

Physics of Life Data Epidemiology

Lect 6: Outbreak analysis - surveillance

Chiara Poletto

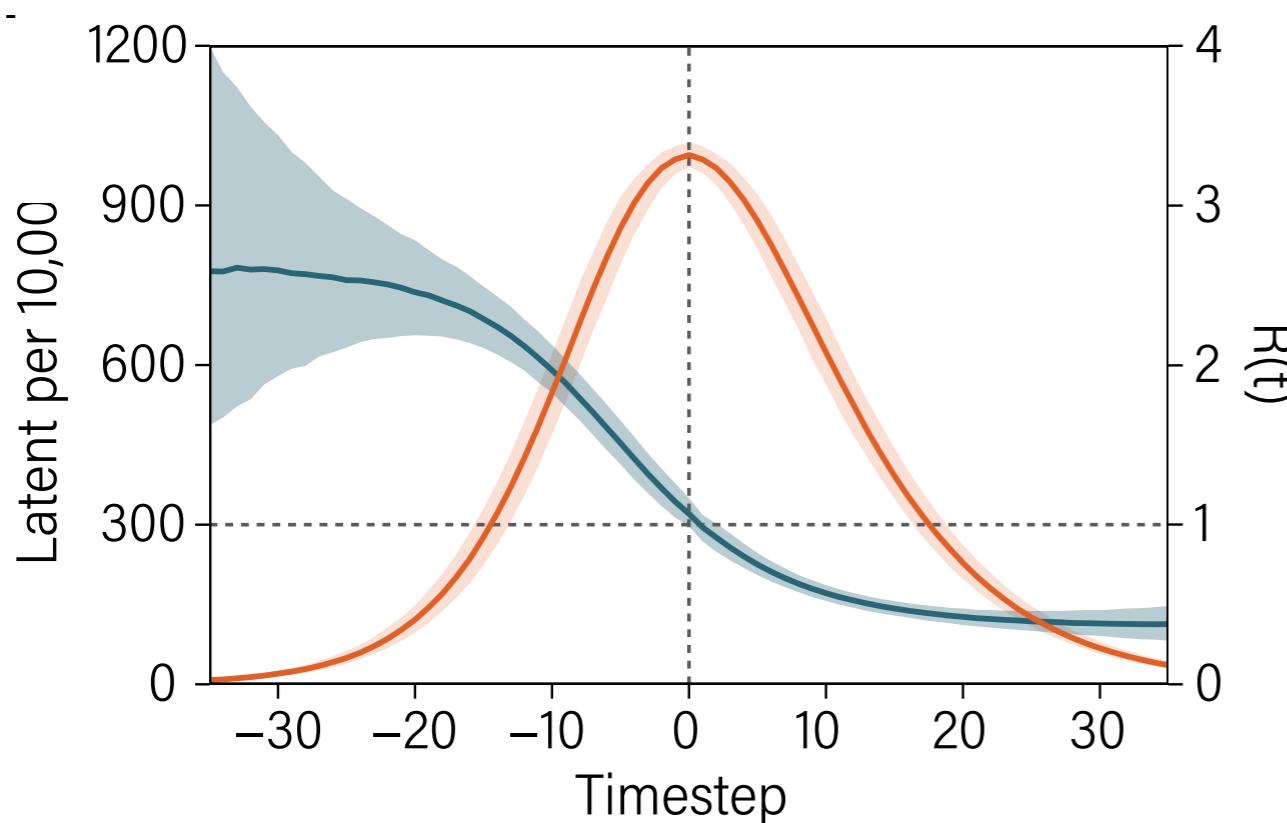
mail: chiara.poletto@unipd.it

web: chiara-poletto.github.io

bsky: @chpoletto.bsky.social

index R_t

Reproductive ratio $R(t)$: # of cases generated by a case in a *partially susceptible* population



[Lu et al doi.org/10.1101/2021.03.19.21253974]
 $\beta = \text{const}, \langle k \rangle_t = \text{const}, \tau = \text{const}$
 $\Rightarrow R(t) = R_0 s(t)$



[<https://www.gouvernement.fr/info-coronavirus/carte-et-donnees>]

$\beta_t, \langle k \rangle_t, \tau_t$ variable in time

index R_t

how to compute it?

[Bettencourt LMA, Ribeiro RM (2008) Real Time Bayesian Estimation of the Epidemic Potential of Emerging Infectious Diseases. PLoS ONE 3(5): e2185. <https://doi.org/10.1371/journal.pone.0002185>]

Departing from the Kermack and McKendrick model

[Cori, Ferguson, Fraser, Cauchemez, A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics, American Journal of Epidemiology, 178, 2013]

Departing from the renewal equation

[Wallinga, Teunis, Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures, American Journal of Epidemiology, 160, 6, 2004]

Cohort reproduction number

index R_t

[Bettencourt LMA, Ribeiro RM (2008) Real Time Bayesian Estimation of the Epidemic Potential of Emerging Infectious Diseases. PLoS ONE 3(5): e2185. <https://doi.org/10.1371/journal.pone.0002185>]

Departing from the Kermack and McKendrick model

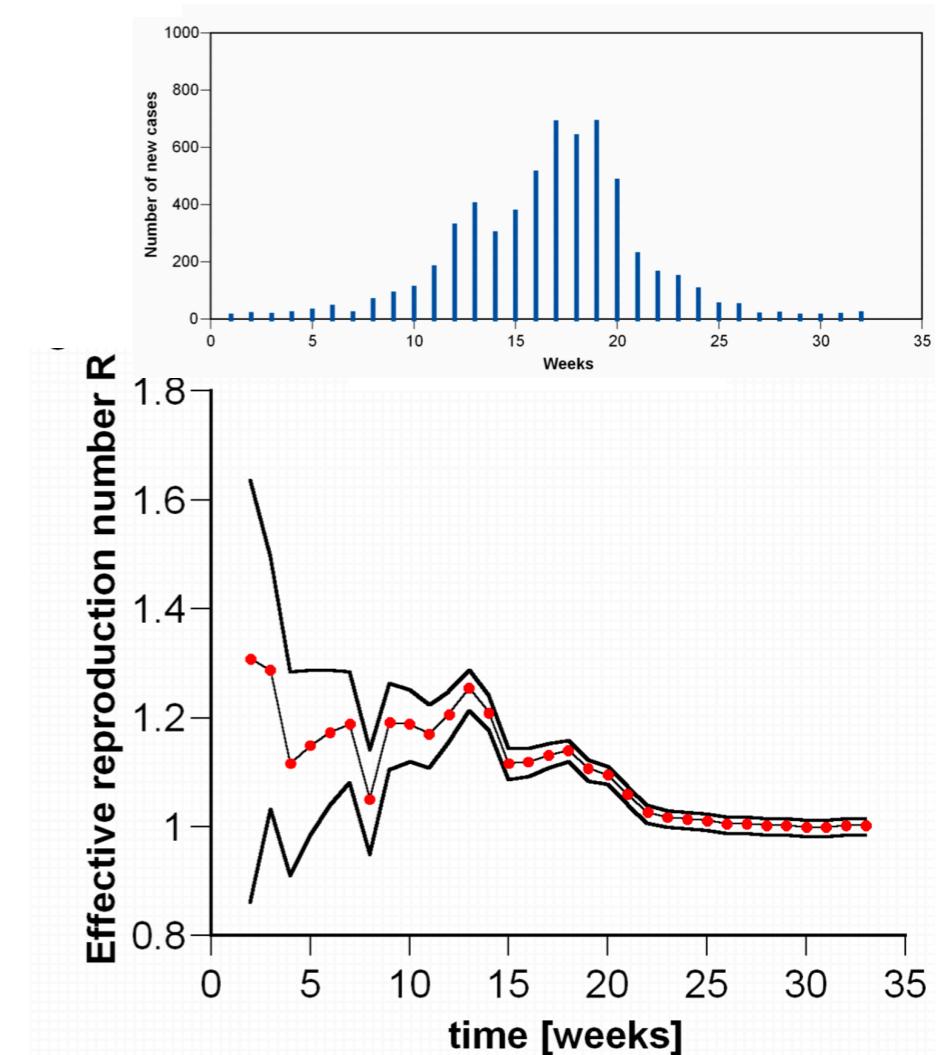
$$I(t + \delta t) = I(t) \exp \left[\delta t \mu \left(R_0 \frac{S(t)}{N} - 1 \right) \right]$$

more generic formulation (accounting for all possible factors that could make R_t changing)

$$I(t + \delta t) \simeq I(t) \exp \left[\delta t \mu (R_t - 1) \right]$$

if $R_t \simeq \text{const}$ during δt , i.e. incidence small

$$R_t \simeq \frac{\ln I(t + \delta t) - \ln I(t)}{\delta t \mu} + 1 \quad (\text{crude approximation})$$



[Bettencourt et al PLOS ONE 2008]

index R_t

[Cori, Ferguson, Fraser, Cauchemez, A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics, American Journal of Epidemiology, 178, 2013]

Departing from the **renewal equation**

-> We are assuming a case is registered the date in which it was infected

generic generation time distribution, $w(\tau)$: Lotka Euler equation

$$I(t) = \int_0^{\infty} I(t - \tau) \beta(\tau) d\tau \text{ with } \beta(\tau) = w(\tau) R_0$$

assume reproductive ratio varies in time $\beta(\tau, t) = w(\tau) R_t$

$$I(t) = \int_0^{\infty} I(t - \tau) w(\tau) R_t d\tau$$

index R_t

[Cori, Ferguson, Fraser, Cauchemez, A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics, American Journal of Epidemiology, 178, 2013]

$$R_t = \frac{I(t)}{\int_0^\infty I(t-\tau)w(\tau) d\tau}$$

we write the equation in discrete time

$$R_t = \frac{I_t}{\sum_{s=1}^t I_{t-s} w_s}$$

$\sum_{s=1}^t I_{t-s} w_s$: total infectiousness of individuals that are infectious at the time t

R_t : **average number of secondary cases that each infected individual will infect if the conditions remain as they are at the time t**

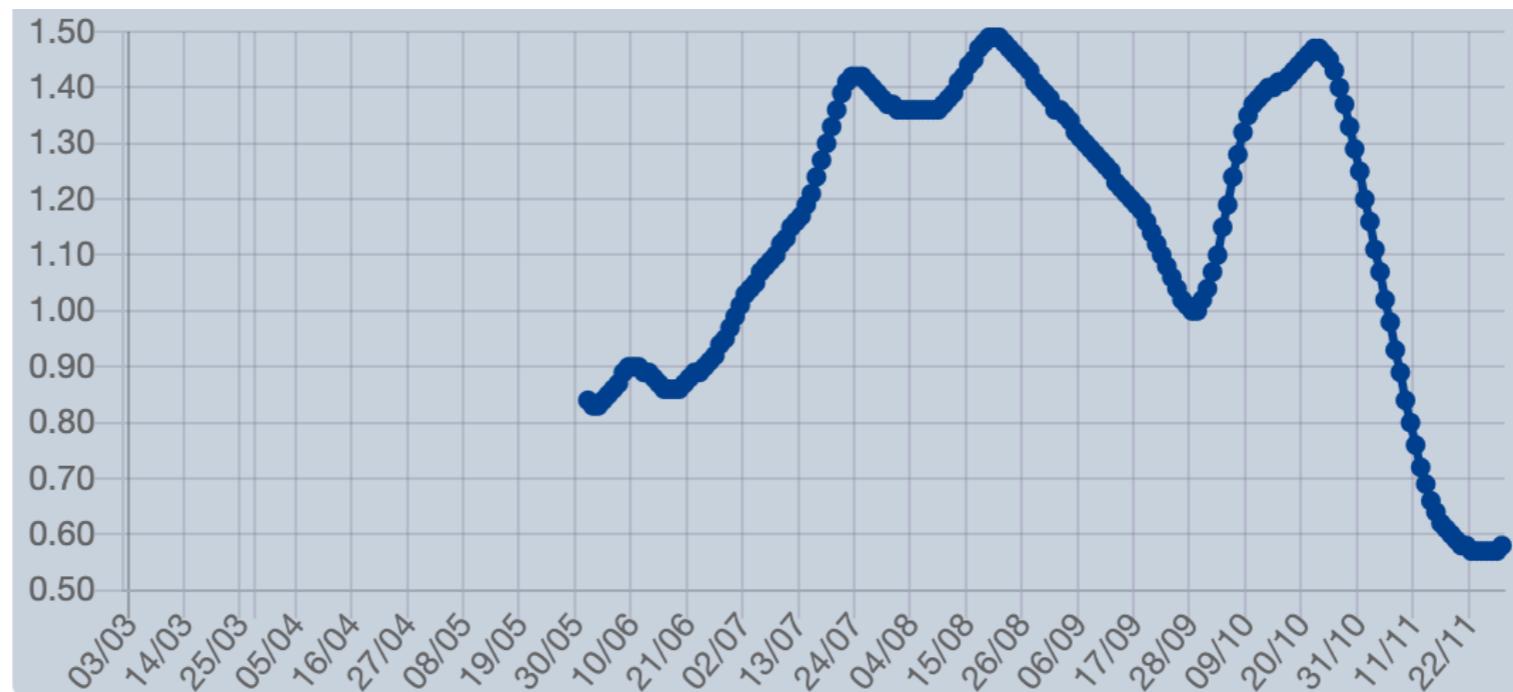
I describe the number of new infections at time t in relation to past infections

index R_t

[Cori, Ferguson, Fraser, Cauchemez, A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics, American Journal of Epidemiology, 178, 2013]

$$R_t = \frac{I_t}{\sum_{s=1}^t I_{t-s} w_s}$$

R_t for France



[<https://www.gouvernement.fr/info-coronavirus/carte-et-donnees>]

index R_t

[Wallinga, Teunis, Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures, American Journal of Epidemiology, 160, 6, 2004]

Infer “who infected whom” from available information. When we have an incidence curve the only information regarding a case is the date in which a case was recorded

p_{ij} = relative probability that case i is infected by case j (given $t_i - t_j$) depends on the generation interval (*assuming a case is registered the date in which it was infected*)

$$p_{ij} = \frac{w(t_i - t_j)}{\sum_{i \neq k} w(t_i - t_k)}$$

$$R_j = \sum_i p_{ij}$$

cohort reproduction number: it counts the average number of secondary transmissions caused by a cohort infected at time step t

index R_t

[Wallinga, Teunis, Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures, American Journal of Epidemiology, 160, 6, 2004]

Infer “who infected whom” from available information. When we have an incidence curve the only information regarding a case is the date in which a case was recorded

p_{ij} = relative probability that case i is infected by case j (given $t_i - t_j$) depends on the generation interval (assuming a case is registered the date in which it was infected)

$$p_{ij} = \frac{w(t_i - t_j)}{\sum_{i \neq k} w(t_i - t_k)}$$

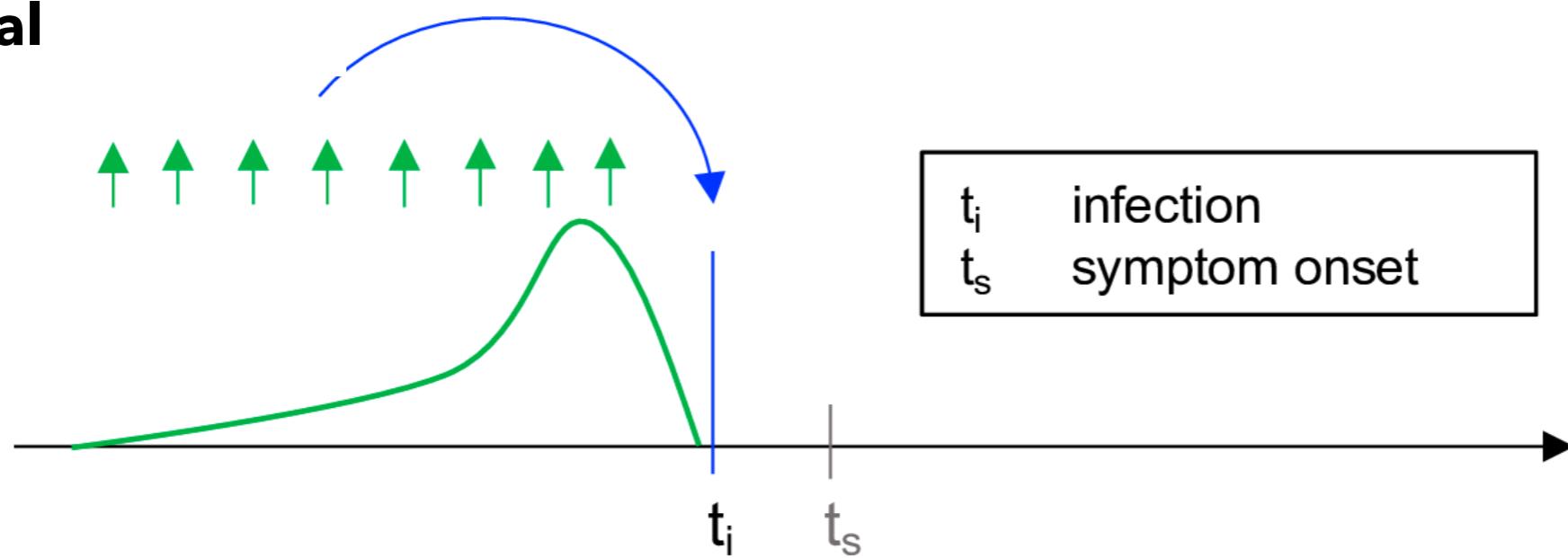
$$R_j = \sum_i p_{ij}$$

the method can be used only in retrospect, once the secondary cases generated by the infected at time t are infected

index R_t

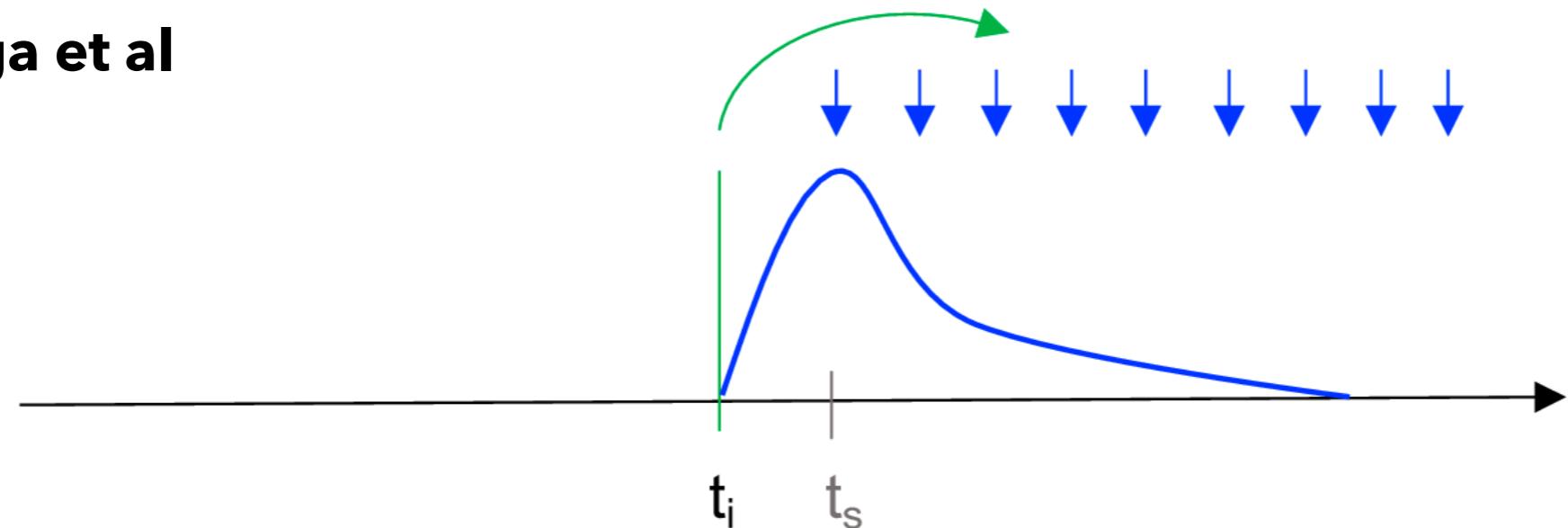
[Gostic KM, McGough L, Baskerville EB, Abbott S, Joshi K, Tedijanto C, et al. (2020) Practical considerations for measuring the effective reproductive number, R_t . PLoS Comput Biol 16(12): e1008409]

Instantaneous R_t - Cori et al



I describe number of new infections at time t in relation to past infections. I look backward in time

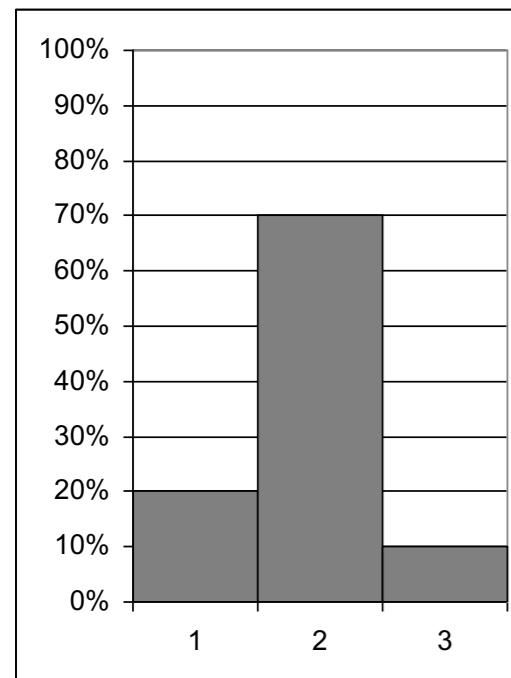
Case R_t - Wallinga et al



I relate number of new infections at time t to the future cases they generate. I look forward in time

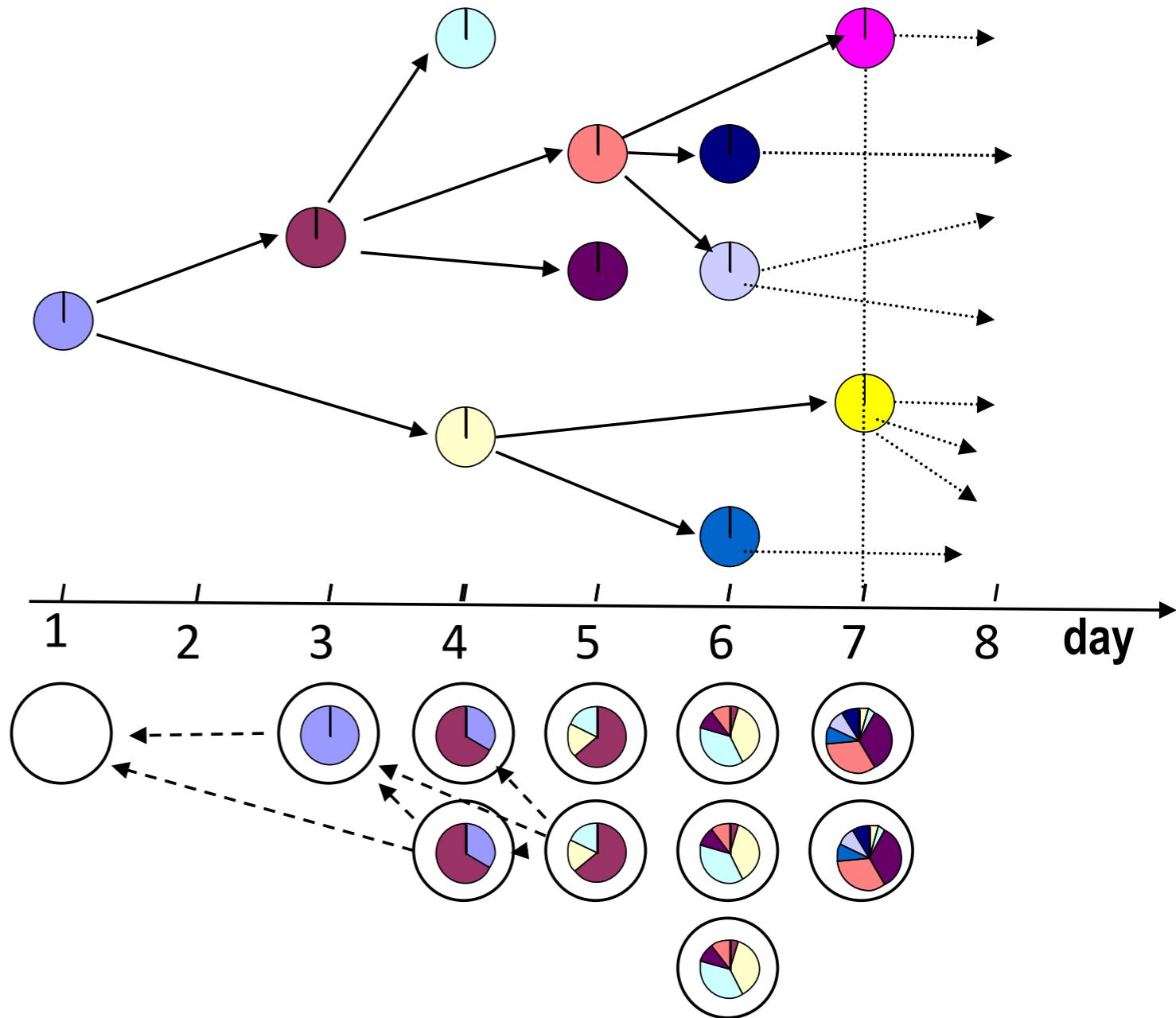
index R_t

generation time



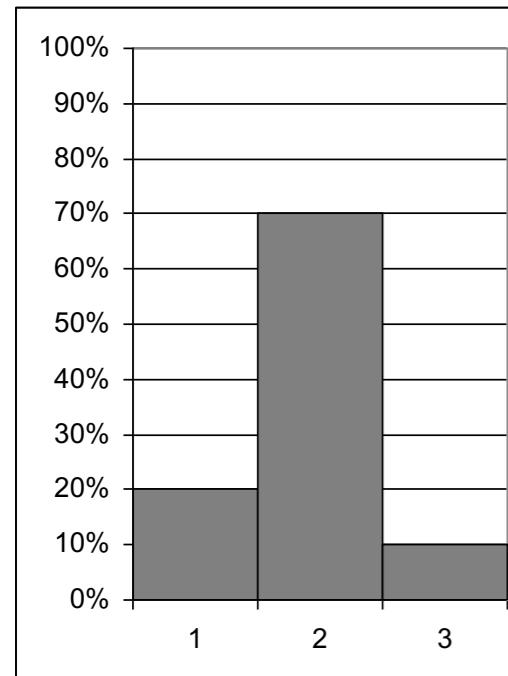
Real
transmission
chain
(*unobserved*)

Incidence, i.e.
untraced
outbreak
(*observed*)



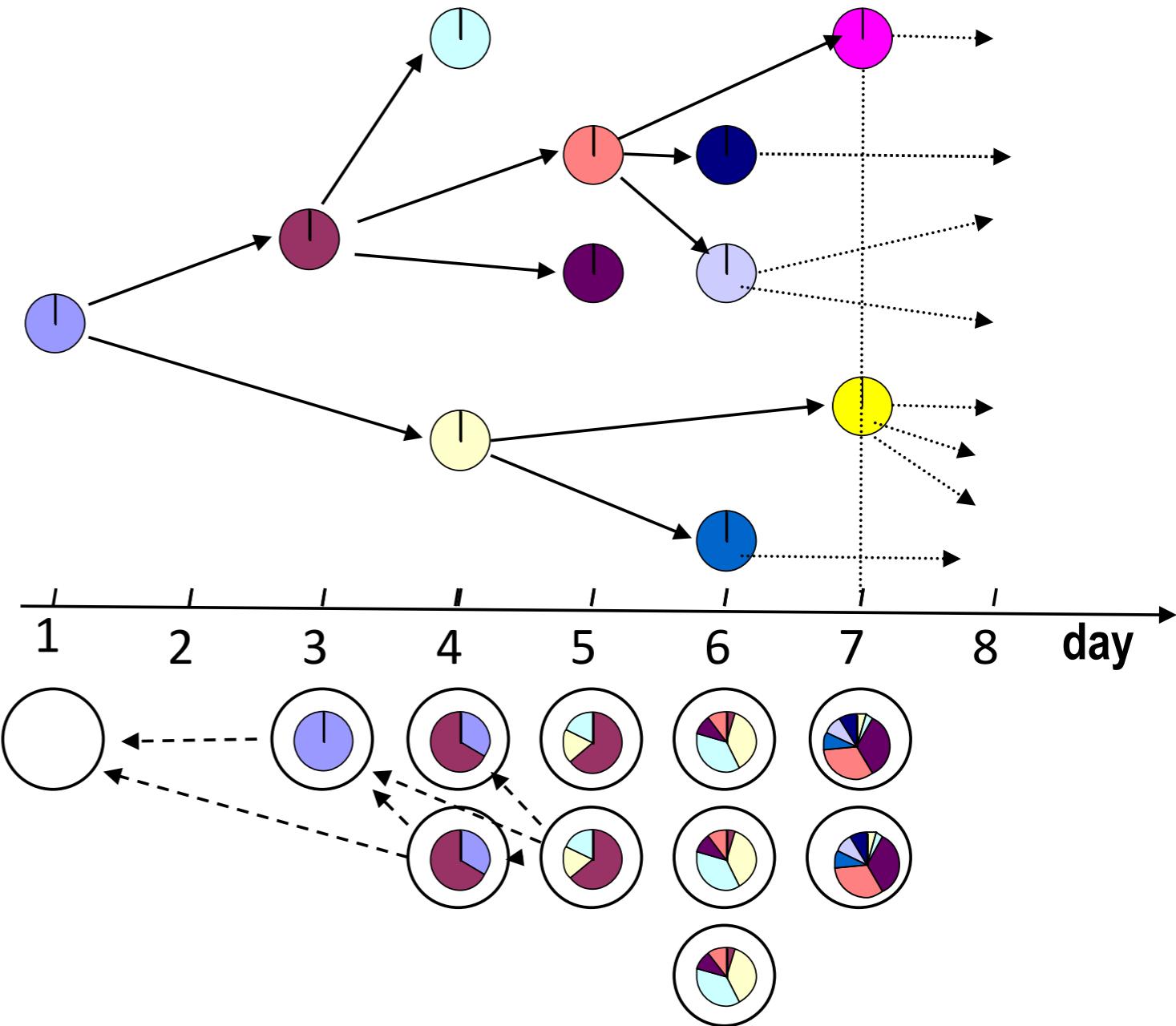
index R_t

generation time



Real
transmission
chain
(unobserved)

Incidence, i.e.
untraced
outbreak
(observed)

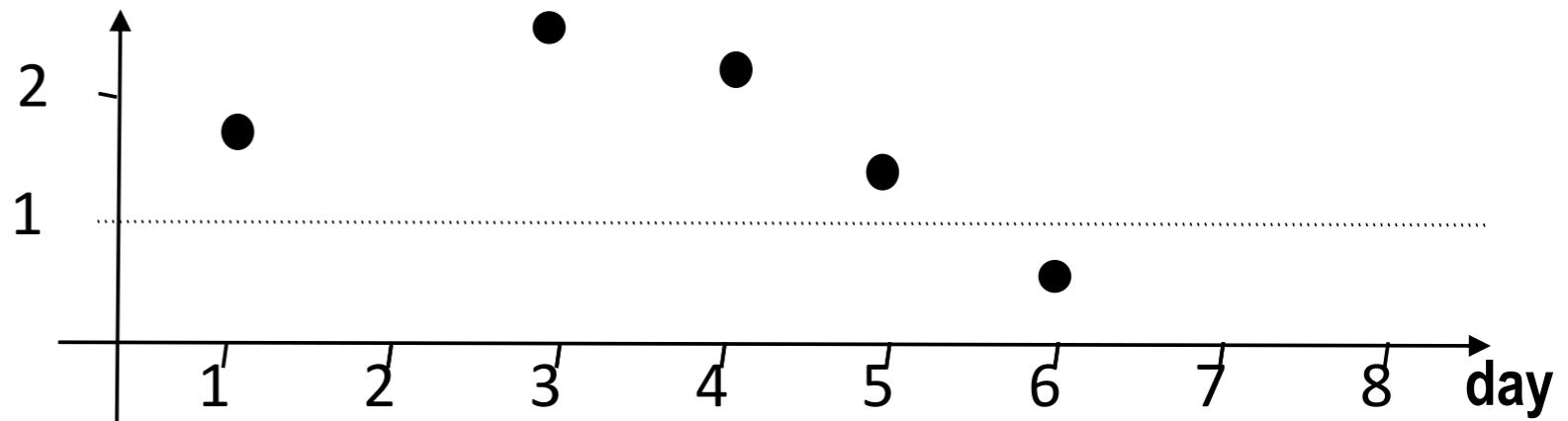


$$p_{ij} = \frac{w(t_i - t_j)}{\sum_{i \neq k} w(t_i - t_k)}$$

$$R_j = \sum_i p_{ij}$$

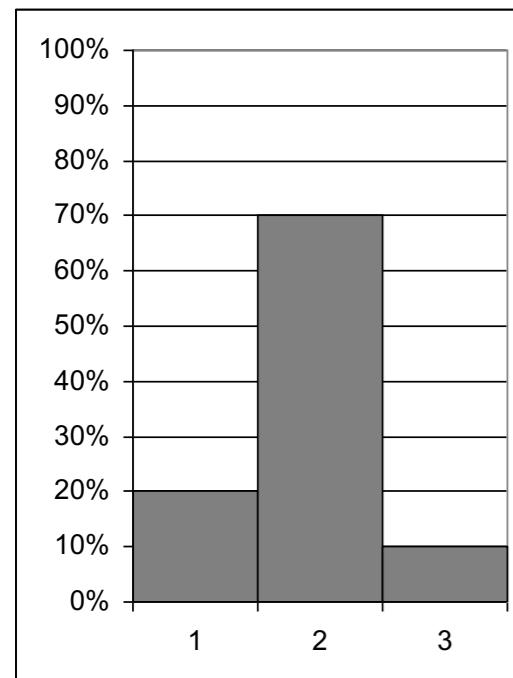
● Wallinga et al.

R



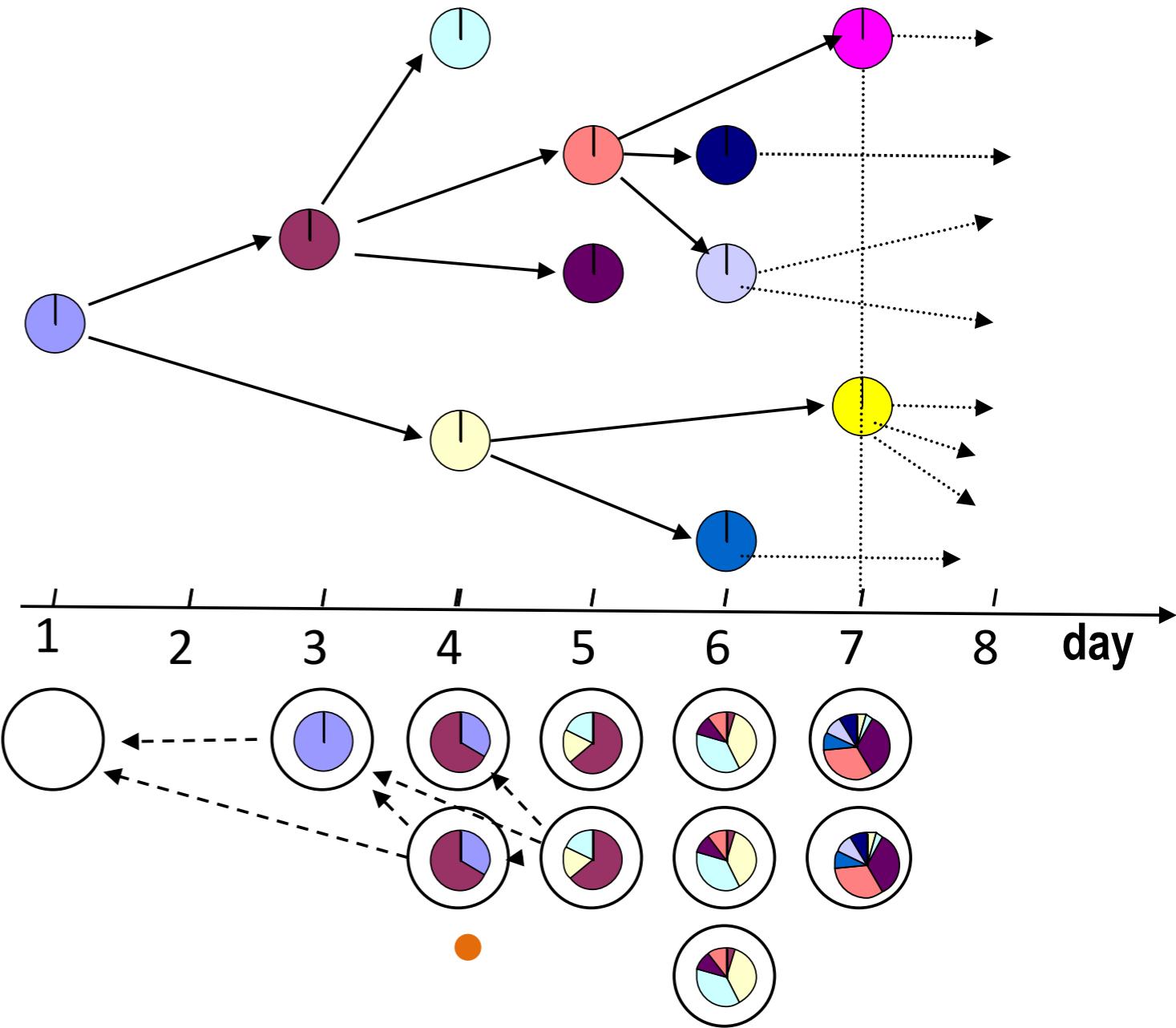
index R_t

generation time

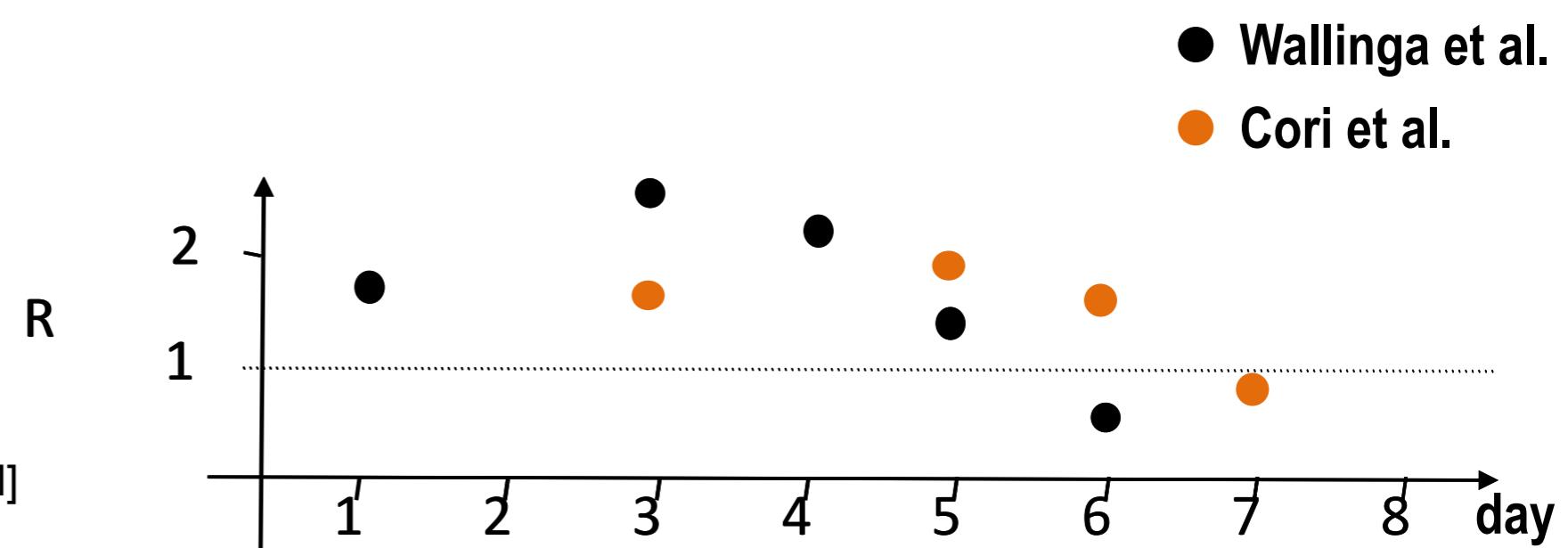


Real
transmission
chain
(unobserved)

Incidence, i.e.
untraced
outbreak
(observed)



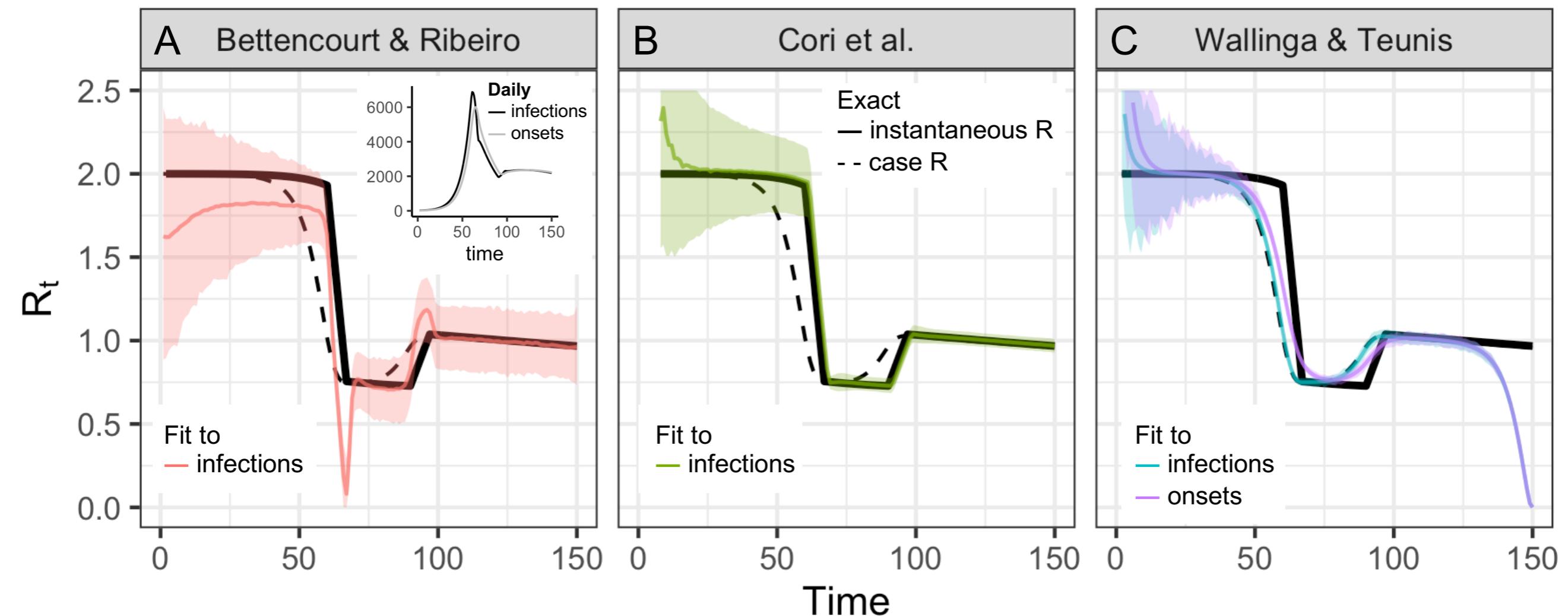
$$R_t = \frac{I_t}{\sum_{s=1}^t I_{t-s} w_s}$$



[Courtesy of Pierre-Yves Boëlle, INSERM]

index R_t

[Gostic KM, McGough L, Baskerville EB, Abbott S, Joshi K, Tedijanto C, et al. (2020) Practical considerations for measuring the effective reproductive number, R_t . PLoS Comput Biol 16(12): e1008409]



index R_t

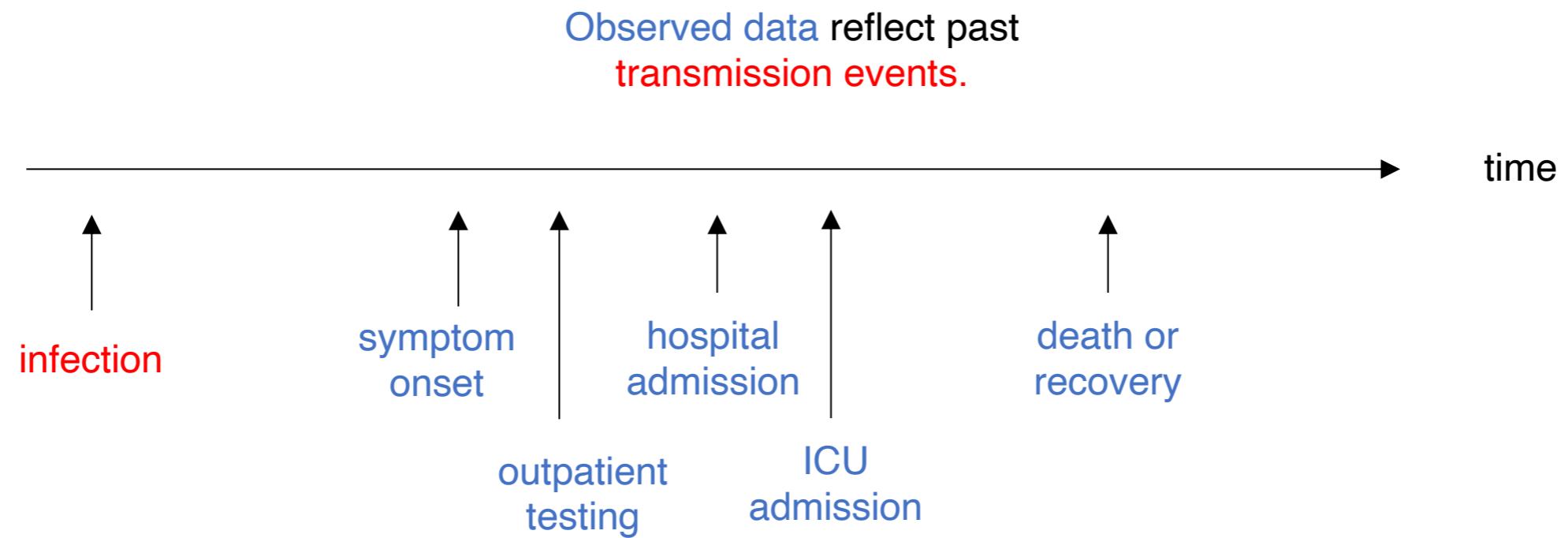
Difference between Wallinga and Cori similar to

actual life span of individuals born in 2013, which we can measure only retrospectively after all individuals have died (Wallinga), and **life expectancy** in 2013, estimated now by assuming that death rates in the future will be similar to those in 2013 (Cori).

index R_t

[Gostic KM, McGough L, Baskerville EB, Abbott S, Joshi K, Tedijanto C, et al. (2020) Practical considerations for measuring the effective reproductive number, R_t . PLoS Comput Biol 16(12): e1008409]

In reality we don't have the time of infection, and in general not even the time of symptoms onset. Solution: deconvolution, inference accounting for latent states



R_0 and R_t are 2 indicators

R_0 = indicator of the strength of an epidemic, i.e. the final outcome in case the epidemic will spread undisturbed, essential for scenario analysis

R_t = indicator of the epidemic trend. It describes the present, essential to provide anticipation

R_0 and R_t : key indicators they provide a lot of information, already

Still, models can go beyond that

Models can simulate the past, present, future of an epidemic, as well as alternative past, present, future

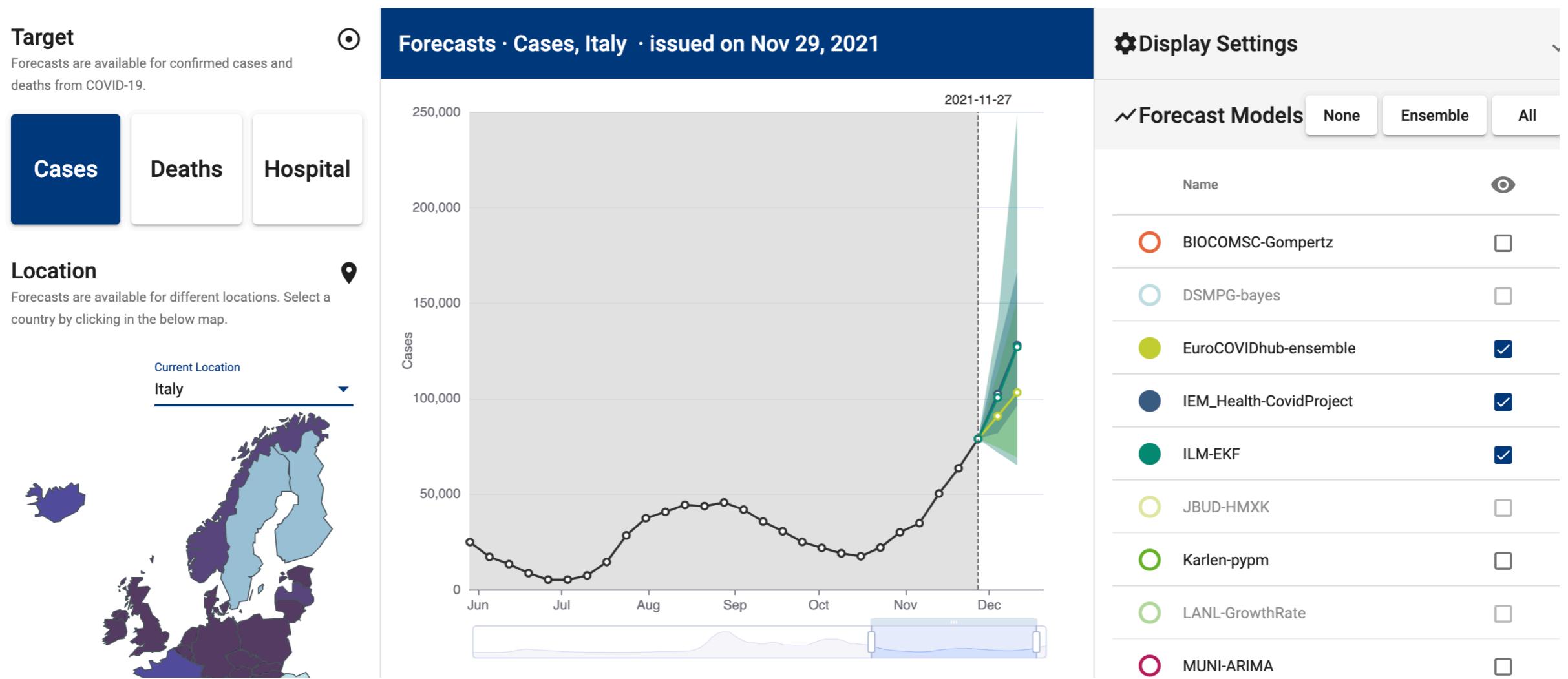
what models can do

forecasting: future projections on the epidemic evolution. Number of cases in the next weeks, months, time of the epidemic peak, height of the epidemic peak (hospital occupancy at the peak) time of the epidemic end, final epidemic size

<https://covid19forecasthub.eu/>

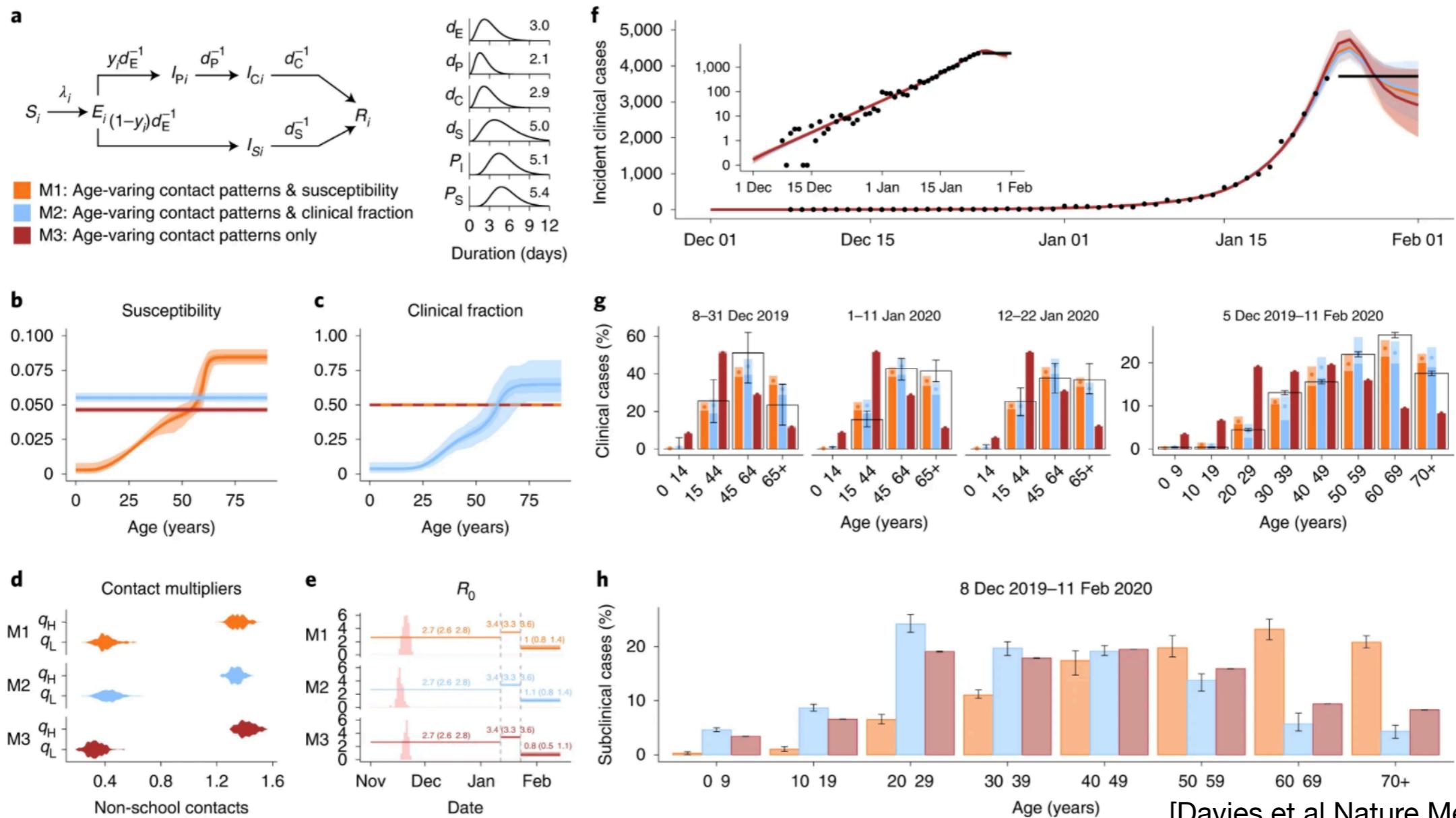
European Covid-19 Forecast Hub Forecasts Reports Community Background Contact

Please be aware of the [limitations](#) when interpreting the forecasts. We recommend to focus on the ensemble which would be expected to give the most reliable forecasts. For horizons beyond one week case forecasts in particular can rapidly become unreliable, especially in times of changing non-pharmaceutical interventions. The individual models that constitute the ensemble have been contributed by independent teams. They are shown for transparency but should not be taken on their own without fully acknowledging their assumptions and limitations and engaging with the listed authors.



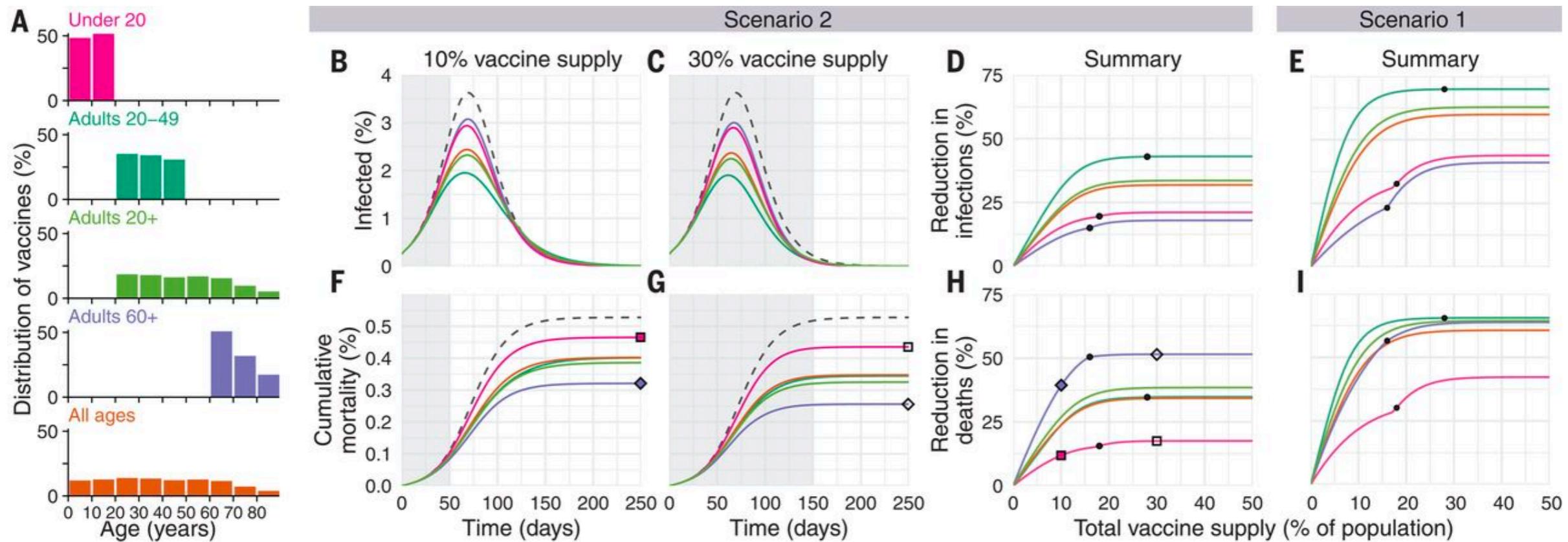
what models can do

medical and biological understanding: e.g. role of different transmission routes (human-to-human, zoonotic, vector-borne, direct transmission, fomite, aerosol, droplets, etc ..), role of asymptomatic/pre-symptomatic in transmission, susceptibility and rate of symptoms by age group

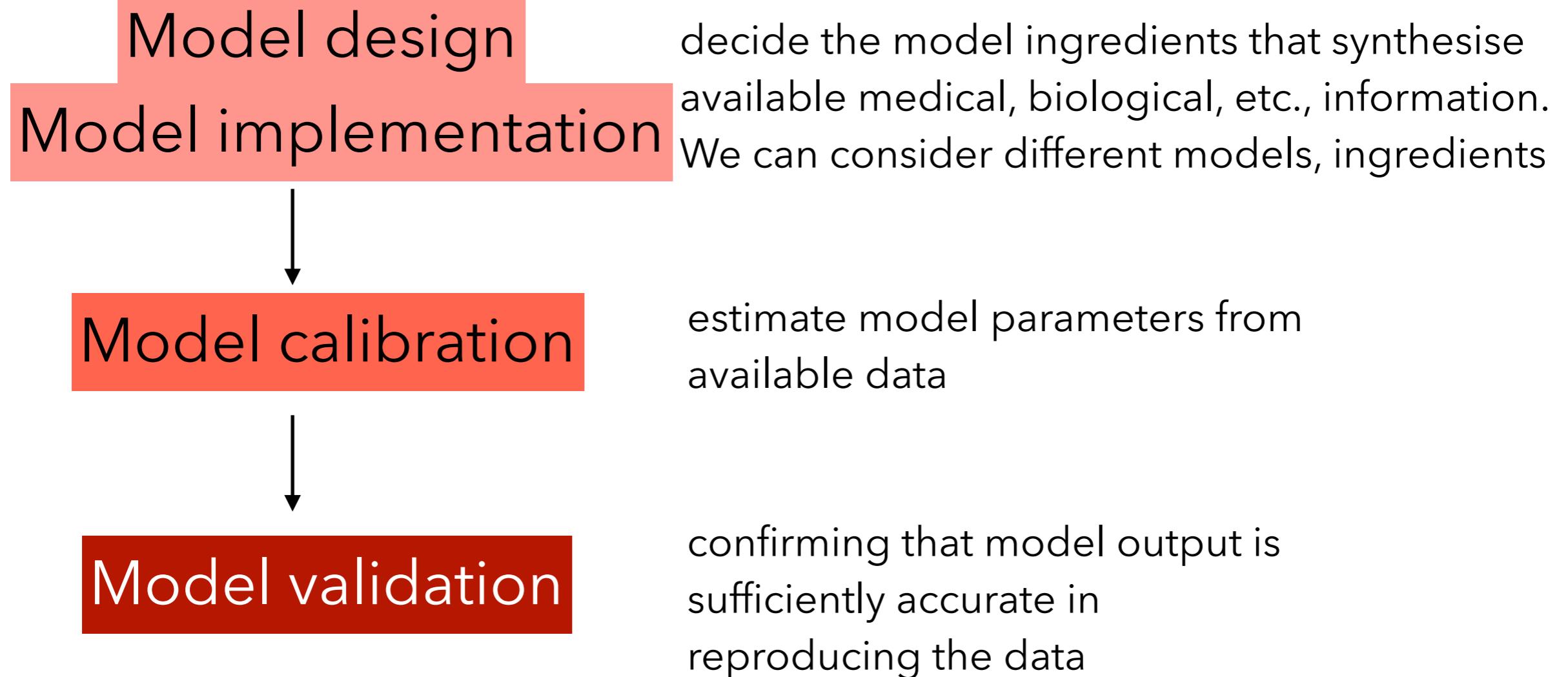


what models can do

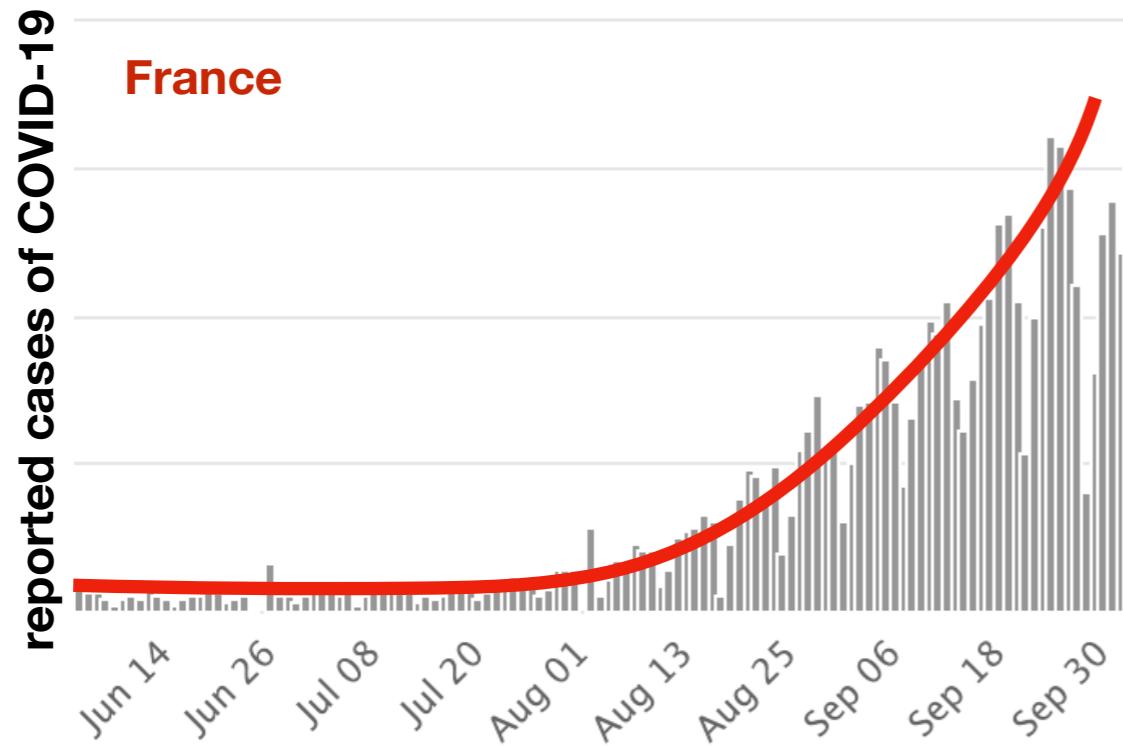
exploration of counterfactuals and hypothetical scenarios: what if scenarios to project the impact of interventions in the future (vaccination, pharmacological interventions, lockdown, travel restrictions). Analysis of counterfactuals scenarios to understand the impact of intervention implemented in the past. Long-term projections.



Epidemic modeling



Model calibration



— modelled trajectory = $F(M, \mathcal{O}, \theta, I_0)$

M = epidemic model

e.g. SIR

\mathcal{O} = observation model

e.g. Binomial process with probability p

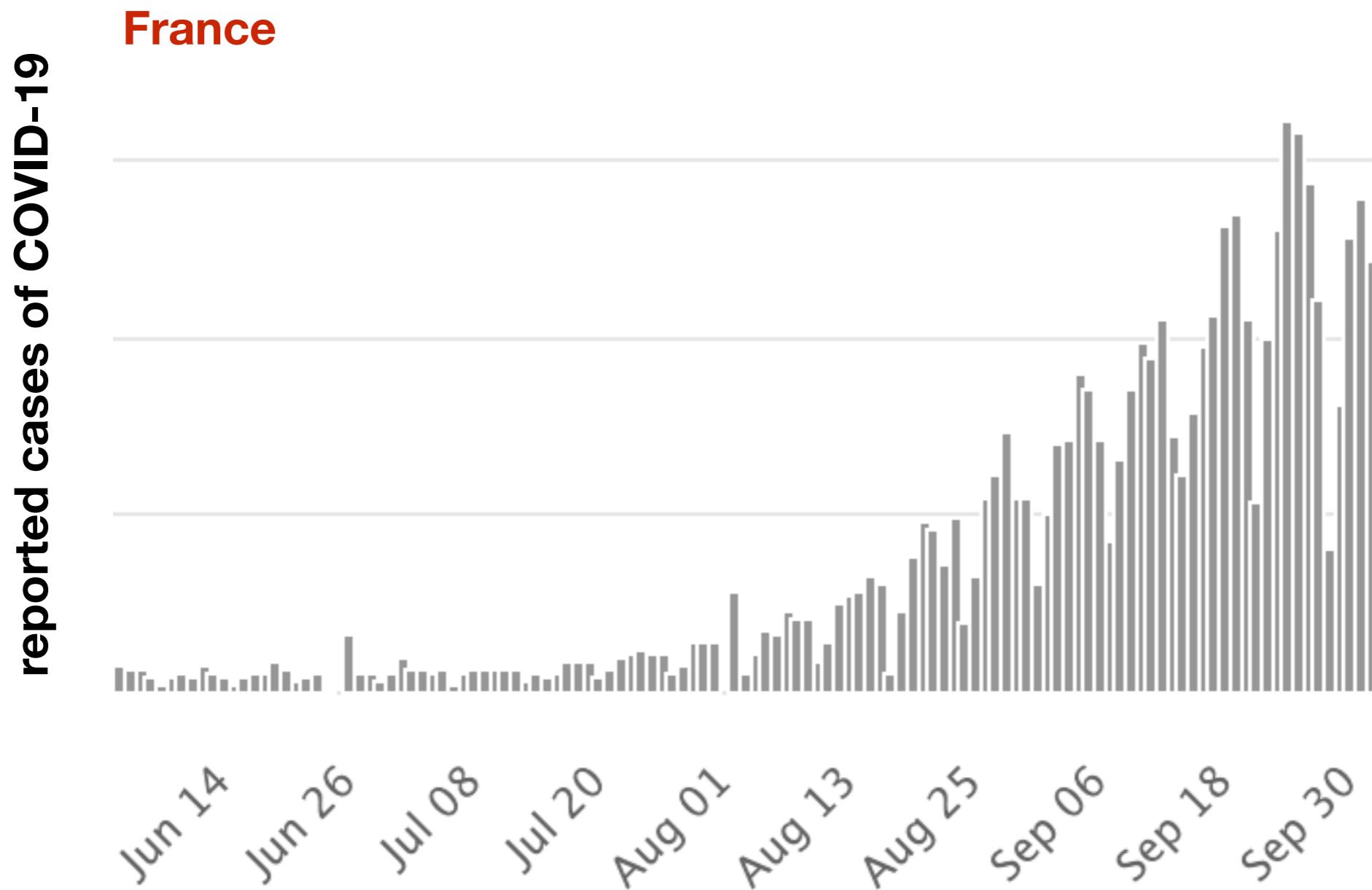
θ = vector of parameters

e.g. (β, μ) = (transmission rate, recovery rate)

I_0 = initial conditions

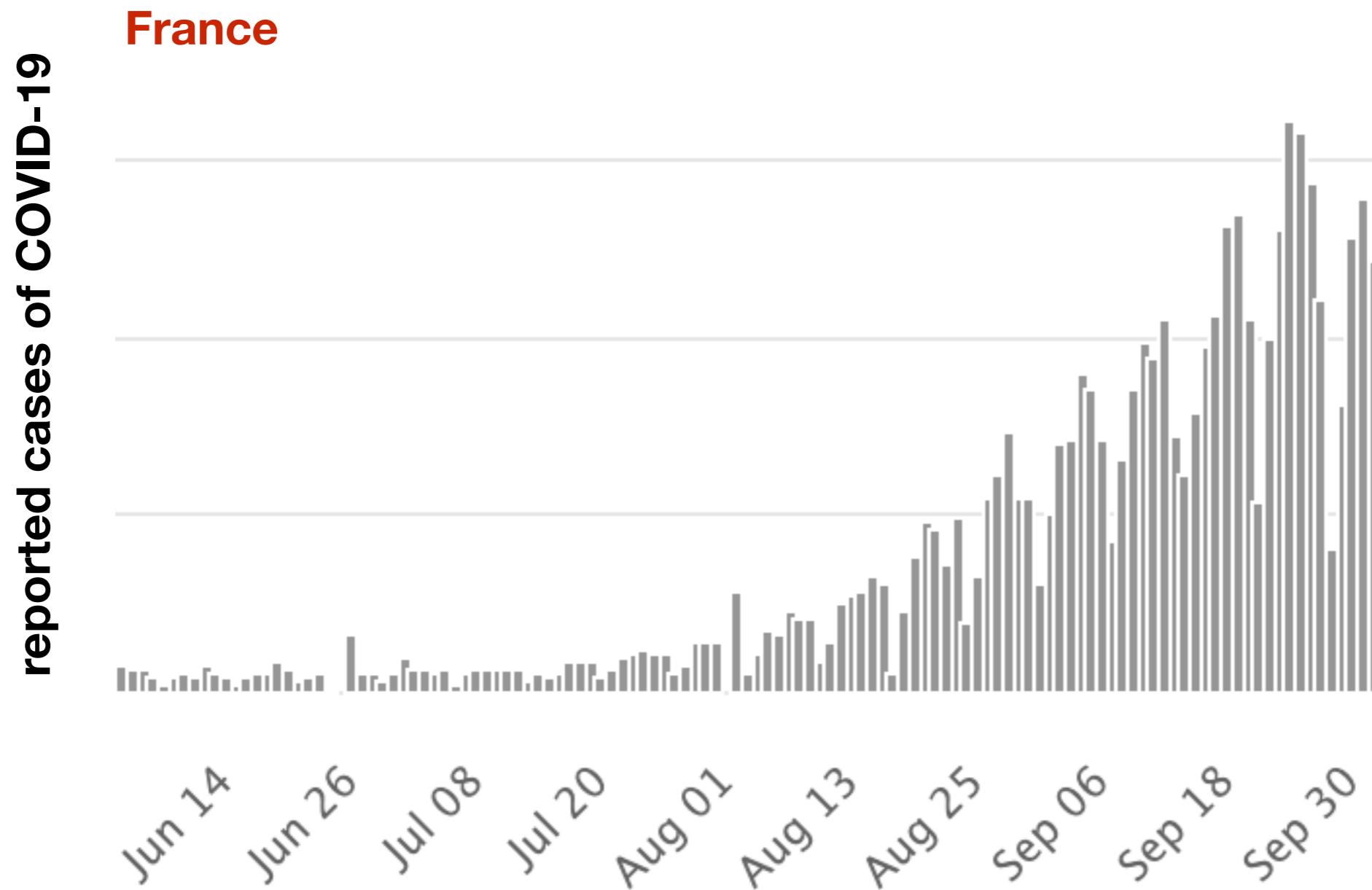
e.g. initial number of infectious

let's talk about the data: what is incidence?



Incidence in a given area a at time t = fraction of the population
catching flu in a at t

let's talk about the data: what is incidence?



Incidence in a given area a = numerator_a/denominator_a

numerator = the number of people hit by flu

denominator = the population at risk

numerator: case definition

case definition = set of criteria used in making a decision as to whether an individual has a disease or health event of interest

possible criteria:

- clinical
- laboratory characteristics
- information regarding the person
- ...

cases can be classified:

- confirmed
- probable
- possible

case definition: sensitivity and specificity

A **sensitive** case definition will detect many cases but may also count as cases individuals who do not have the disease.

A **specific** case definition is more likely to include only persons who truly have the disease under investigation but also more likely to miss some cases

	Disease is truly present	Disease is truly absent	Total
complies to case definition	a	b	all cases
does not comply to case definition	c	d	all non-cases
	all 'diseased'	all 'non-diseased'	all people in the study sample

$$\text{Sensitivity} = [a / (a+c)]$$

$$\text{Specificity} = [d / (b+d)]$$

numerator: example of flu

numerator for a area a= cases seen by General Practitioners defined based on *clinical criteria*. These are in fact possible cases. Laboratory confirmation available only for a small proportion of cases.

symptoms of flu

- no symptoms (~30%)
- upper respiratory symptoms, e.g. nasal stuffiness, runny nose, sore throat, sneezing, hoarseness, ear pressure, or earache (~60%)
- lower respiratory symptoms, e.g. cough, breathing difficulty, and chest discomfort (~2%)
- fever (~35%)

numerator: example of flu

numerator for area a = cases seen by General Practitioners defined based on *clinical criteria*. These are in fact possible cases. Laboratory confirmation available only for a small proportion of cases.

symptoms of flu

- no symptoms (~30%)
- upper respiratory symptoms, e.g. nasal stuffiness, runny nose, sore throat, sneezing, ear pressure, or earache (~60%)
- lower respiratory symptoms, e.g. cough, breathing difficulty (~2%)
- fever (~35%)
- from early to peak symptoms ~1 day



clinical case definition (Influenza-like-illness, ILI)

- fever $> 39^{\circ}\text{C}$ AND myalgia
- sudden onset
- respiratory symptoms

Sentinelles
Réseau Sentinelles

numerator: example of flu

numerator for area a = cases seen by General Practitioners defined based on *clinical criteria*. These are in fact possible cases. Laboratory confirmation available only for a small proportion of cases.

symptoms of flu

- no symptoms (~30%)
- upper respiratory symptoms, e.g. nasal stuffiness, runny nose, sore throat, sneezing, hoarseness, ear pressure, or earache (~60%)
- lower respiratory symptoms, e.g. cough, breathing difficulty, and chest discomfort (~2%)
- fever (~35%)

(Carrat et al. Am J Epidemiol 2008)



clinical case definition (Acute-respiratory-infection, ARI)

- fever OR malaise OR headache OR myalgia
- sudden onset
- cough OR sore throat OR shortness of breath

numerator: example of flu

numerator for area a = cases seen by General Practitioners defined based on *clinical criteria*. These are in fact possible cases. Laboratory confirmation available only for a small proportion of cases.

clinical case definition



- fever > 39 °C AND myalgia
- sudden onset
- respiratory symptoms

higher specificity

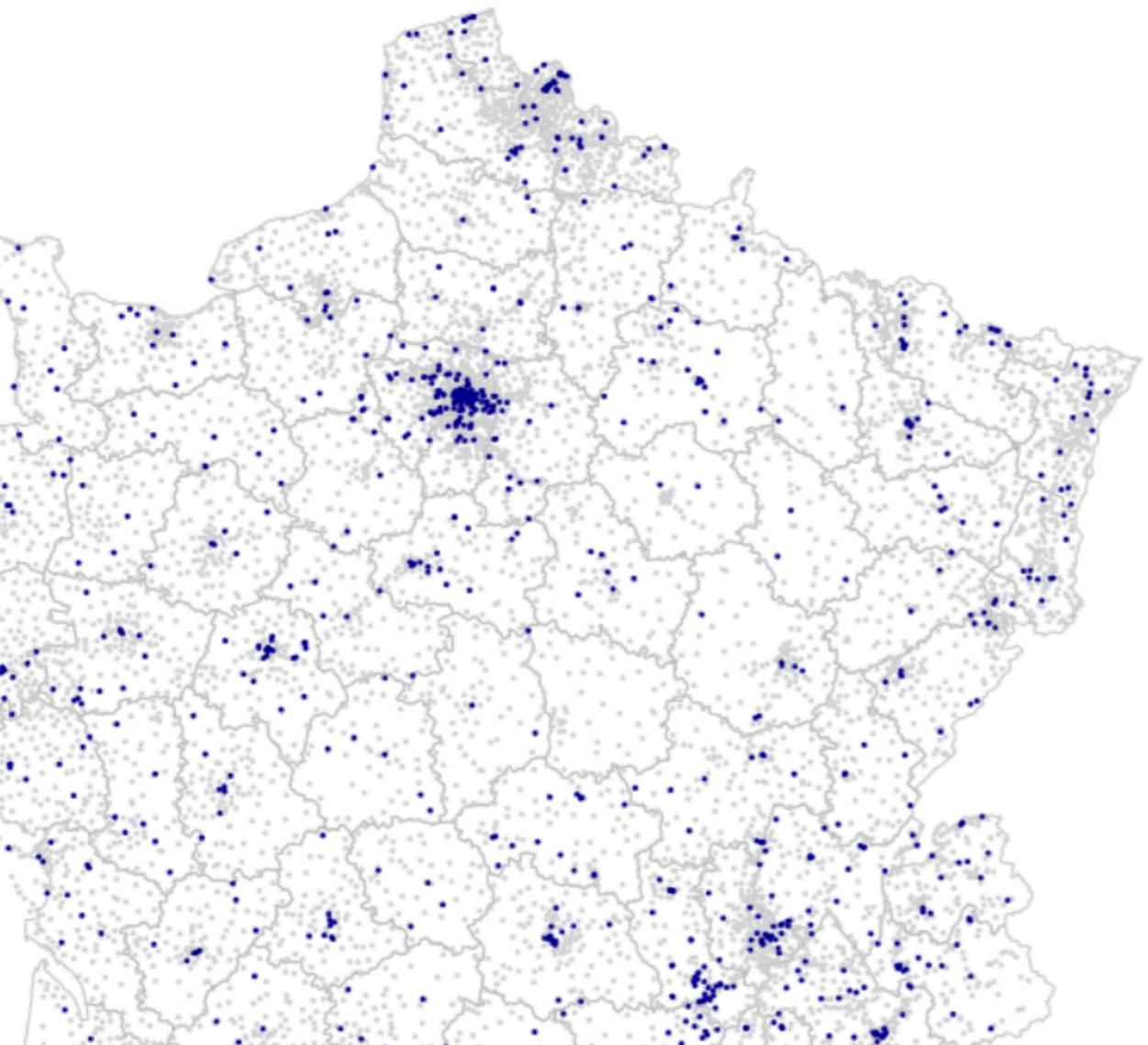


- fever OR malaise OR headache OR myalgia
- sudden onset
- cough OR sore throat OR shortness of breath

higher sensitivity

denominator: example of flu

denominator for the area a = catchment population, i.e. all the people living in the catchment area of the General Partitioner reporting the cases, who would usually seek healthcare at the site when they get sick



The Surveillance Networks (SN) is based on a *fraction of General Partitioner* (~1%), who are volunteers

denominator $a = \text{Population}_a \text{ GP}_{\text{SN},a} / \text{GP}_a$

(Horvitz DG, Thompson DJ. A JASA. 1952;47:663–85)

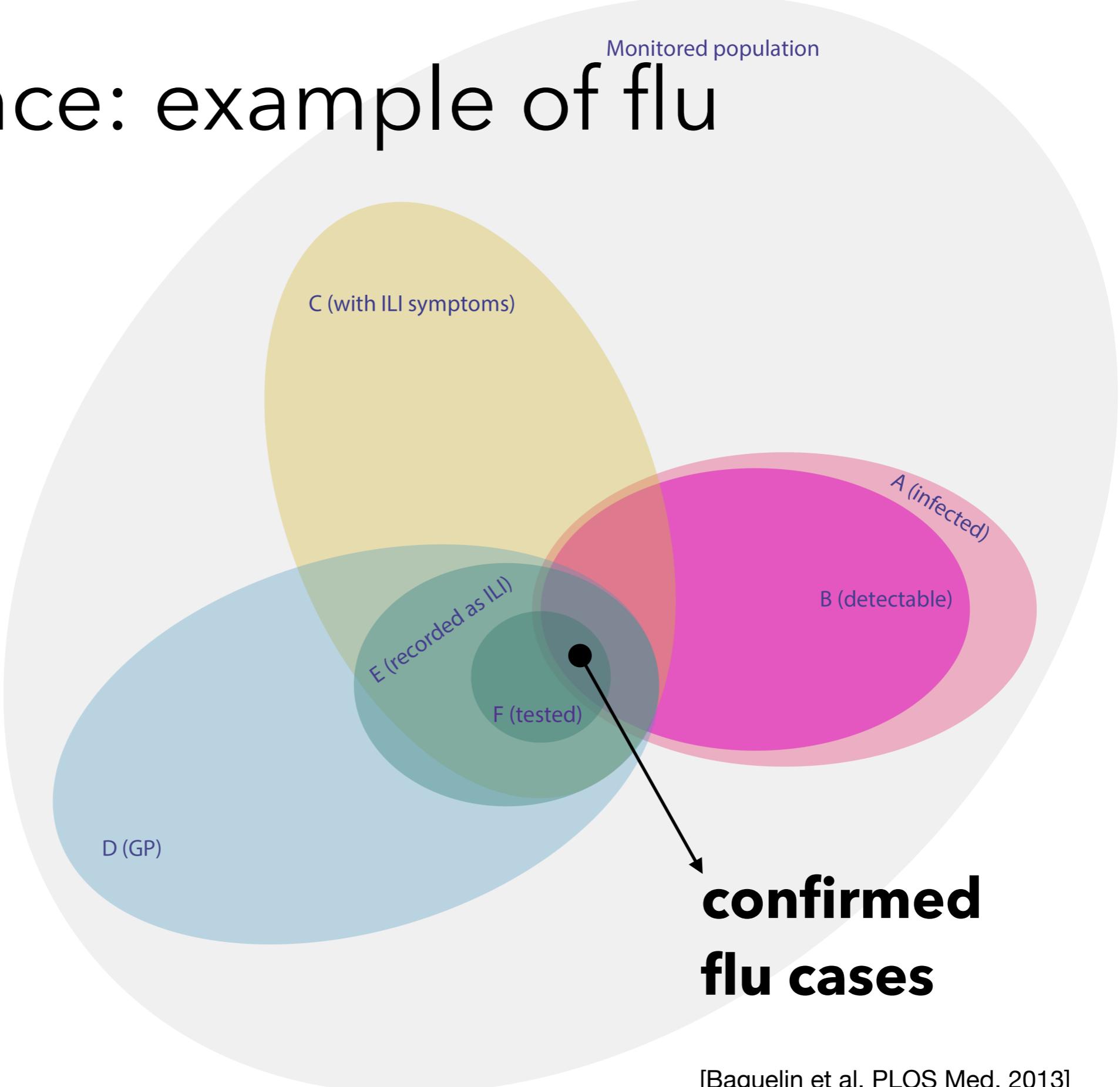
incidence: example of flu

Incidence in a given area a= numerator_a/denominator_a

but ... consultancy rate

- many people are asymptomatic or paucisymptomatic
- highly variable by age
- variable geographically (e.g. dependent on the GP density)
- variable according to the period of the year
- highly dependent on the health-care system (how expensive is going to the GP? Do you need a permit for staying at home from work?)

incidence: example of flu



incidence: example of COVID-19 (early stage)

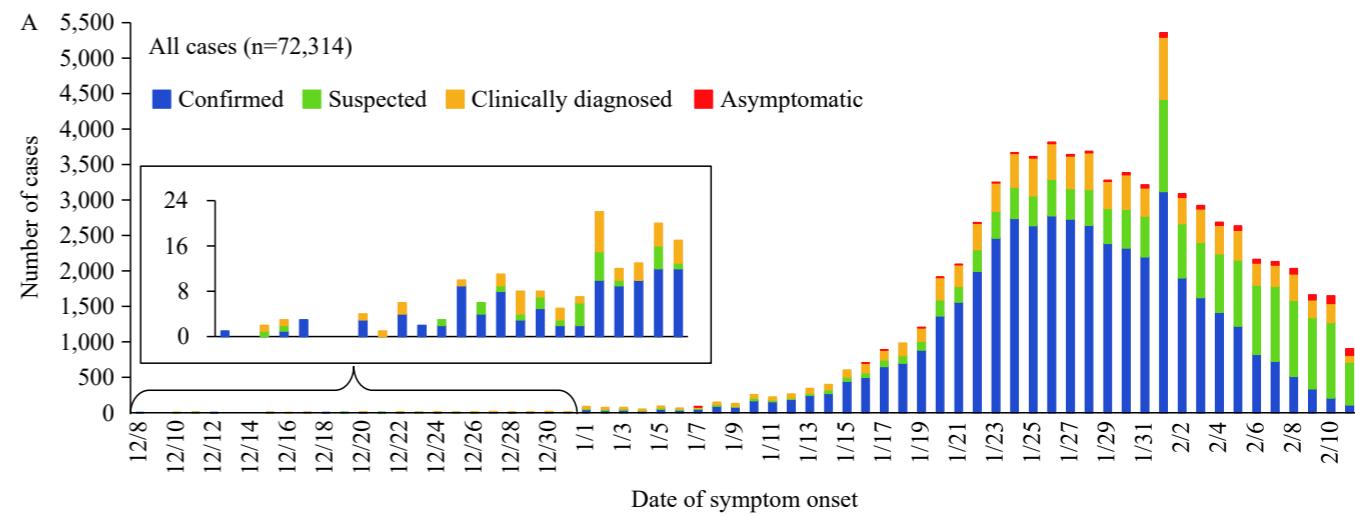
- pathogen and disease unknown (range of symptoms unknown/ lack of diagnostic testing): case definition varying in time

incidence: example of COVID-19 (early stage)

- pathogen and disease unknown (range of symptoms unknown/ lack of diagnostic testing): case definition varying in time
- not wide spread: unclear the region where it is spreading, thus denominator hard to estimate
- reporting rate highly variable in time: from passive to active surveillance, to saturation of the surveillance system

incidence: example of COVID-19 (early stage)

- pathogen and disease unknown (range of symptoms unknown/ lack of diagnostic testing): case definition varying in time
- not wide spread: unclear the region where it is spreading, thus denominator hard to estimate
- reporting rate highly variable in time: from passive to active surveillance, to saturation of the surveillance system
- there can be a retrospective correction of the number of cases: further source of biases in case of a real time analysis



[China CDC, weekly vol 2 number 8]