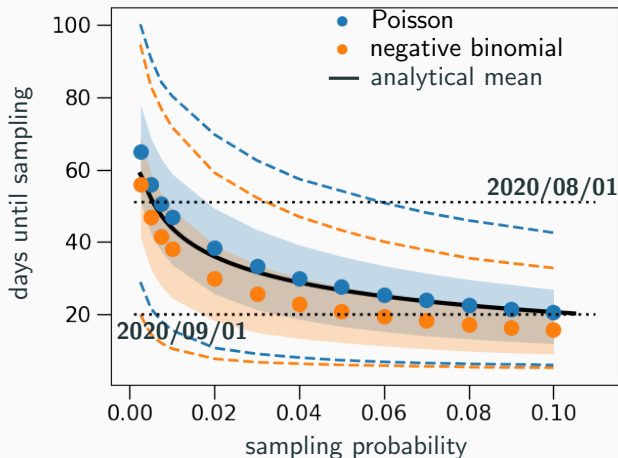
## Approximation:

- Assumption: infection age distribution has reached stationarity
- Detection probability is given by  $Q(a)$
- Proportion of detection among the (cumulative) infected individuals:

$$I_{\text{detected}}(t) = qI(t) = I(t) \int_0^{\infty} Q(a)re^{-ra} da$$

- $R = 1.3$ :  $q = 0.54$  (with rapid tests:  $q = 0.26$ )
- $R = 2.7$ :  $q = 0.71$  (with rapid tests:  $q = 0.26$ )

## Varying the detection probability of B.1.1.7



- shaded region = 50% inter-quantile range
- dashed line = 5- and 95-percentile