Emergence and evolution of the SARS-CoV-2 pandemic

Samuel Alizon

ETE team, MIVEGEC, CNRS, IRD, Université de Montpellier

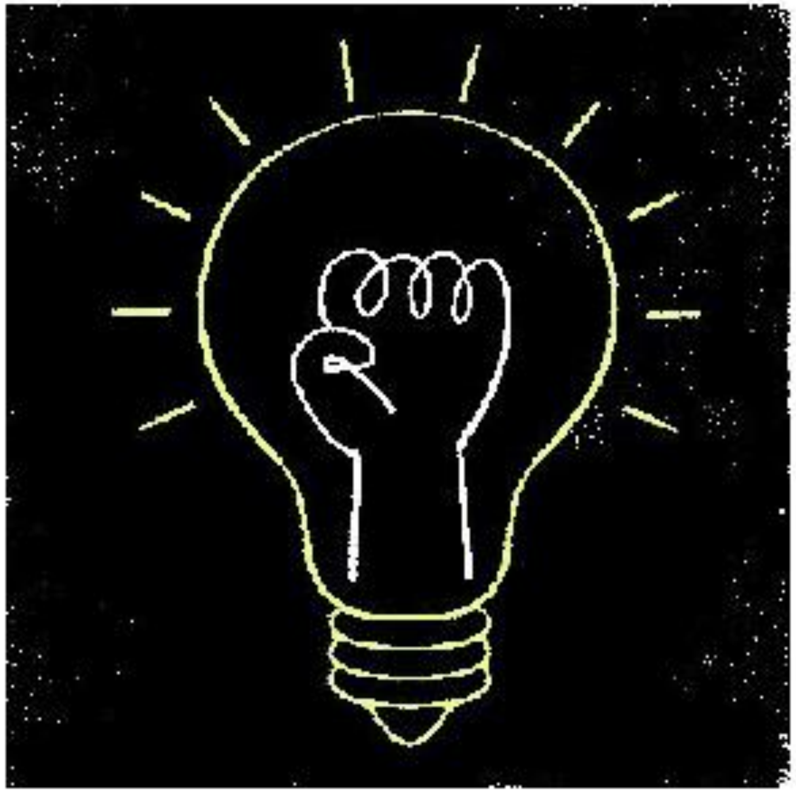
covid-ete.ouvaton.org

XXXIe Congrès National de la Société Française d'Hygiène Hospitalière, 10 Juin 2021

Disclaimer & acknowledgments

- I am a research director at the CNRS in epidemiology & evolution.
- I do not own shares in any company and did not receive any personal money for this work.
- The ETE team is composed of Thomas Bénéteau, Corentin Boennec, Marc Choisy, Gonché Danesh, Ramsès Djidjou-Demasse, Baptiste Elie, Yannis Michalakis, Bastien Reyné, Quentin Richard, Christian Selinger, Mircea T. Sofonea
- CHU Montpellier, CERBA laboratory and the SFM shared data and expertise on RT-PCRs.
- Santé Publique France and the Agences Régionales de Santé compiled and shared data.
- GISAID participants shared sequence data.
- The iTrop cluster from IRD Montpellier (`http://bioinfo.ird.fr`) and France Bioinformatique host our applicationss.
- We received 120k€ from the Région Occitanie (ANR PhyEpi).

French research is drowning...



for detailed analyses, see www.lecese.fr

Emergence of infectious diseases

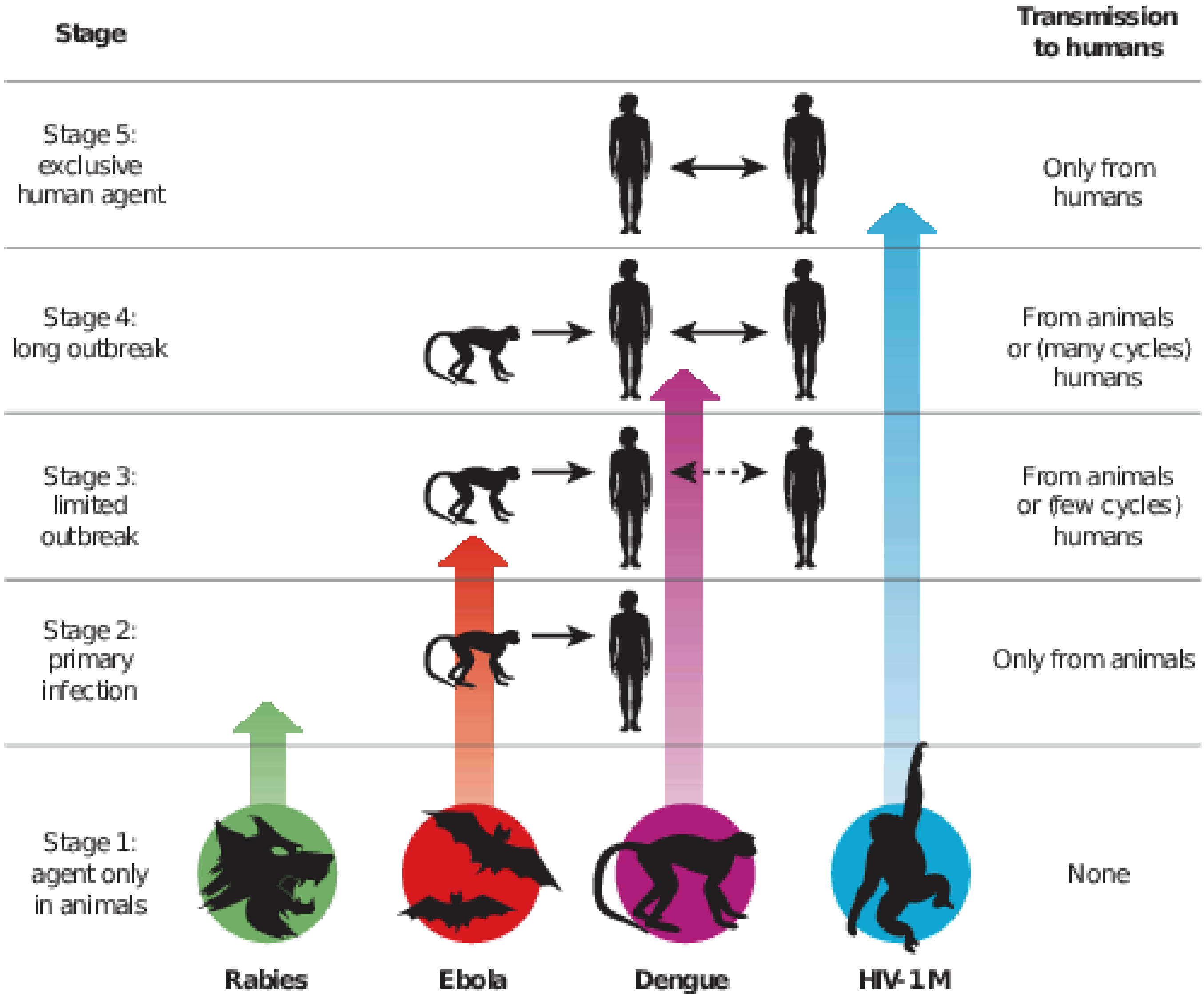
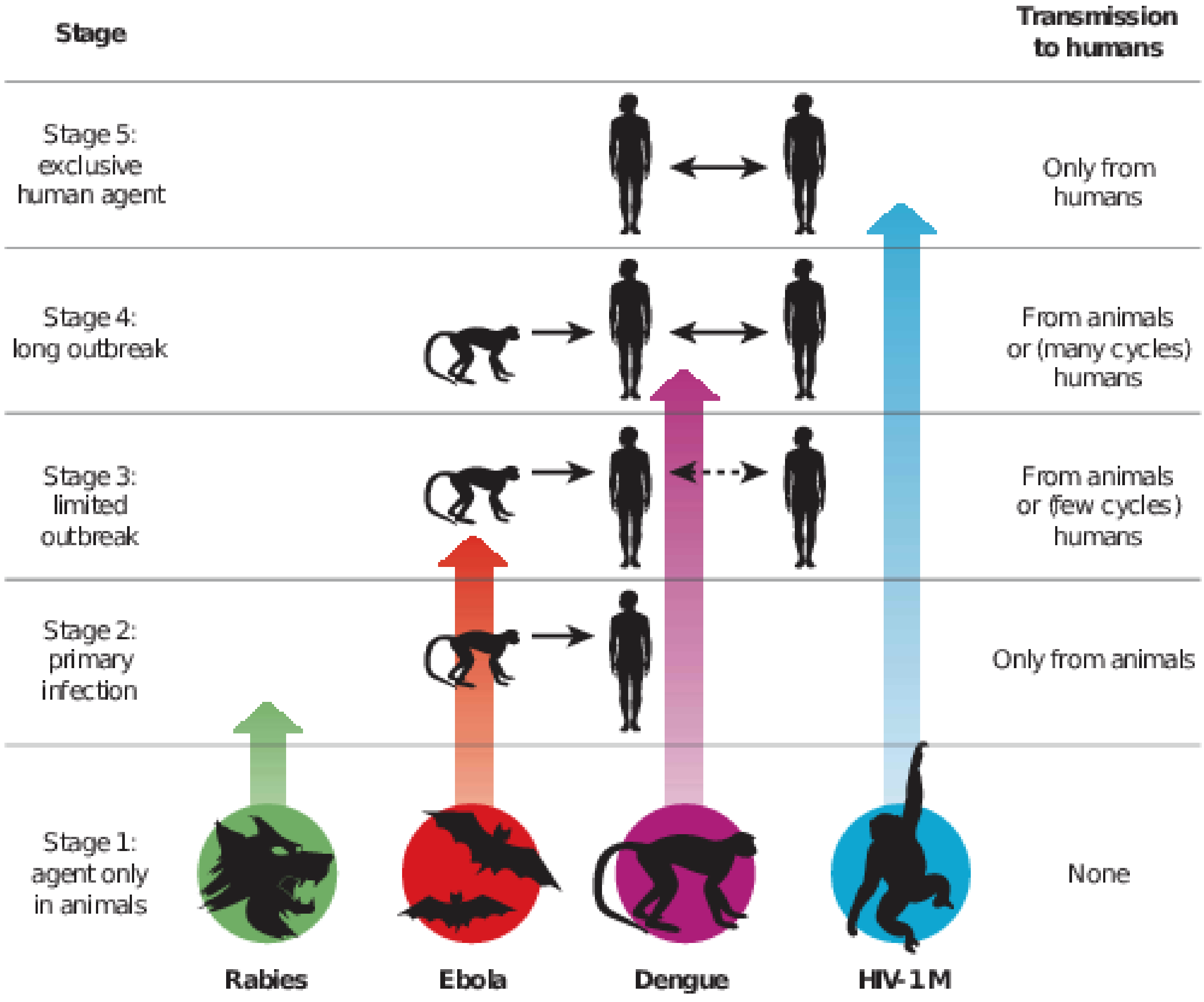


figure from Wolfe *et alii* (2007, *Nature*)

Emergence of infectious diseases



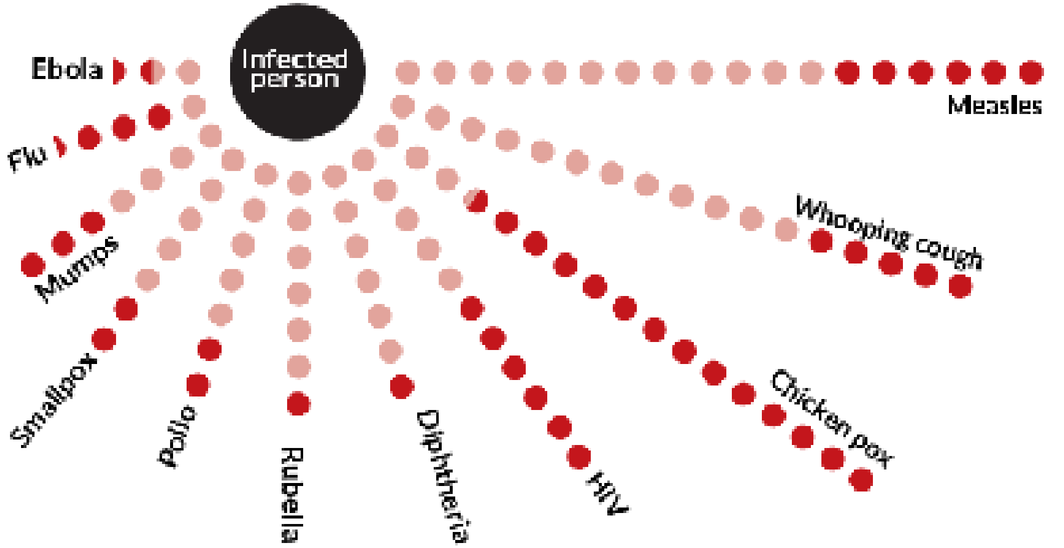
Risk factors such as:

- contacts between humans and wild animals
- contacts between cattle and wild animals
- wild landscape fragmentation
- individual and goods movements
- poor public health systems

figure from Wolfe *et alii* (2007, *Nature*)

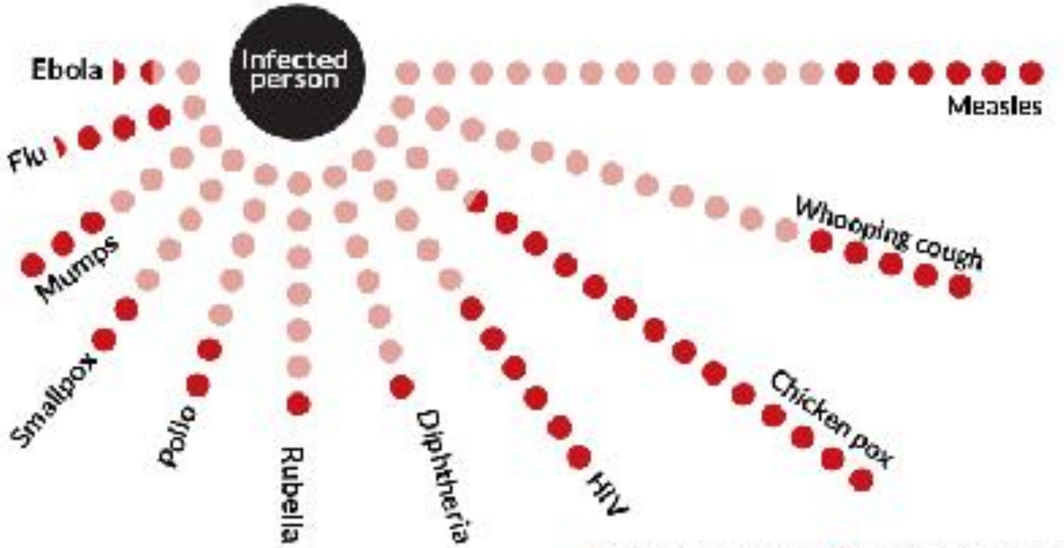
R_0

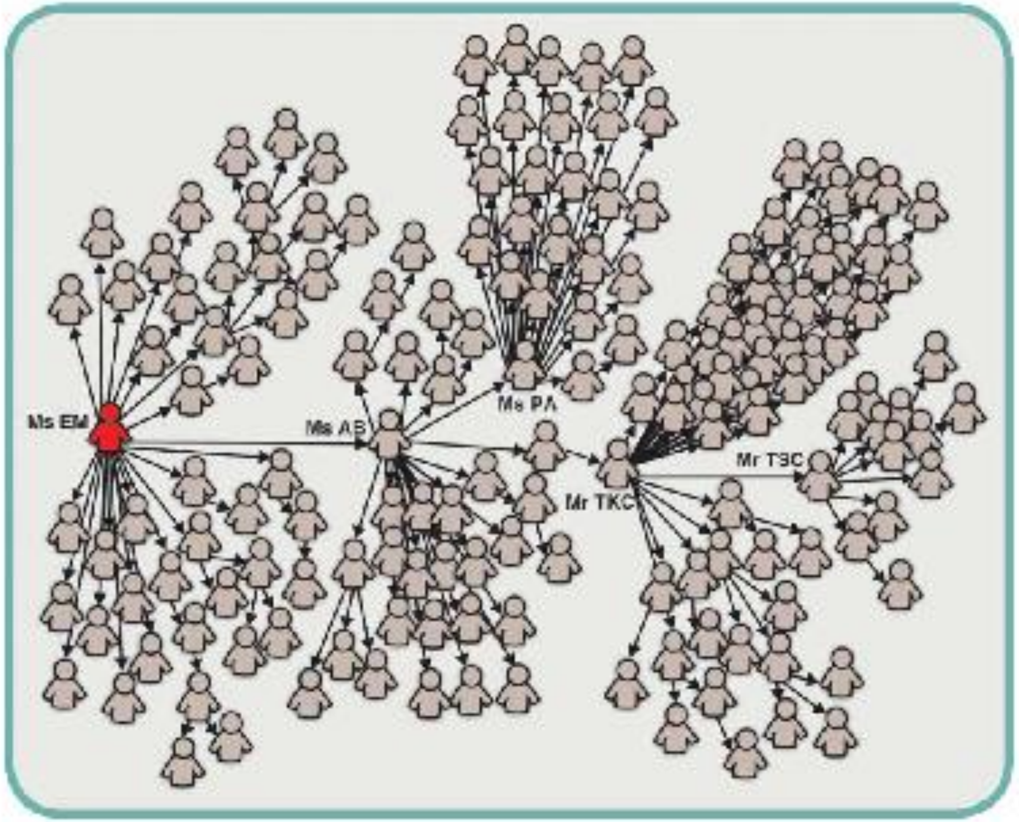
Average number of secondary infections:



R_0

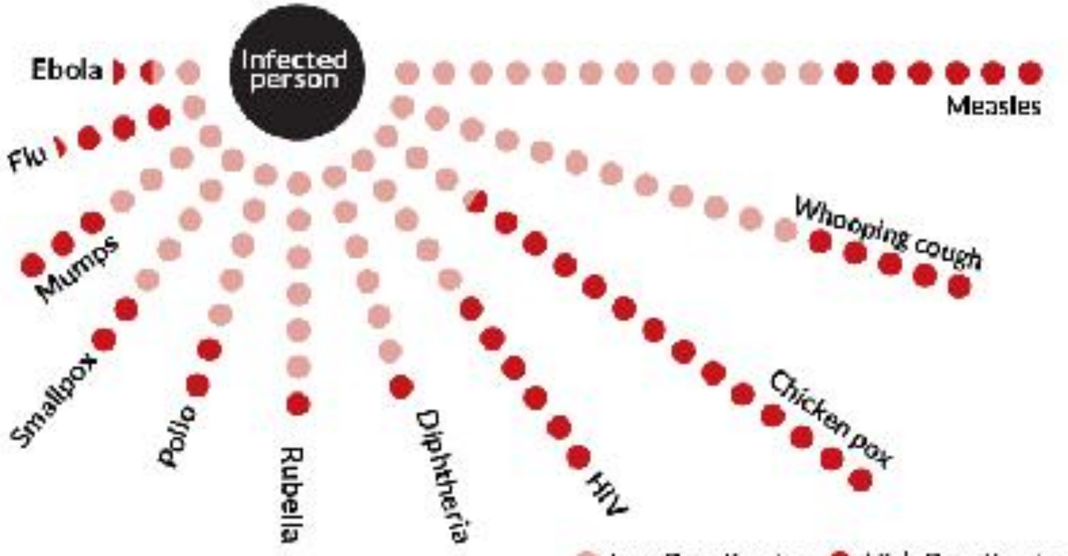
Average number of secondary infections:

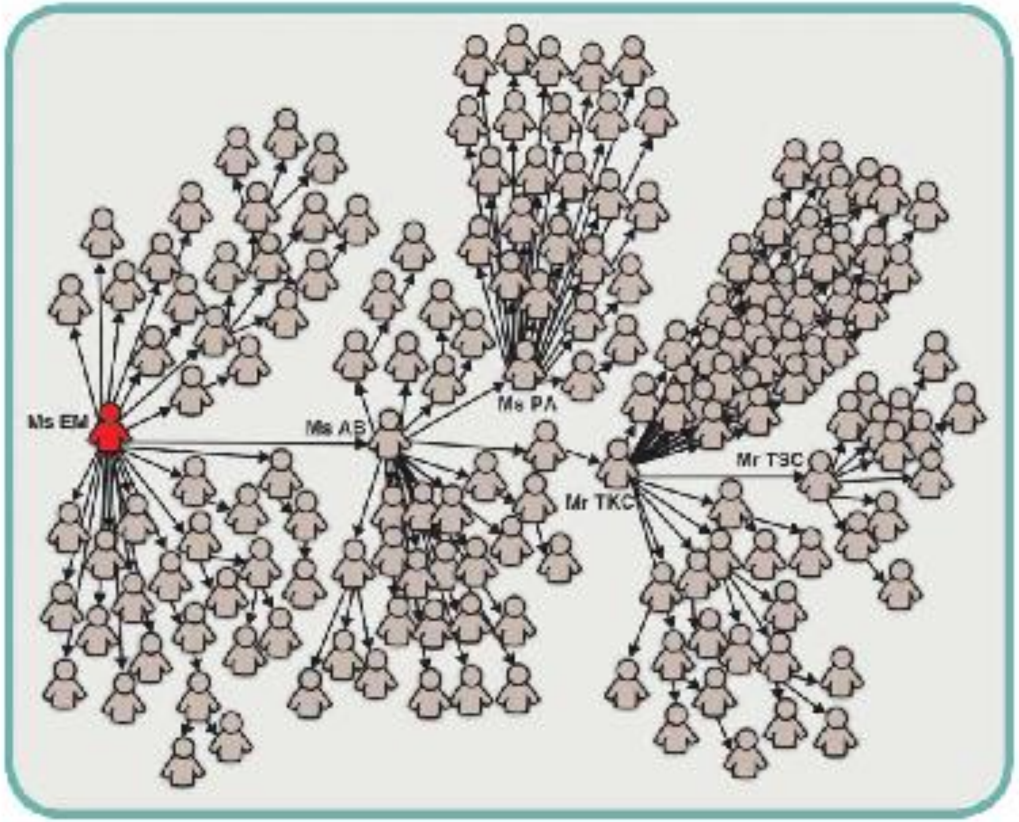




R_0

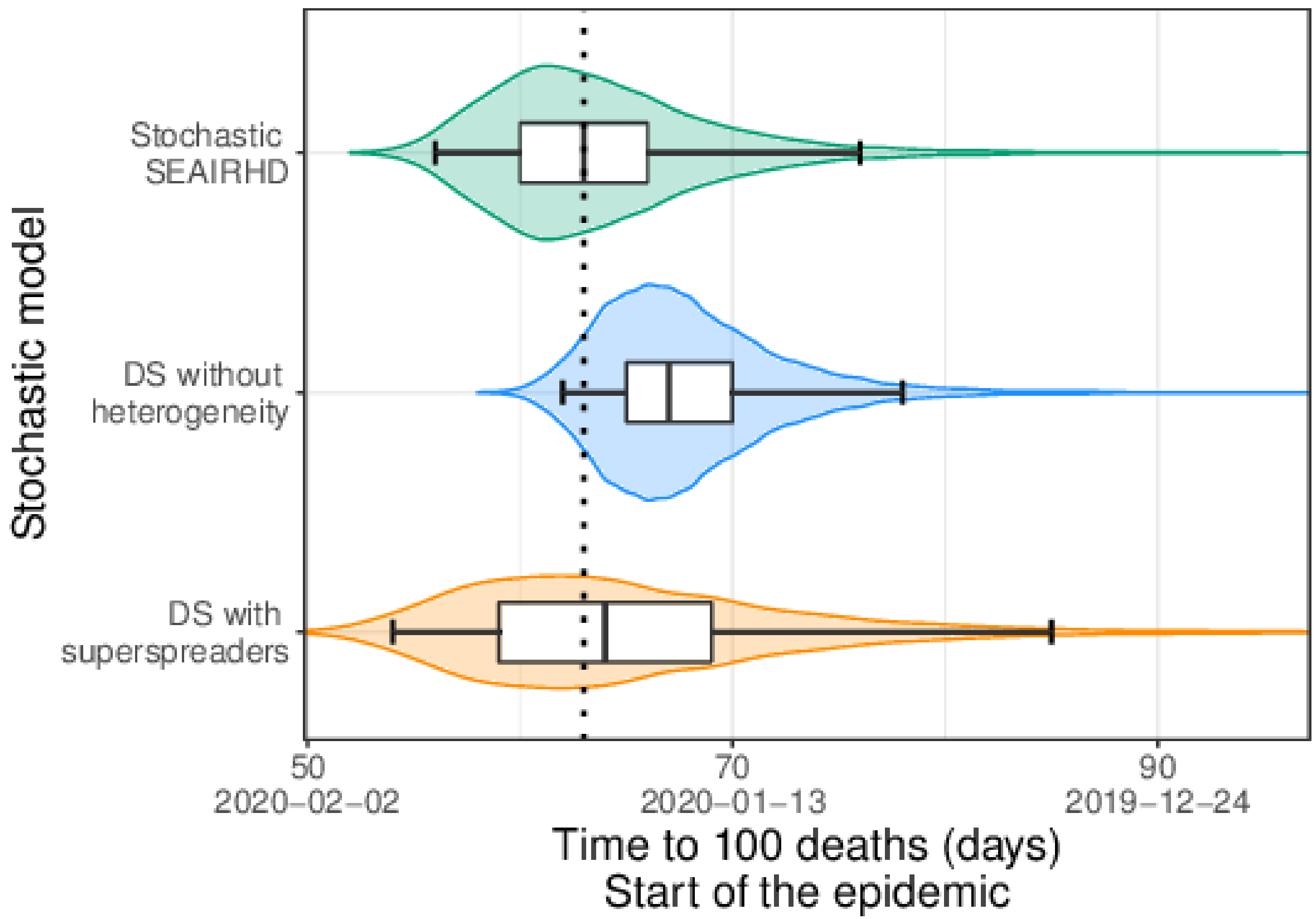
Average number of secondary infections:





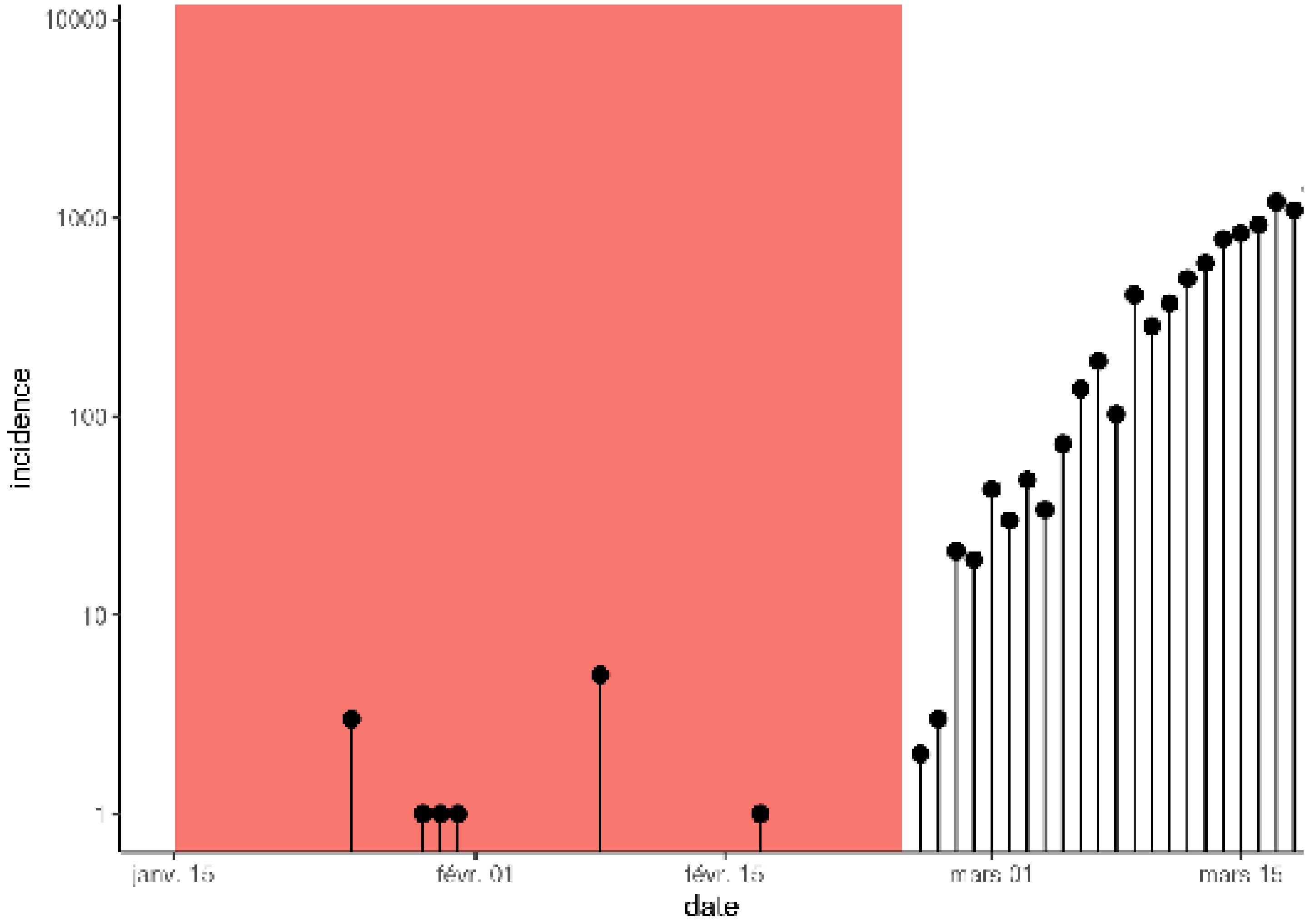
In simple models, the probability of emergence is proportional to $1 - \frac{1}{R_0}$.

Dating an outbreak origin (and end)

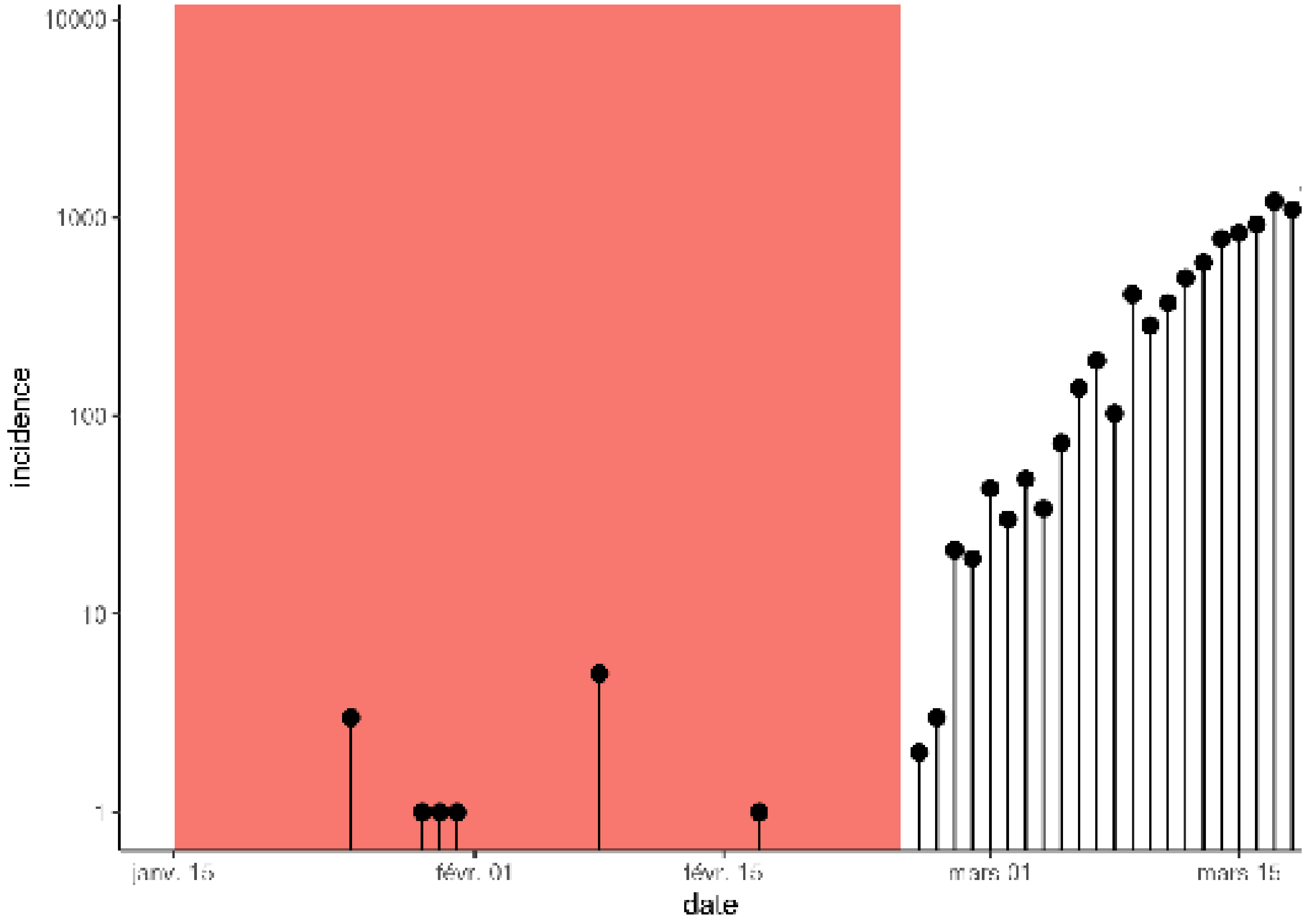


Bénéteau *et alii*
(*Peer Community In Journal, in press*)

Stochastic vs. deterministic



Stochastic vs. deterministic

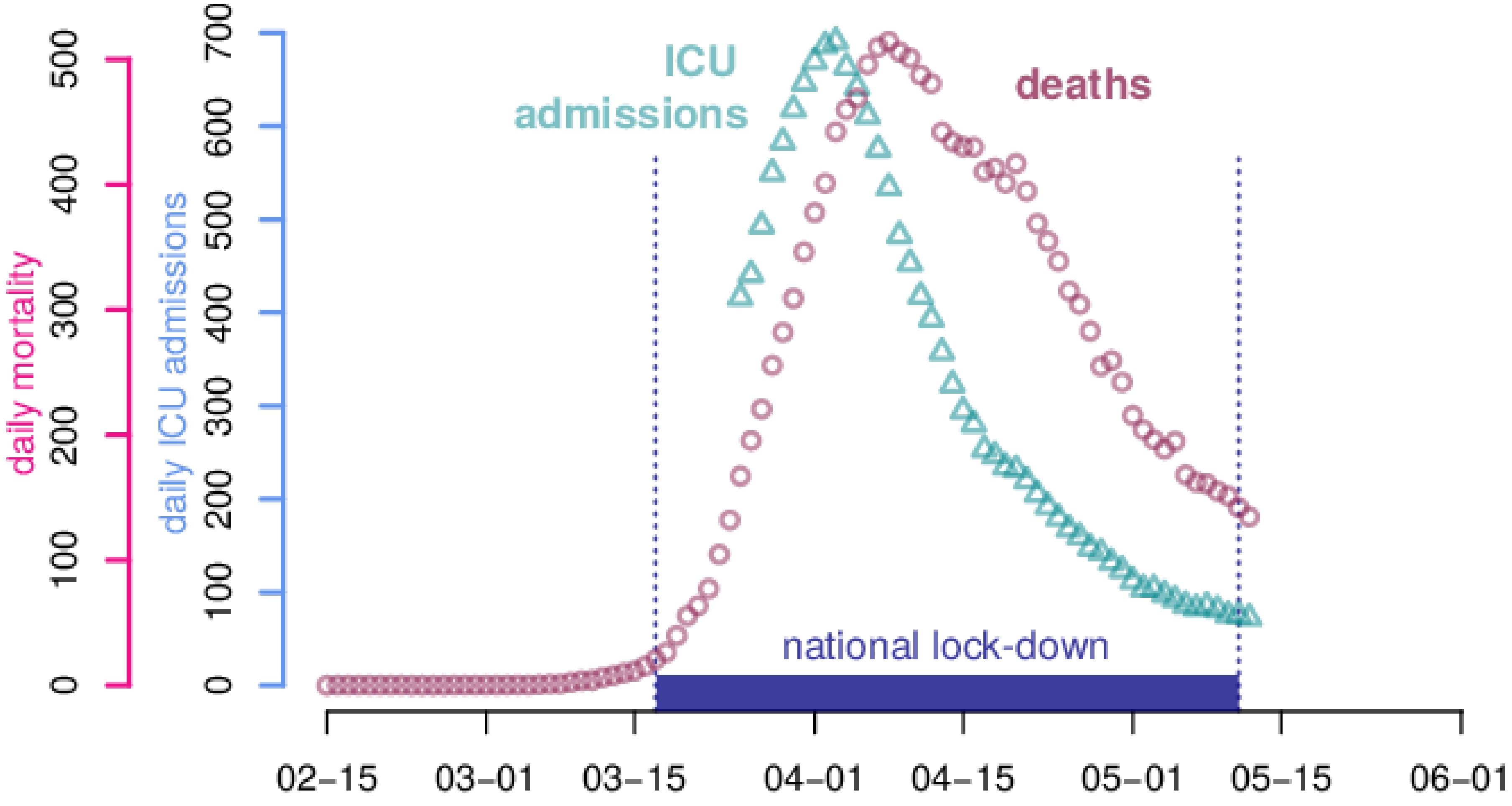


- doubling time:
  $t_D = \frac{r}{\ln(2)} \approx 3$ days

- $\mathcal{R}_0 = 2.49$ [2.39 ; 2.58]

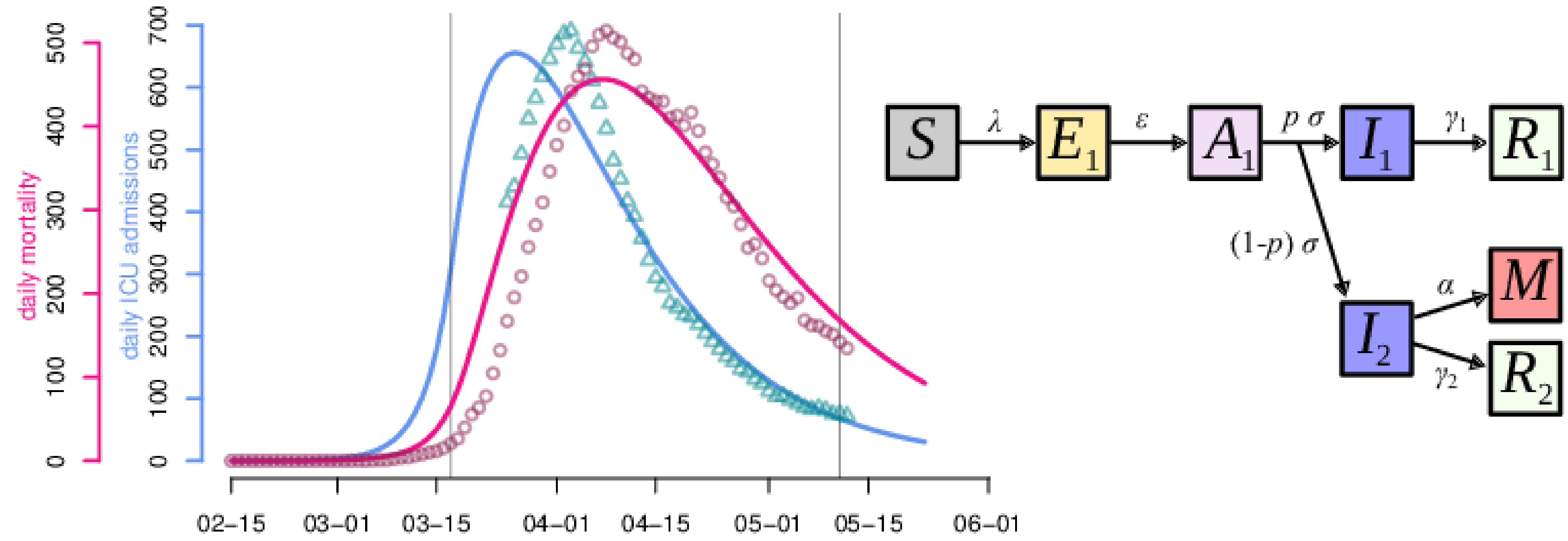
Report 1 (Mar 14, 2020)

Goal: predict ICU admissions



Sofonea *et alii* (2021, *Epidemics*)

Classical 'memoryless' models fail

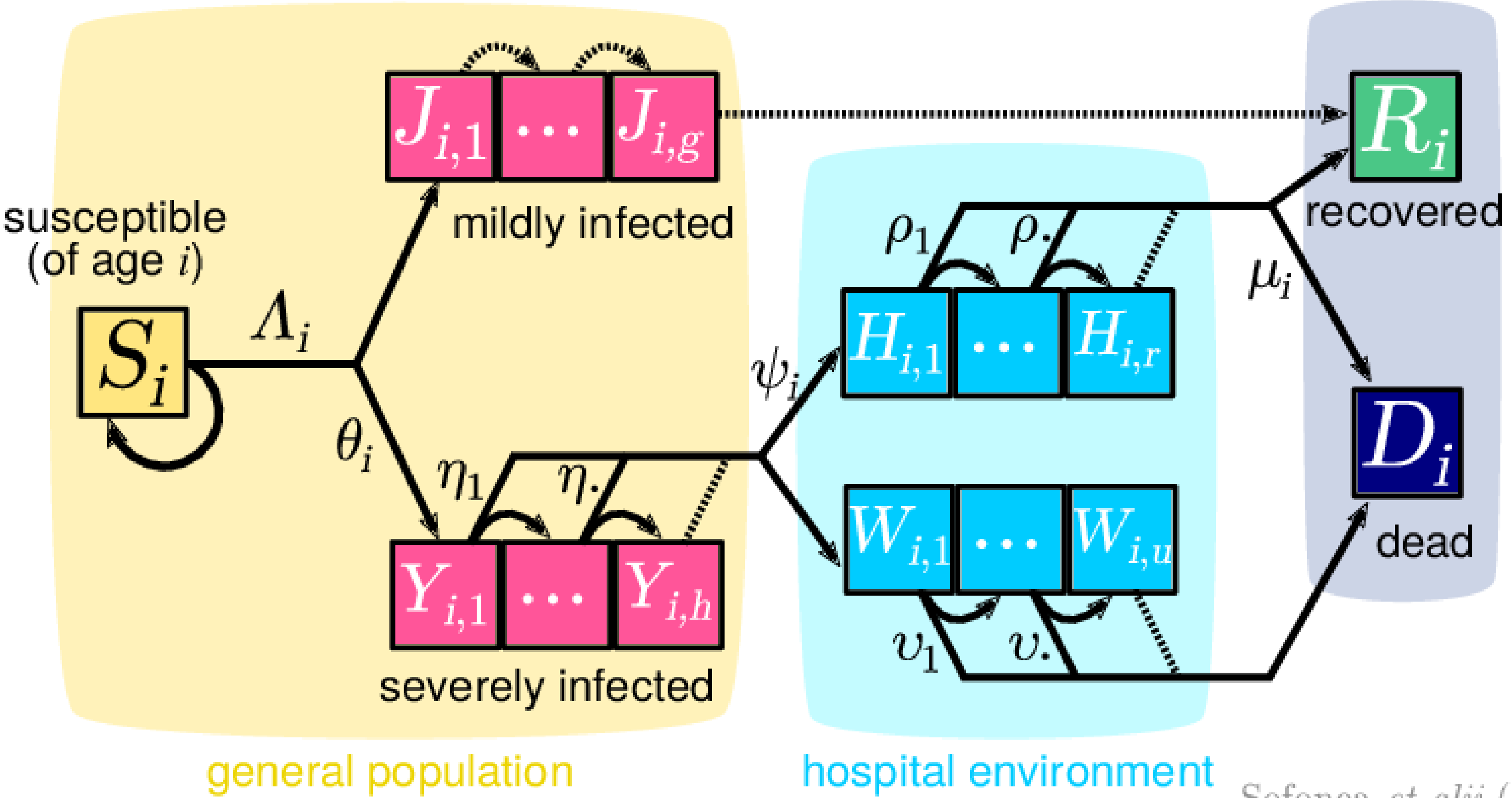


Report 3 (March 24, 2020)

Data's the key

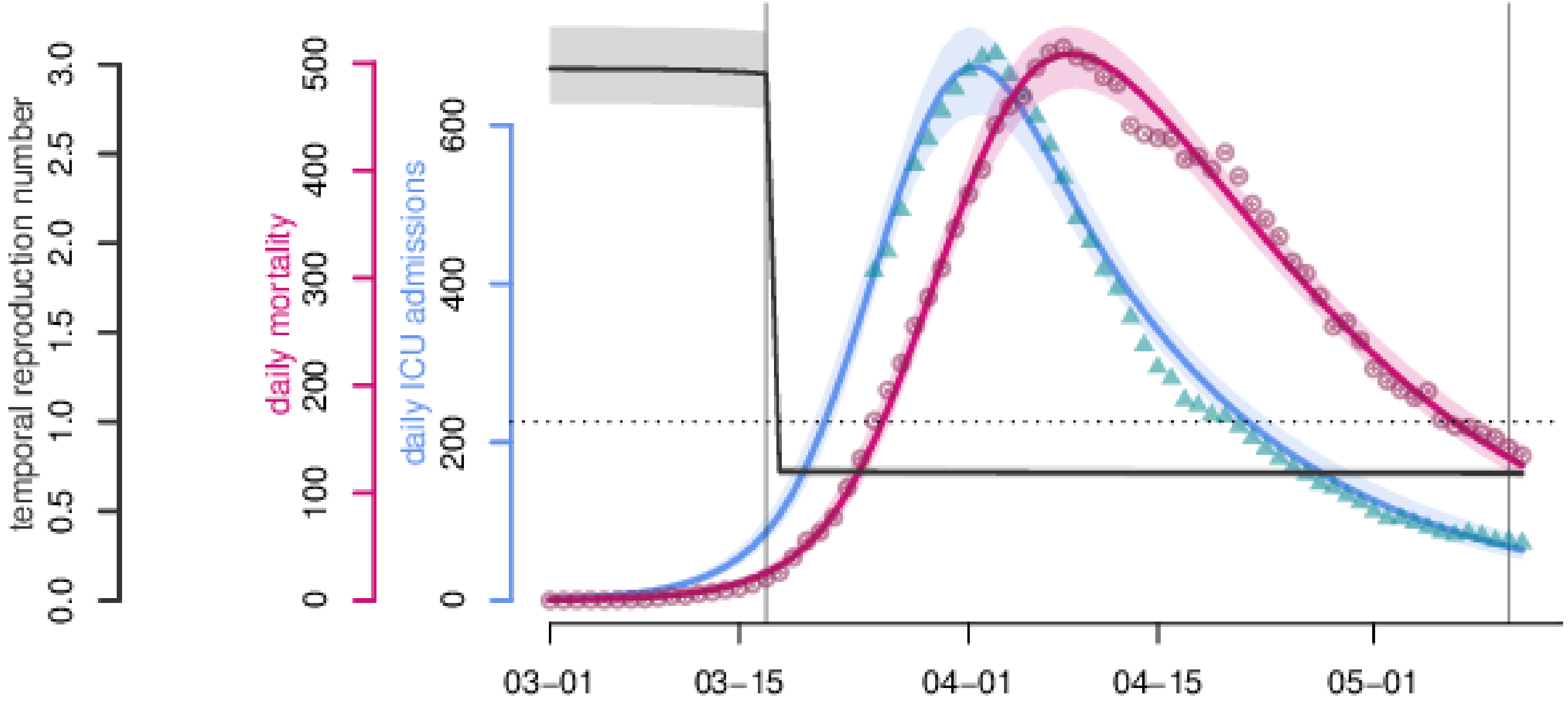
- Mar 27: sharing of hospital ICU data
- Apr 2: sharing of nursing homes data
- Apr 23: publication of lengths of stay in ICUs

Discrete time model



Sofonea *et alii* (2021, *Epidemics*)

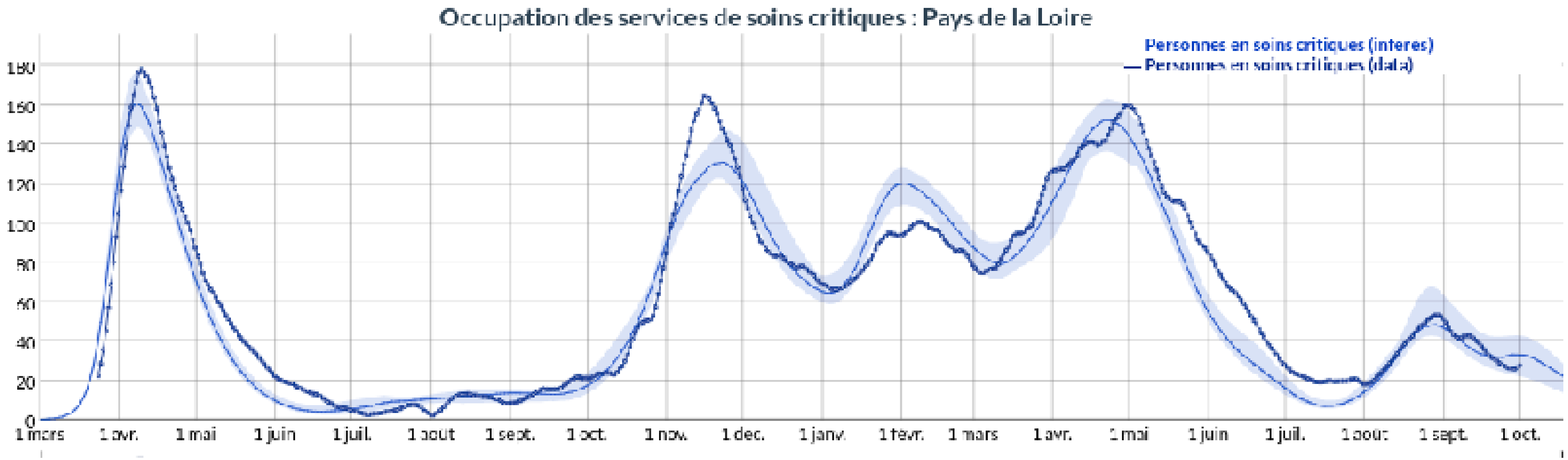
Capturing hospital dynamics



- statistical challenge
- $\mathcal{R}_0 = 3.0$ [2.6, 3.4]
- origin: 20 jan [12-28 jan]
- lock-down effect: $\mathcal{R}_t = 0.71$ [0.69, 0.74]
- $\approx 3,2$ % pop. infected on May 11

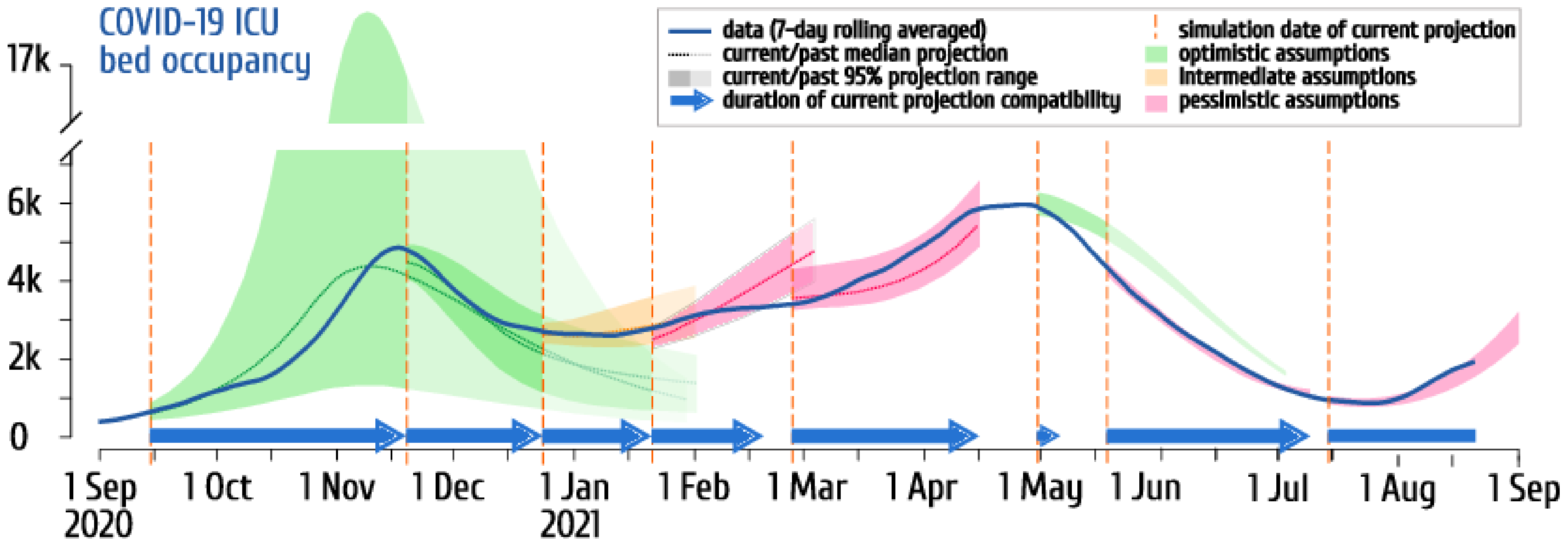
Sofonea *et alii* (2021, *Epidemics*)

COVIDici

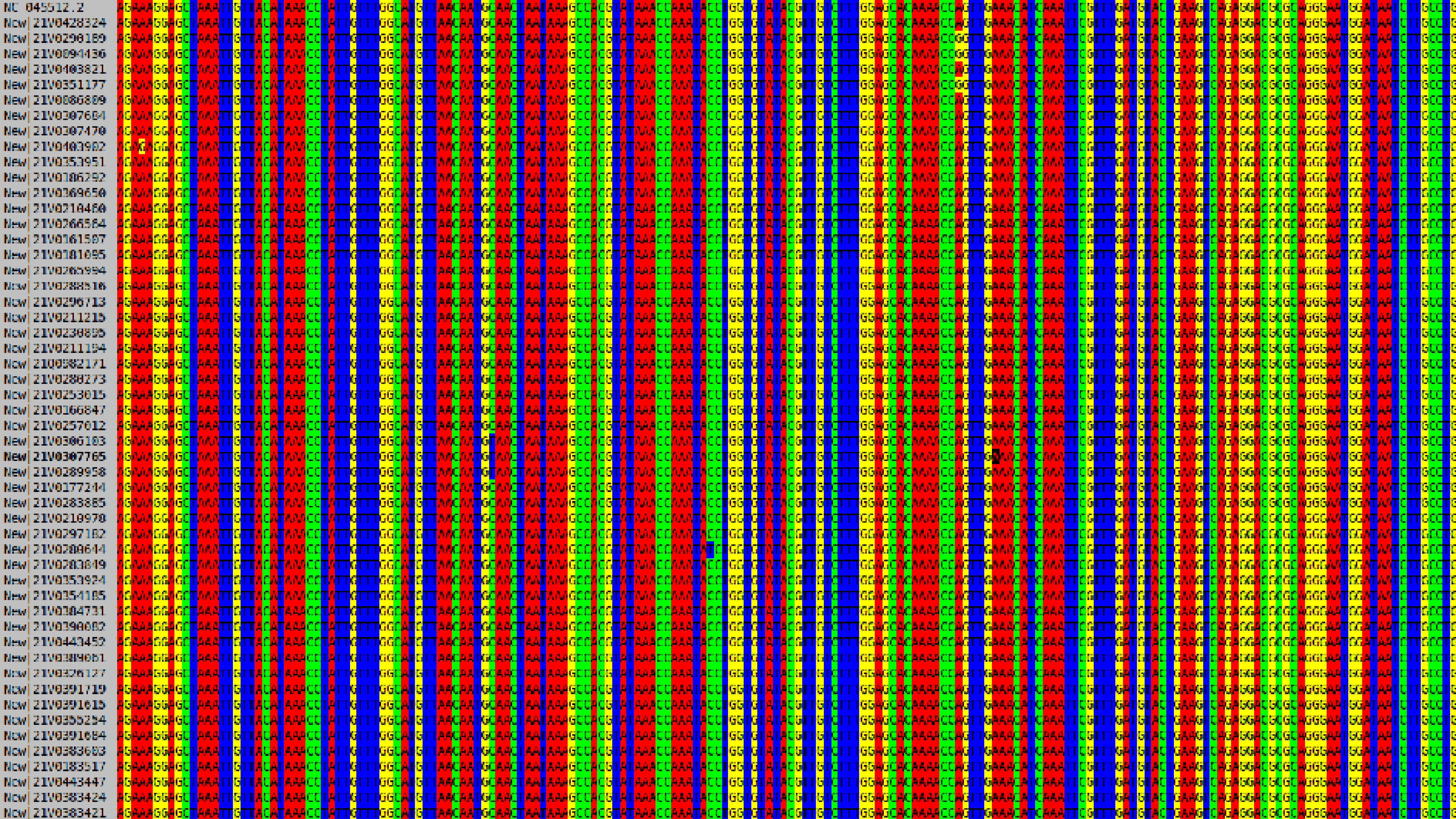


Boennec, Alizon, Sofonea (*in prep.*)
https://cloudapps.france-bioinformatique.fr/covidici/

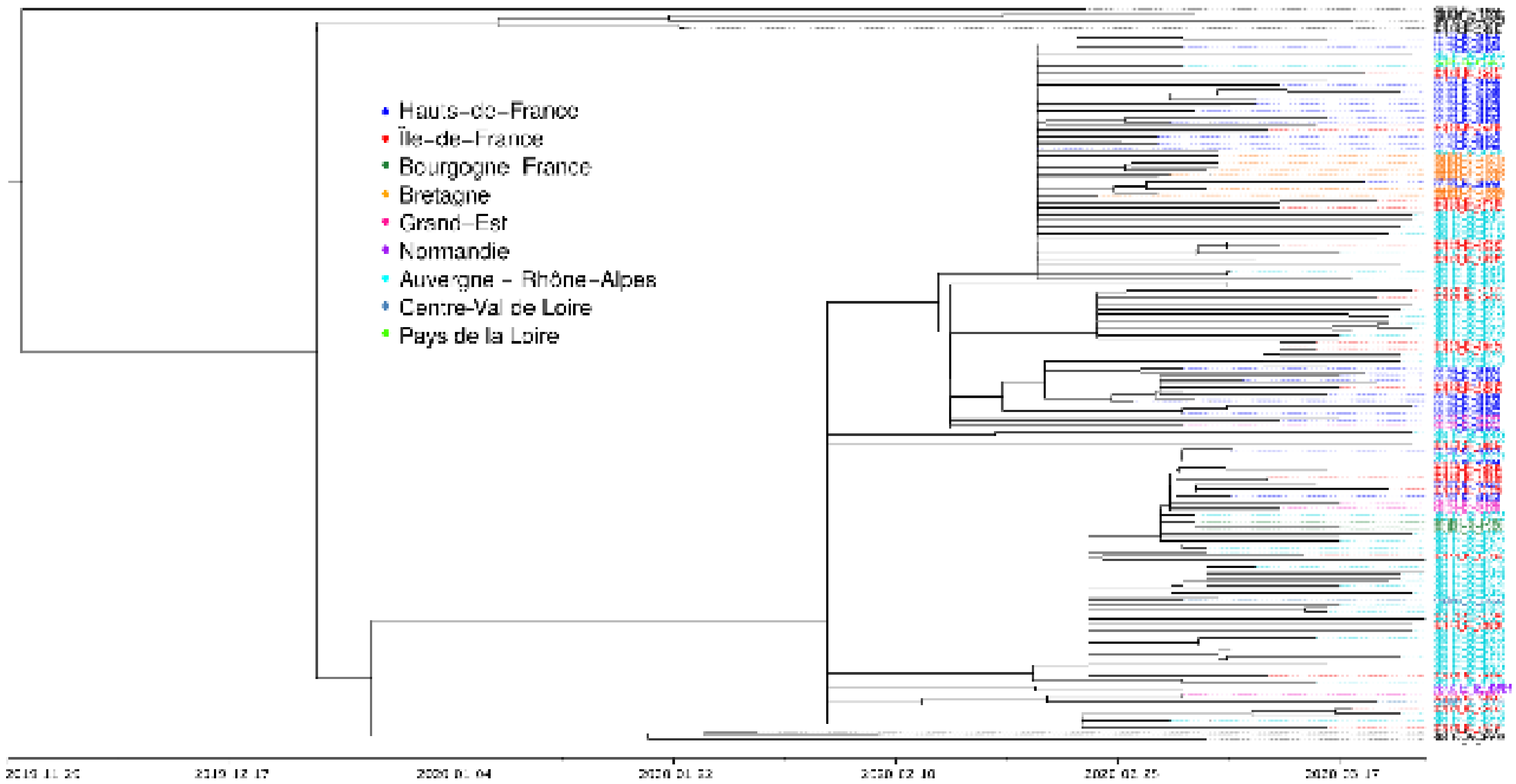
5 WEEKS INSIGHTS ON PROSPECTIVE TRENDS



Sofonea *et alii* (2021, *ACCPM*)


NC 045512.2
New 21V0428324
New 21V0290189
New 21V0094436
New 21V0403821
New 21V0351177
New 21V0086809
New 21V0307604
New 21V0307470
New 21V0403902
New 21V0353951
New 21V0106292
New 21V0369650
New 21V0210460
New 21V0266564
New 21V0161507
New 21V0181095
New 21V0265994
New 21V0288510
New 21V0290713
New 21V0211215
New 21V0230895
New 21V0211194
New 21V0382171
New 21V0280273
New 21V0253015
New 21V0166847
New 21V0257012
New 21V0306103
New 21V0307765
New 21V0289958
New 21V0177244
New 21V0283885
New 21V0210970
New 21V0297182
New 21V0280644
New 21V0283849
New 21V0353924
New 21V0354105
New 21V0304731
New 21V0390082
New 21V0443452
New 21V0389061
New 21V0326127
New 21V0391719
New 21V0391615
New 21V0355254
New 21V0391684
New 21V0383603
New 21V0183517
New 21V0443447
New 21V0383424
New 21V0383421

French SARS-CoV-2 phylogeny

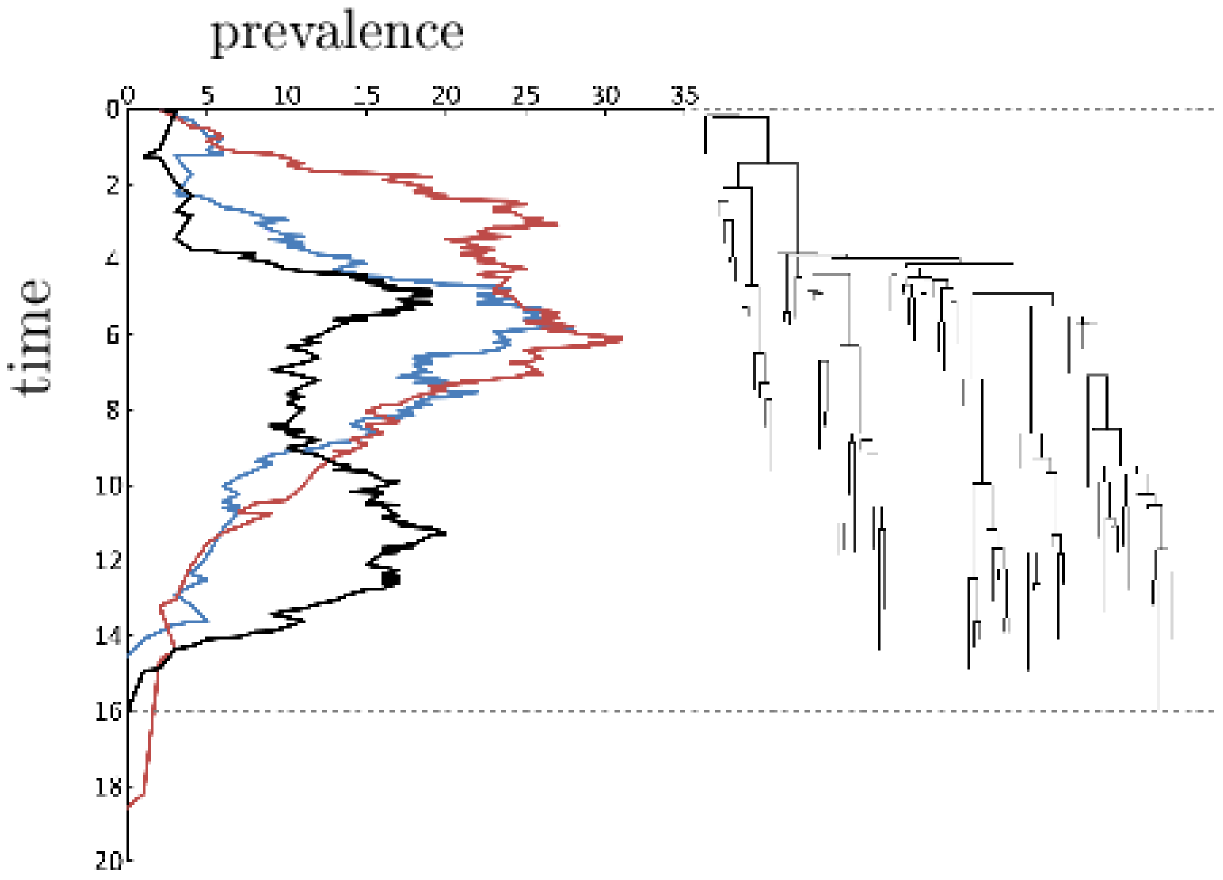


- 196 sequences
- 186 in the epidemic wave clade
- origin between Jan 19 and Feb 15

Report 6 (Apr 9, 2020)

Danesh *et alii* (*Peer Community In Journal, in press*)

Viral phylodynamics

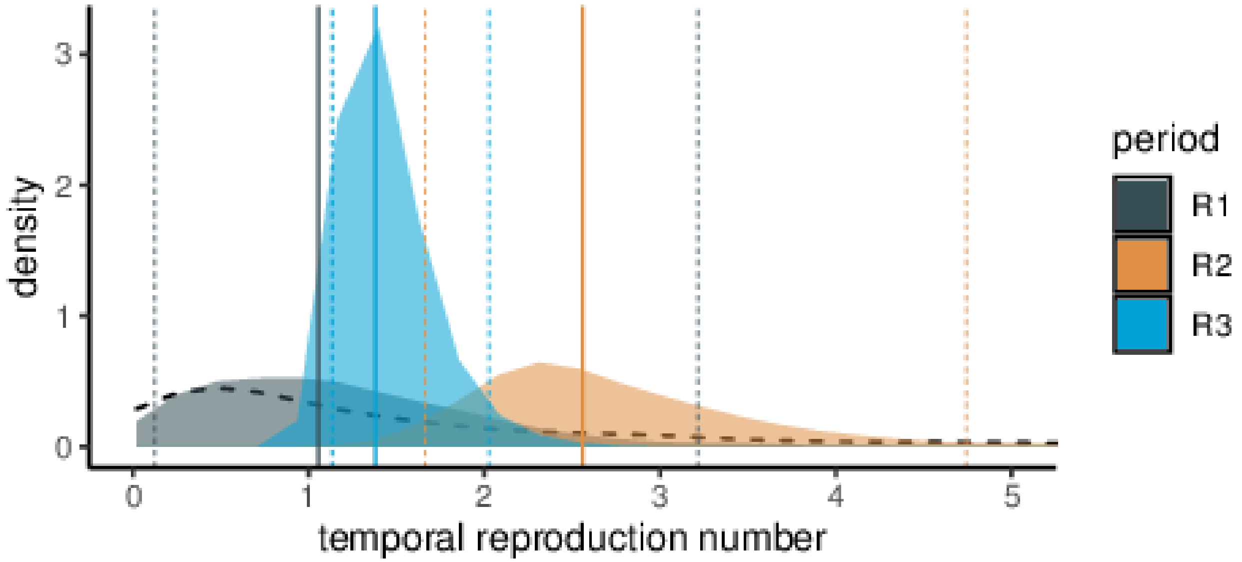


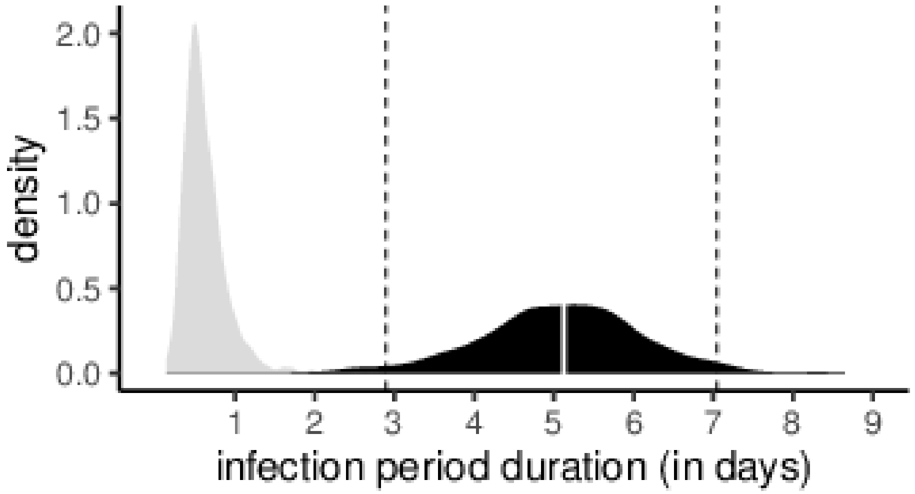
- **dating** common ancestors
- measure **spreading** $(r, \mathcal{R}_t)$
- estimate infection **durations**
- uncover spatial **structure**

Volz *et alii* (2013, *PLoS Comput Biol*)

French phylodynamics

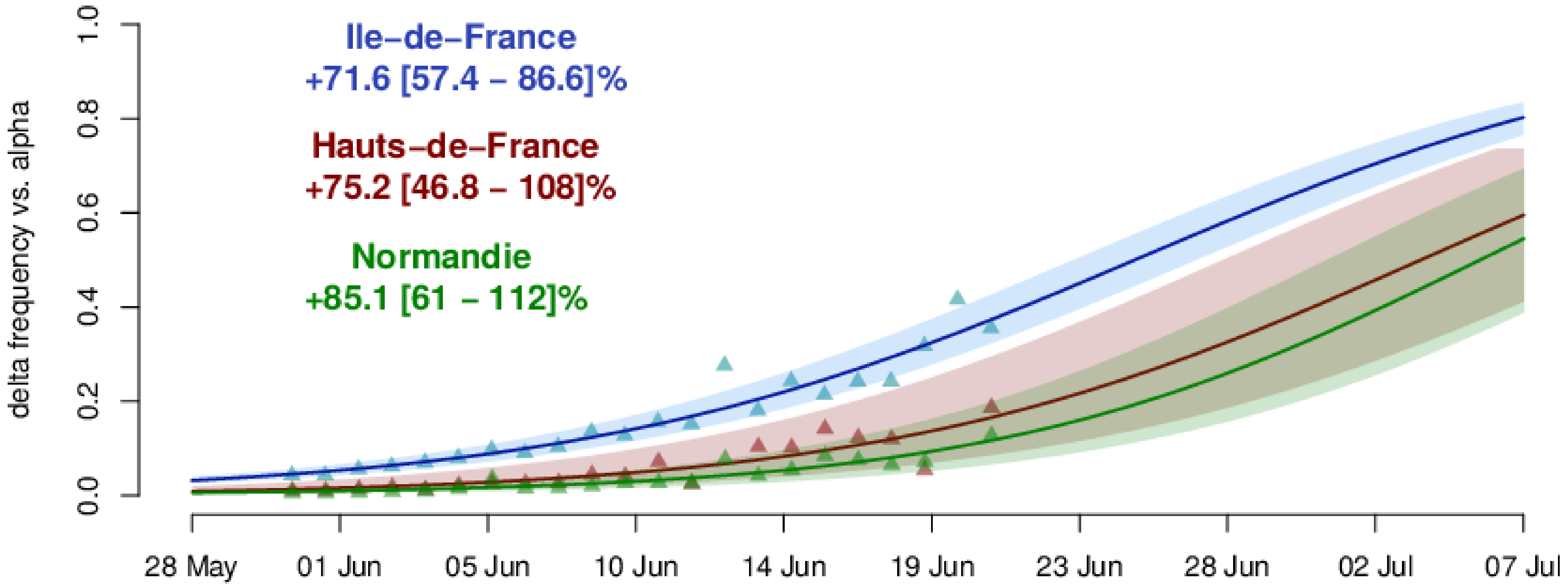
Using a *birth-death skyline plot* model (BDSKY):





Report 6 (Apr 9, 2020)
Danesh *et alii* (2020, *PCI Evol Biol*)

Phenotypic evolution



Alizon *et alii* (2021, *Eurosurveillance*)

Phenotypic evolution

Since 2021, we see "variants" with one (or more) properties:

- increased virulence (Alpha, Beta, Delta)
- increased contagiousness (Alpha, Beta, Gamma, Delta)
- immune escape (Beta, Gamma)
- resistance to treatments (?)

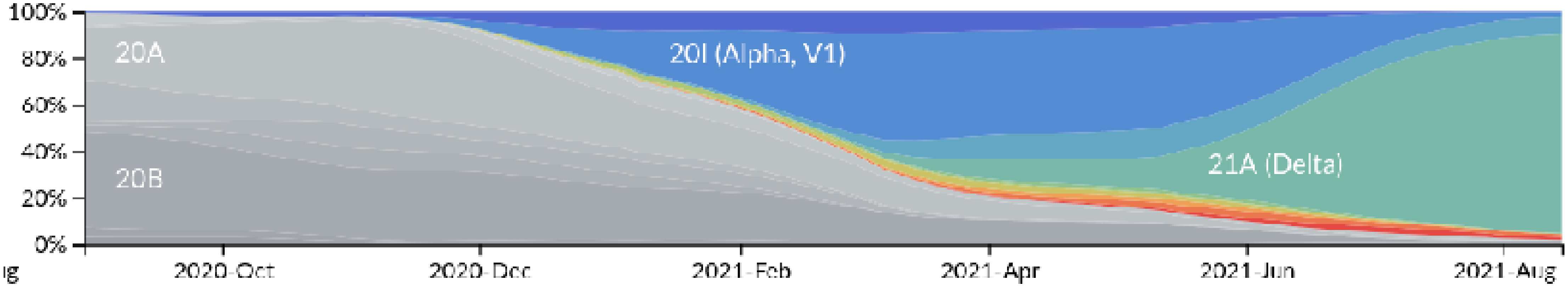
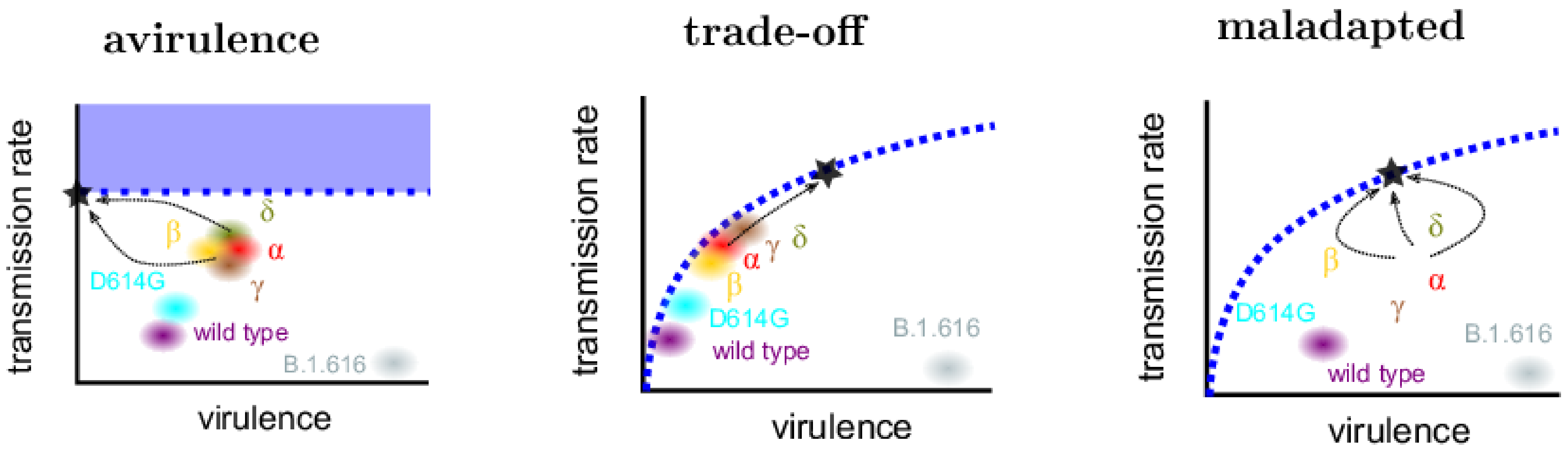


Image from Nextstrain

Evolutionary scenarios



Alizon & Sofonea (2021, *J Evol Biol*)

Long-term hope

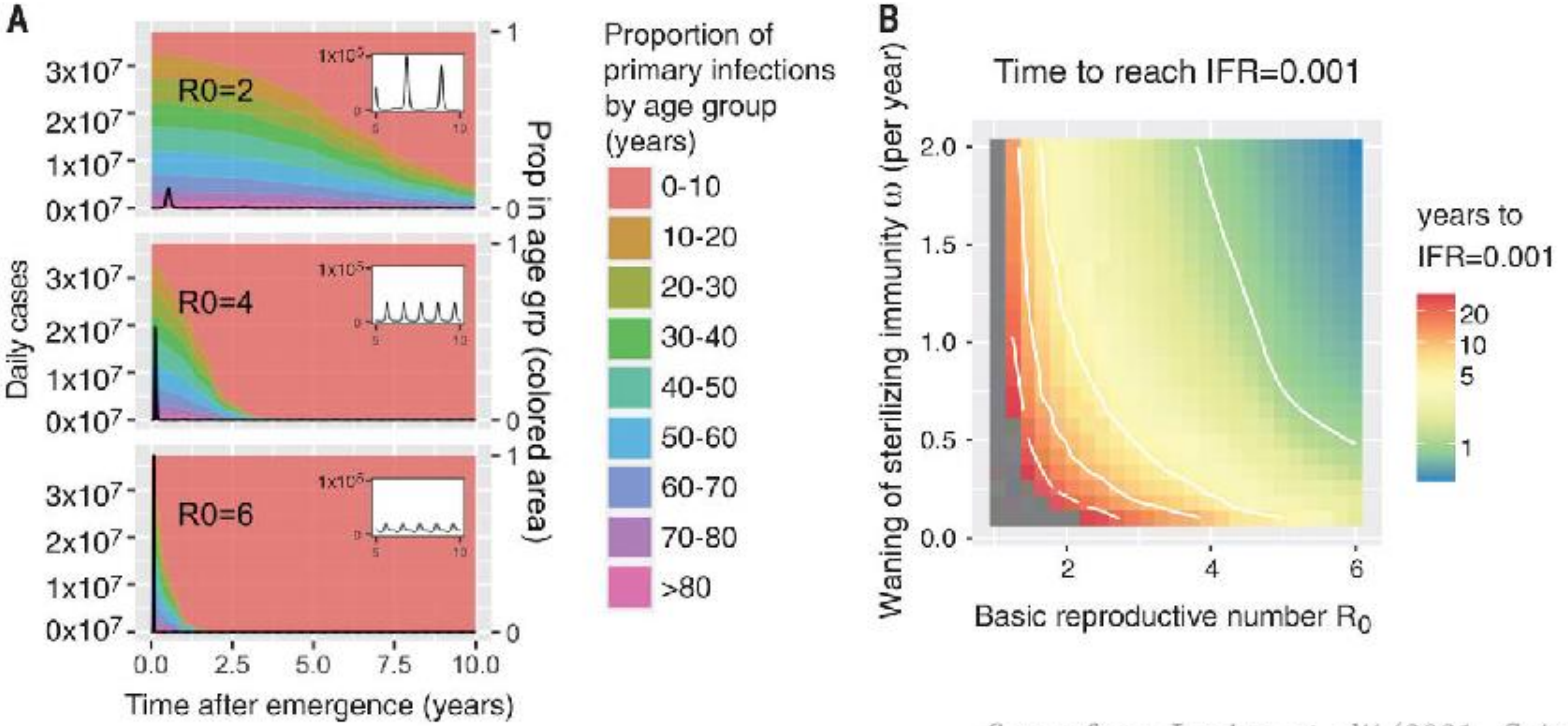
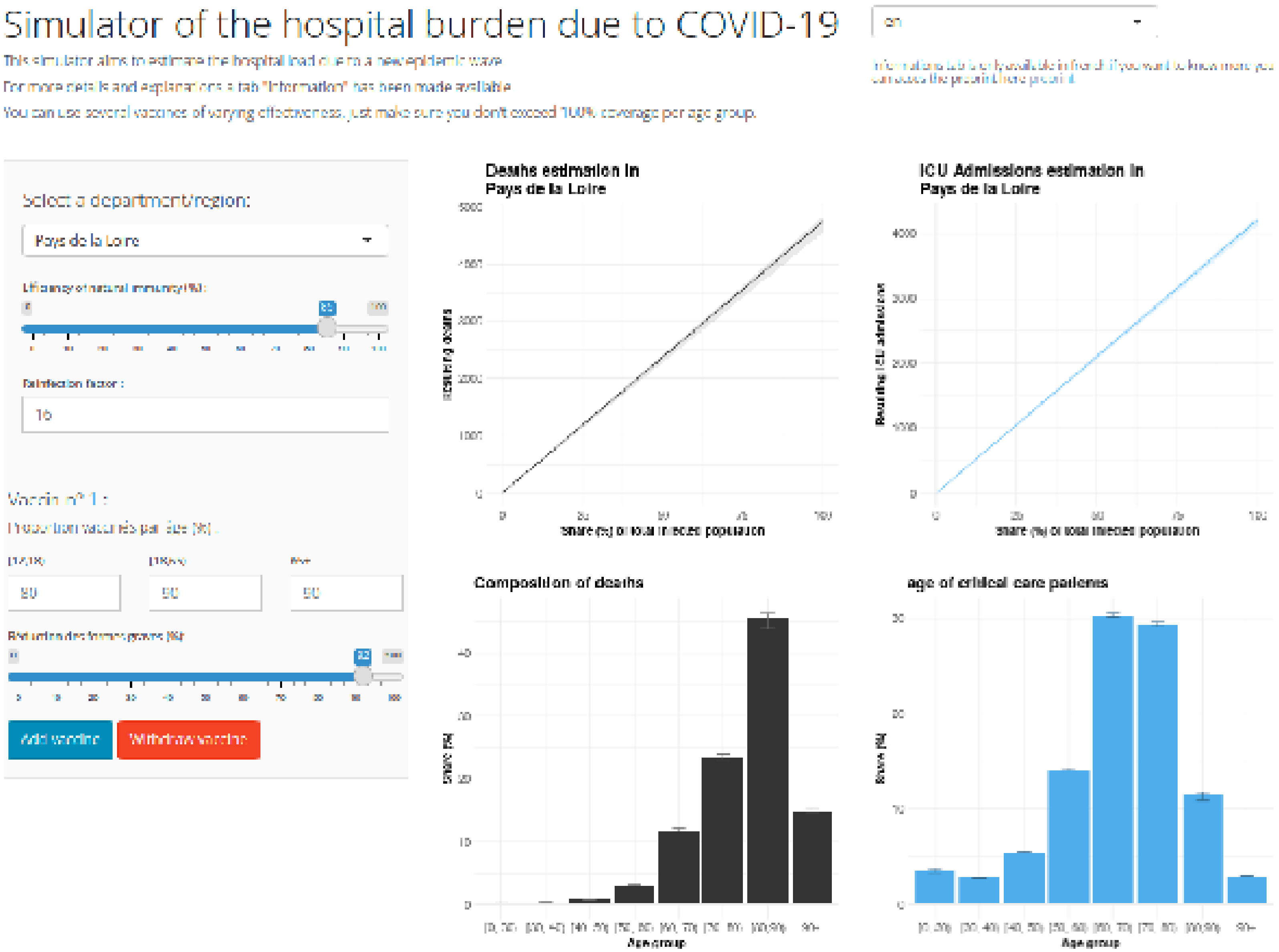


figure from Lavine *et alii* (2021, *Science*)

Current vaccines might not be enough



SARS-CoV-2 emergence and evolution

- Early (and final) outbreal dynamics are **stochastic**.
- We can **anticipate** hospital dynamics to some extent (5 weeks).
- Neutral evolution can be used to track the epidemic (**phylodynamics**).
- Phenotypic evolution (‘**variants**’) threatens epidemic control
- **Vaccination** is probably not enough (to control and to let a wave pass)

Appendix

2. ICU modelling

3. Phylodynamics

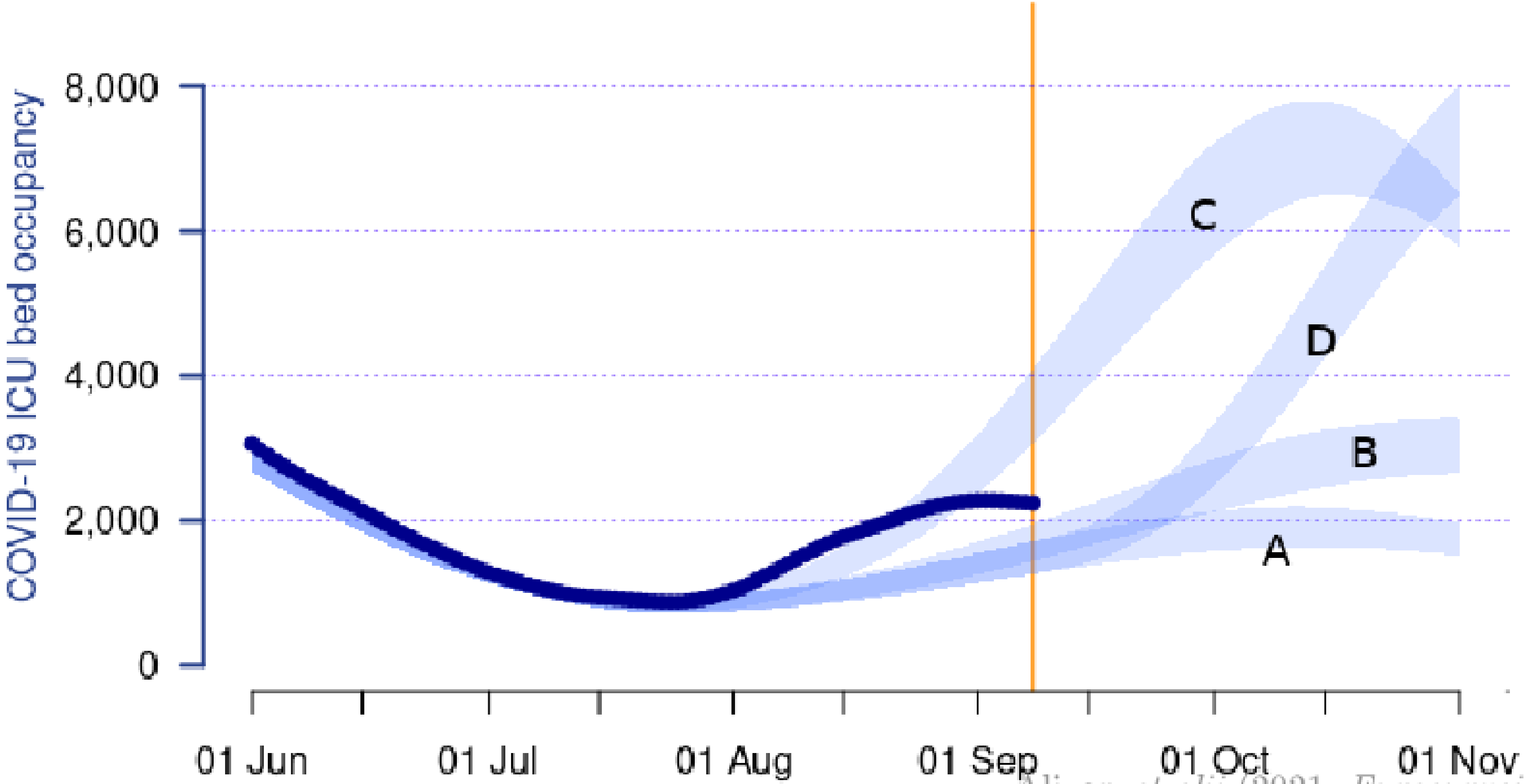
4. EPIDEMAP

5. Optimal age-based control

6. Within-host kinetics

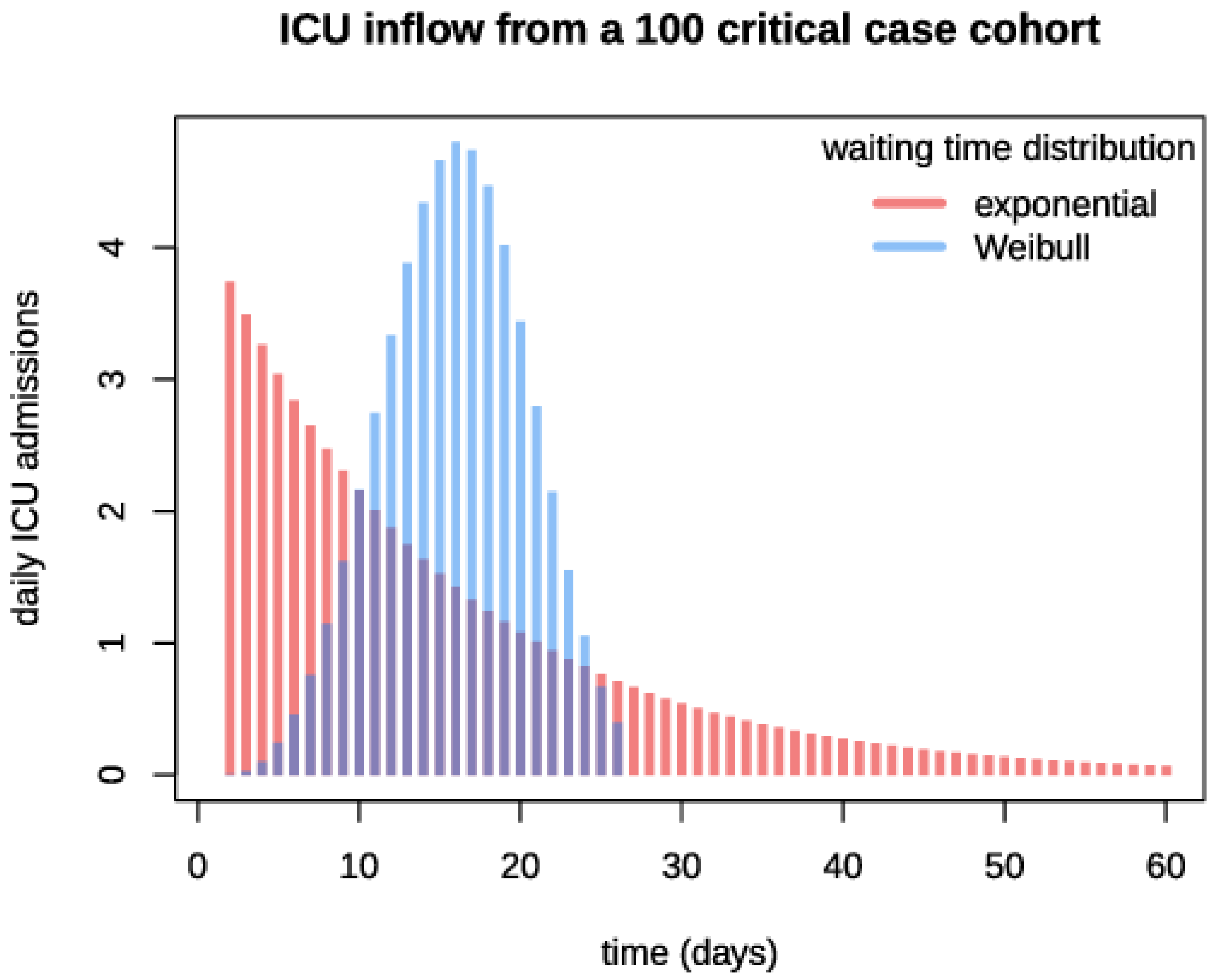
7. Variant transmission advantage

Prospective scenarios



Alizon *et alii* (2021, *Eurosurveillance*)

Delays in ICU dynamics



Bayesian phylodynamics

- data: dated sequences
- T: phylogenetic tree
- η : parameters from the epidemiological model
- θ : parameters from the evolutionary model (substitutions)
- $f\,[\text{data}|T,\theta]$: phylogenetic likelihood
- **$f\,[T|\eta]$: transmission model (coalescent, birth-death model)**
- $f\,[\eta,\theta]$: prior distribution
- $f\,[\text{data}]$: normalisation term

Posterior probability distribution:

$$f\,[T,\eta,\theta|\text{data}] = \frac{f\,[\text{data}|T,\theta]\;\; f\,[T|\eta]\;\; f\,[\eta,\theta]}{f\,[\text{data}]}$$

Drummond *et alii* (2002, *Genetics*)
for a review, see Kühnert *et alii* (2011, *Inf Genet Evol*)

French COVID modelling

2. ICU modelling

3. Phylodynamics

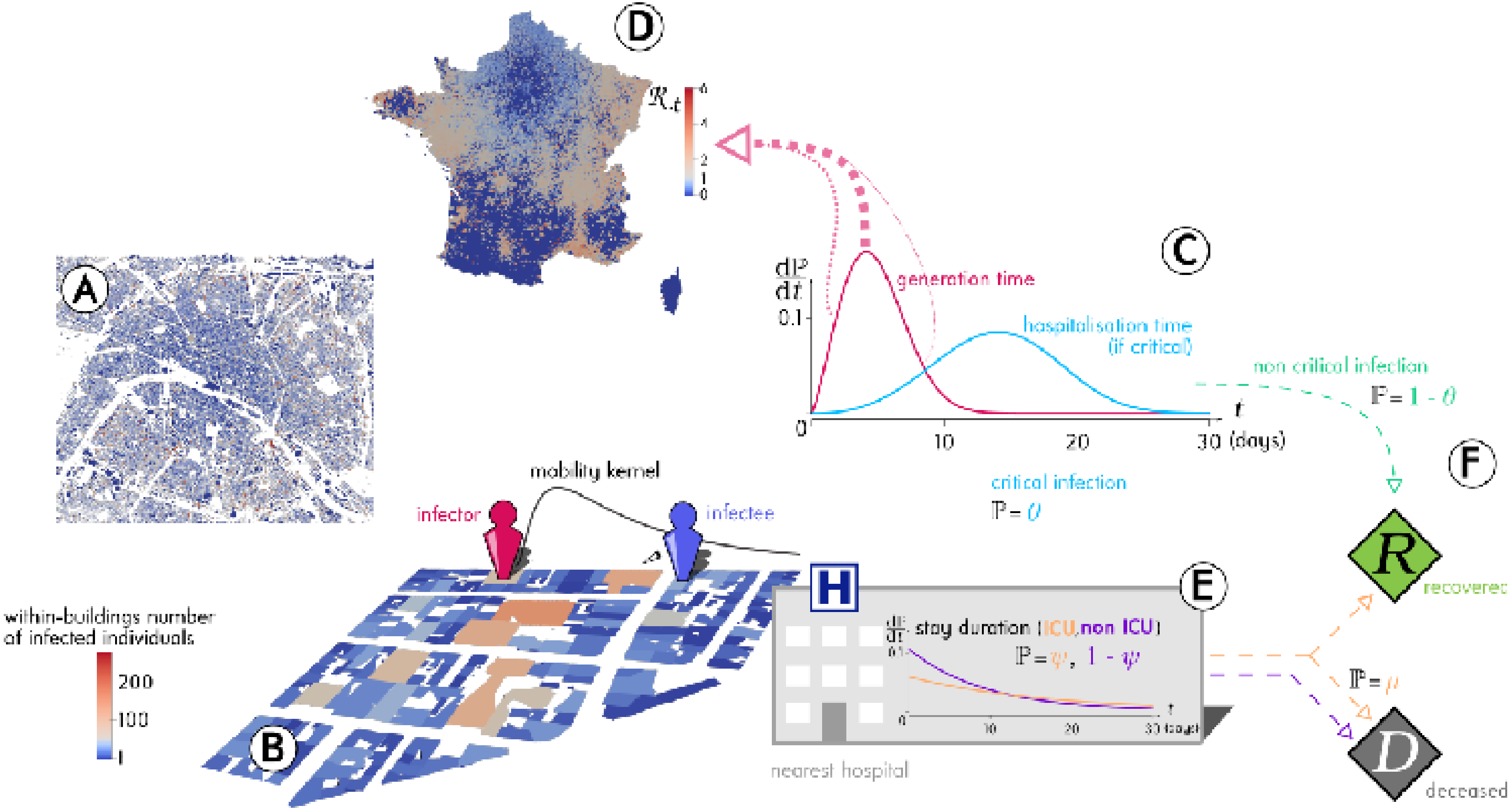
4. EPIDEMAP

5. Optimal age-based control

6. Within-host kinetics

7. Variant transmission advantage

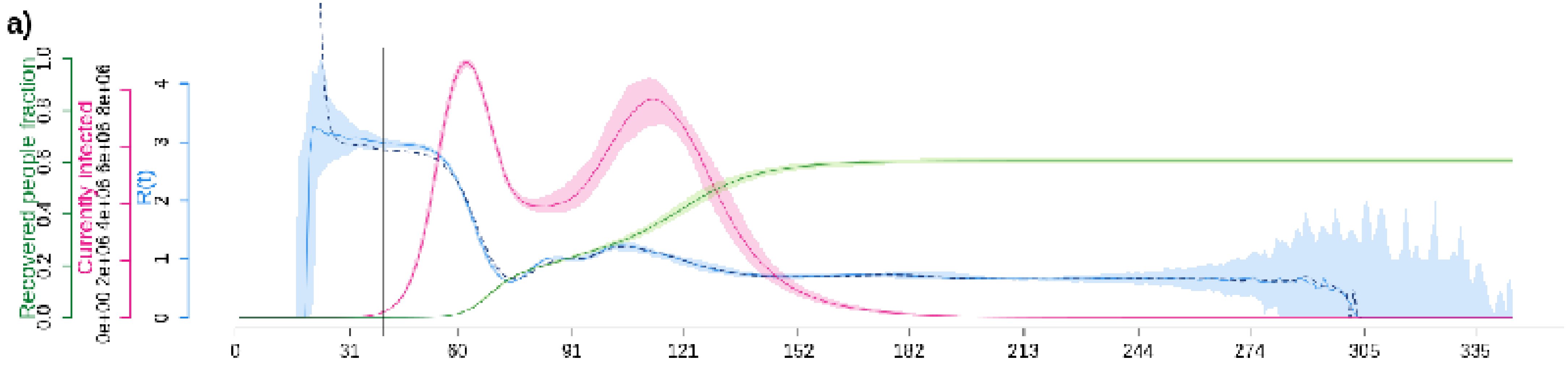
Thomine *et alii* (*Zenodo*)

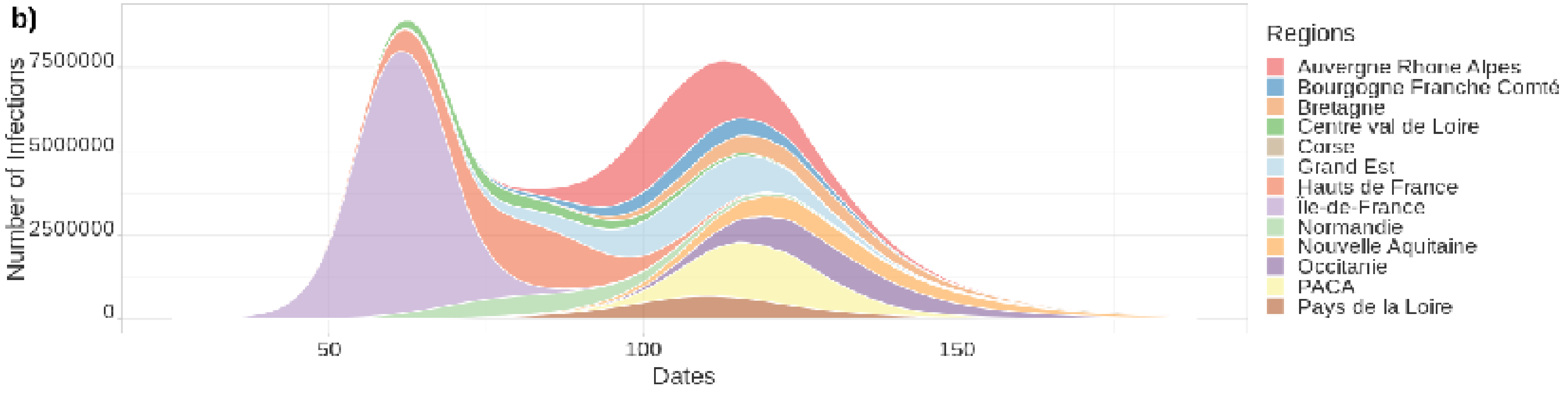


EPIDEMAP

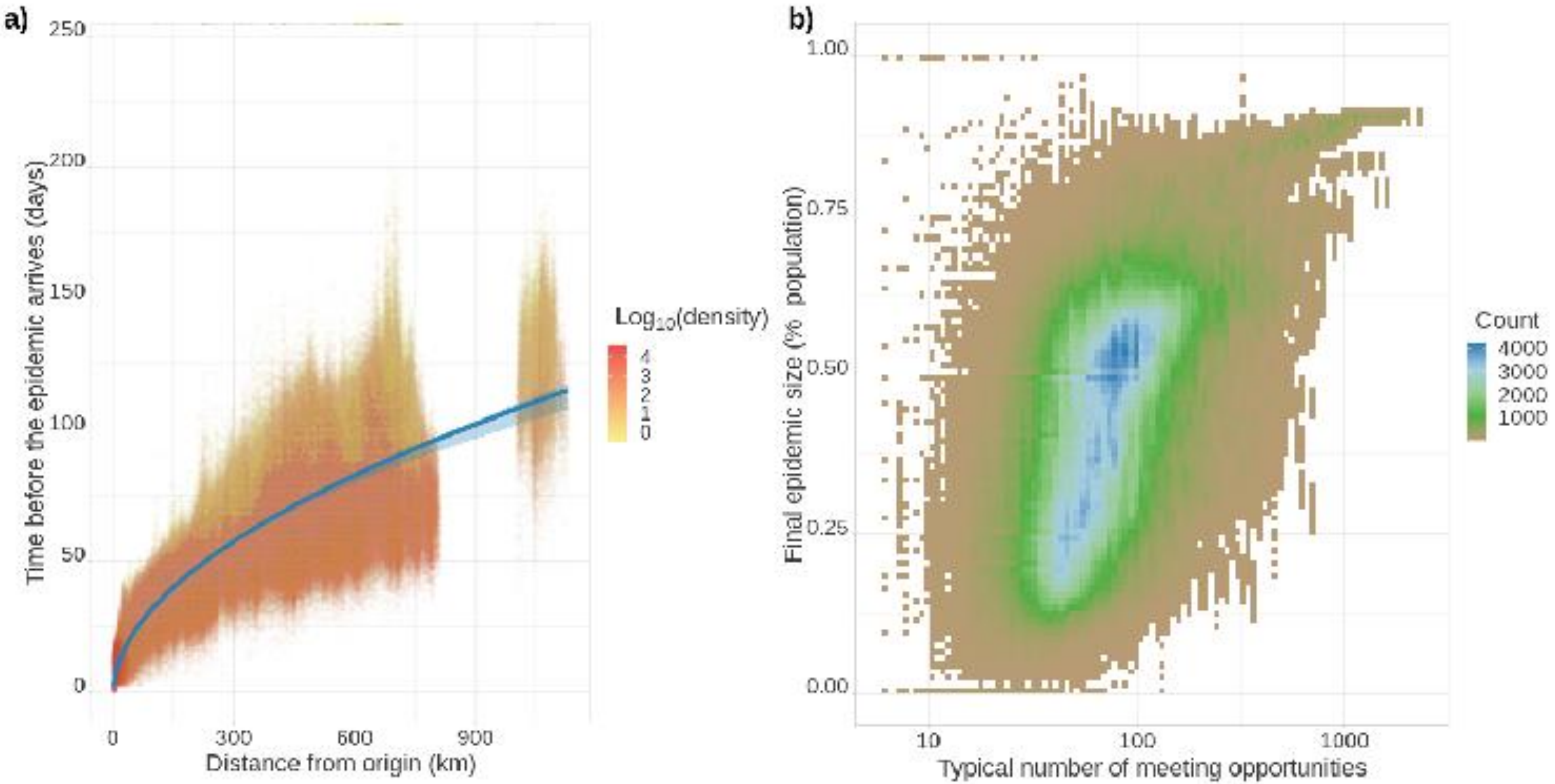
Thomine *et alii* (*Zenodo*)

Epidemic course

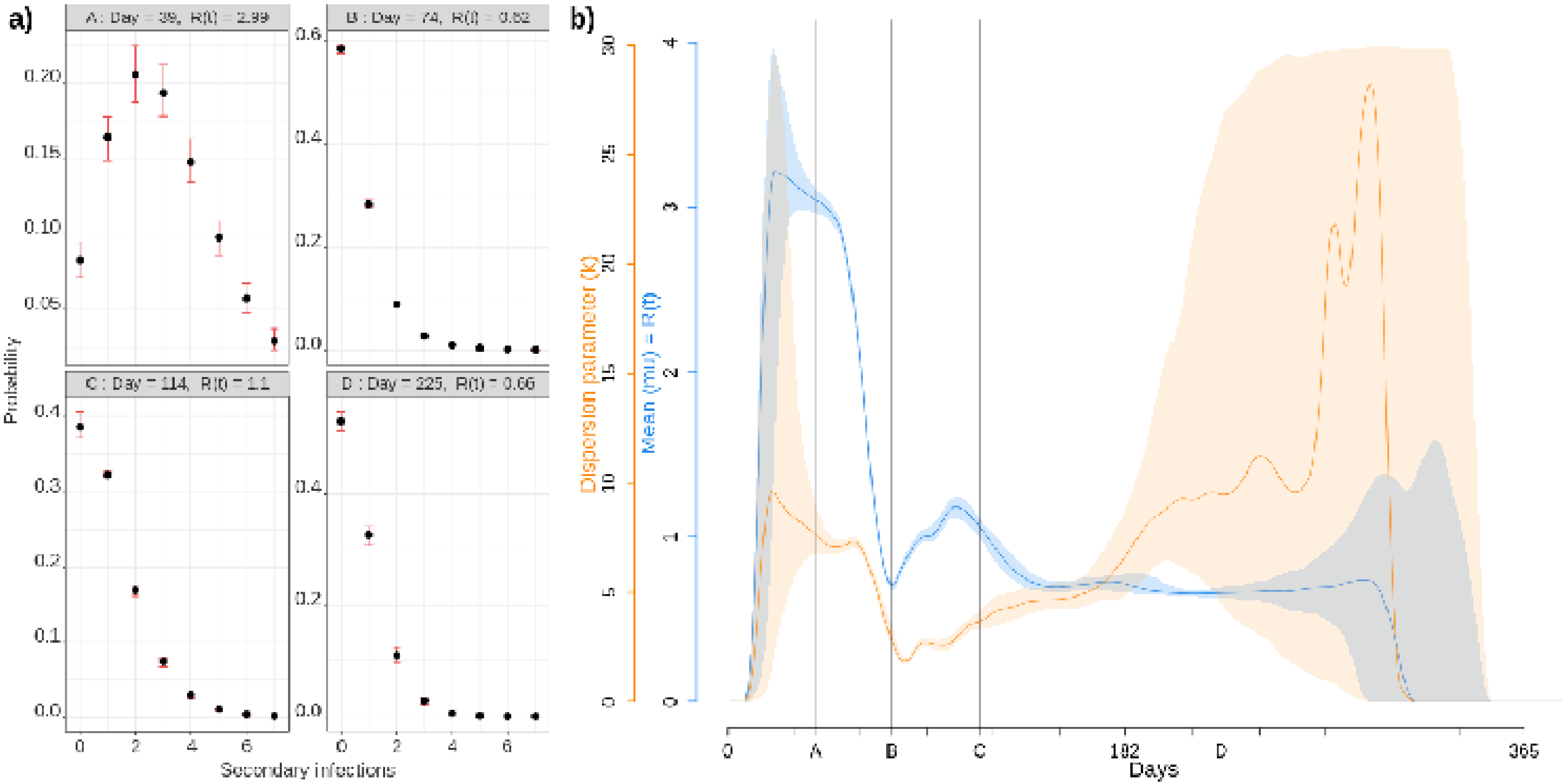




Spatial resolution



Transmission chains



French COVID modelling

2. ICU modelling

3. Phylodynamics

4. EPIDEMAP

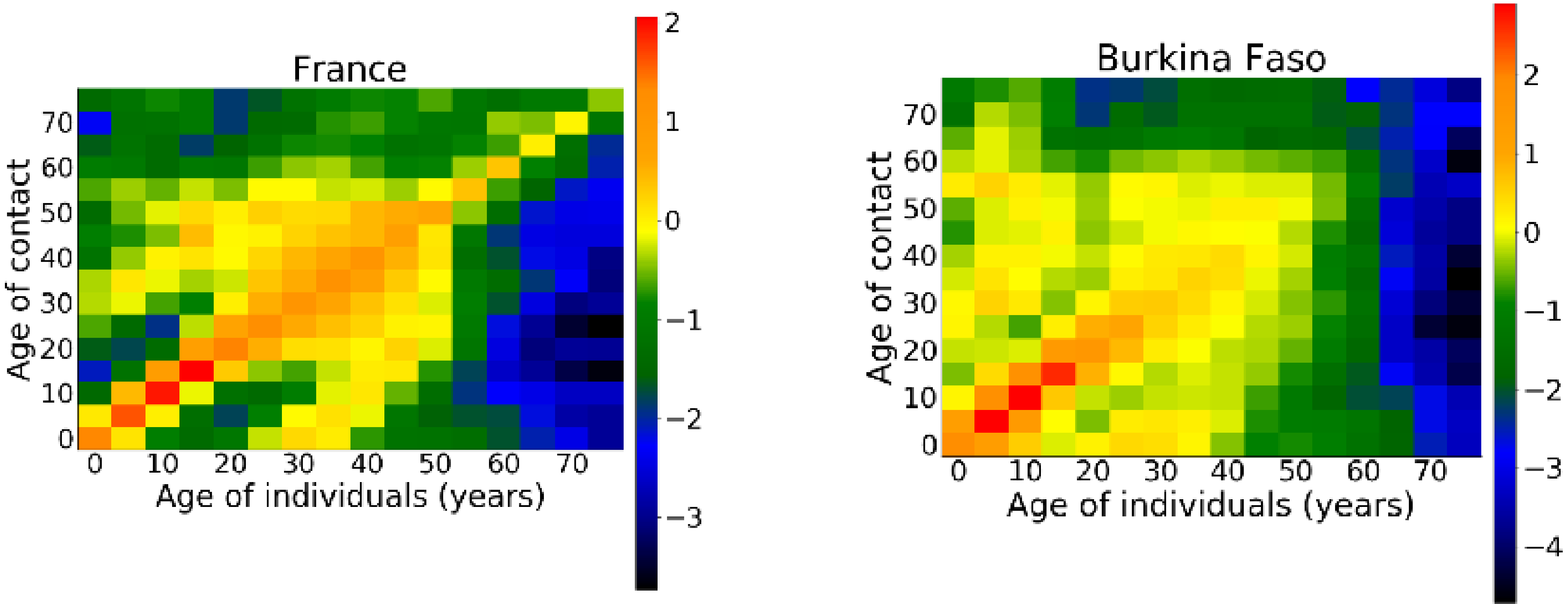
5. Optimal age-based control

6. Within-host kinetics

7. Variant transmission advantage

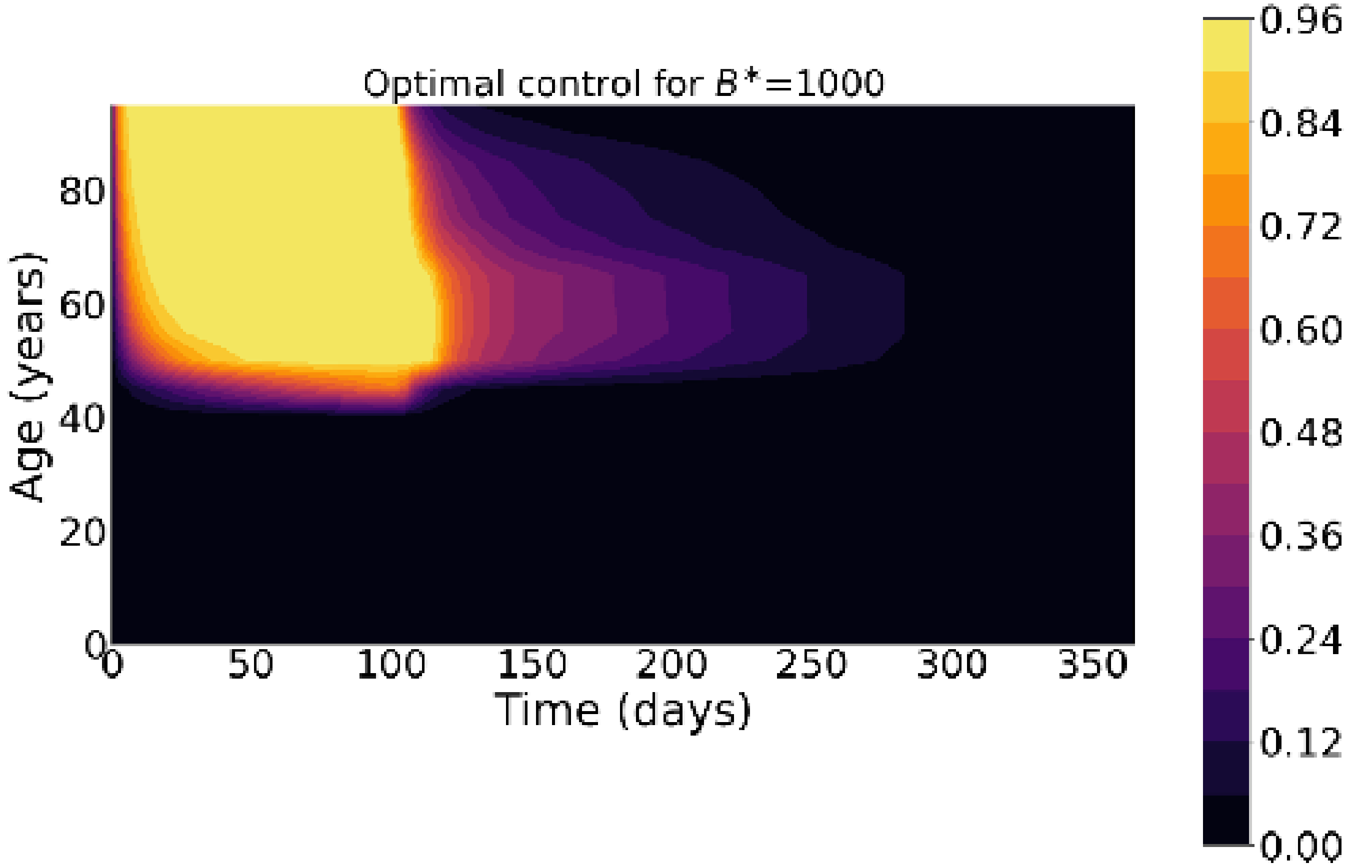
Richard *et alii* (2021, *PLoS Comput Biol*)

Age-stratified contact matrix



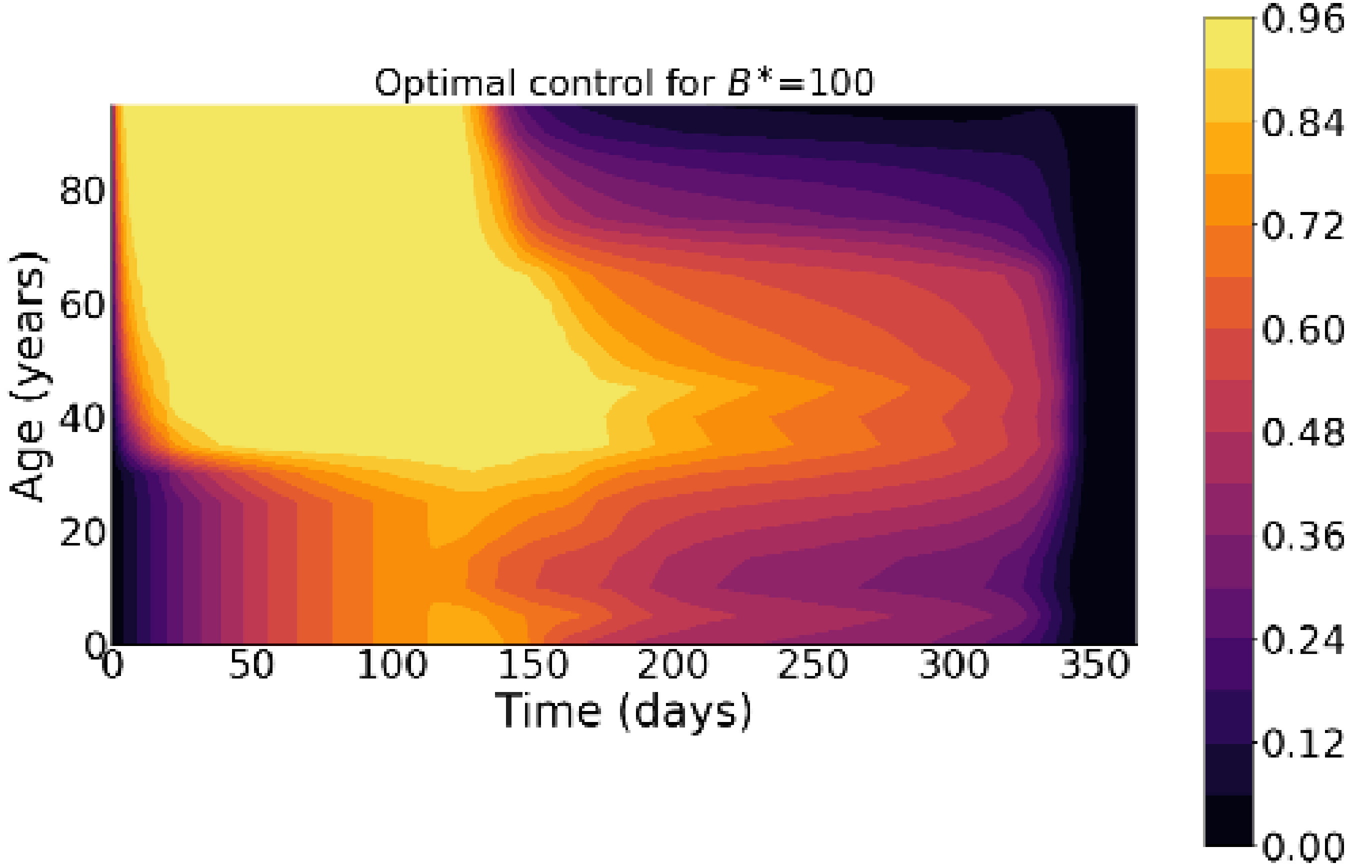
Richard *et alii* (2021, *PLoS Comput Biol*)

Optimal control (costly)



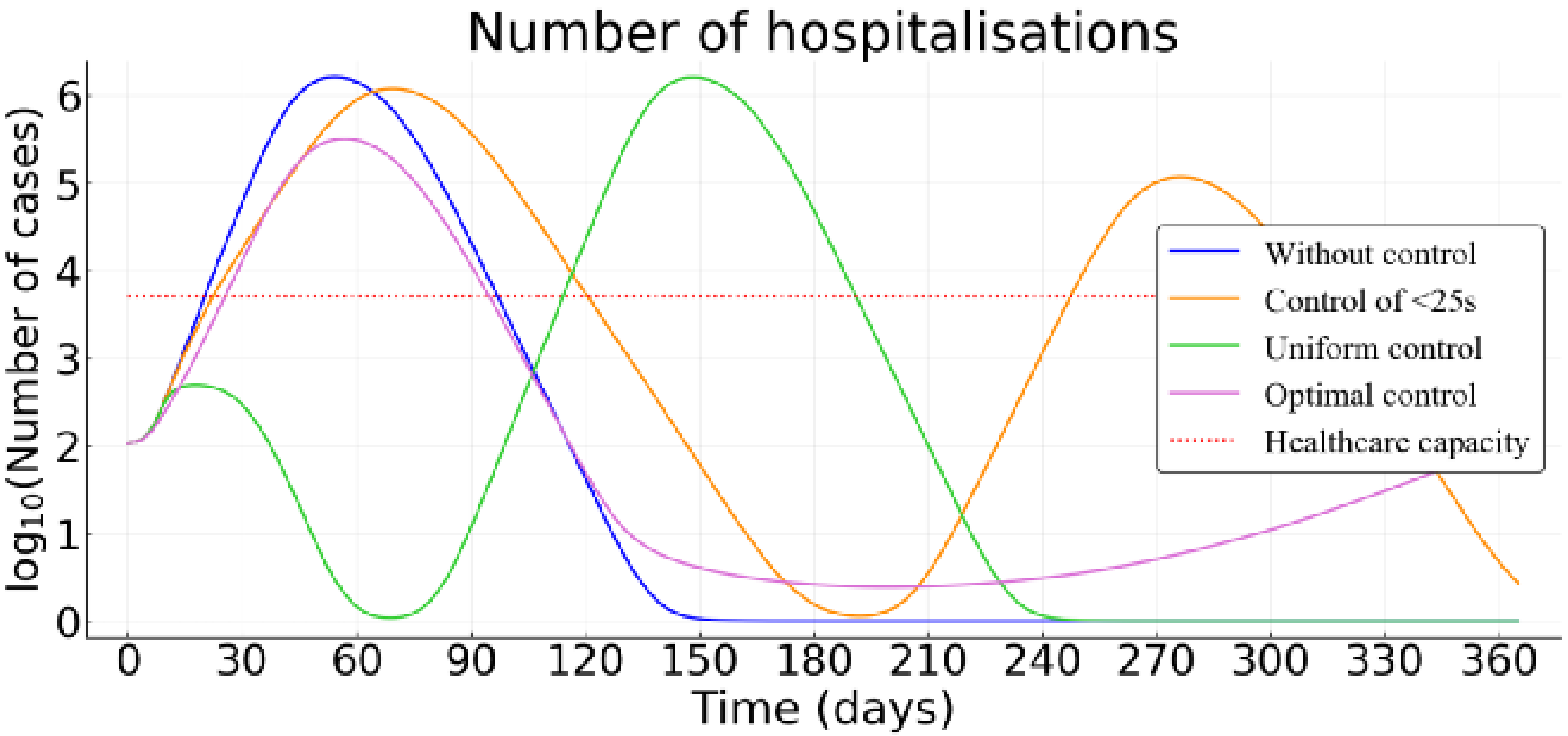
Richard *et alii* (2021, *PLoS Comput Biol*)

Optimal control (cheap)



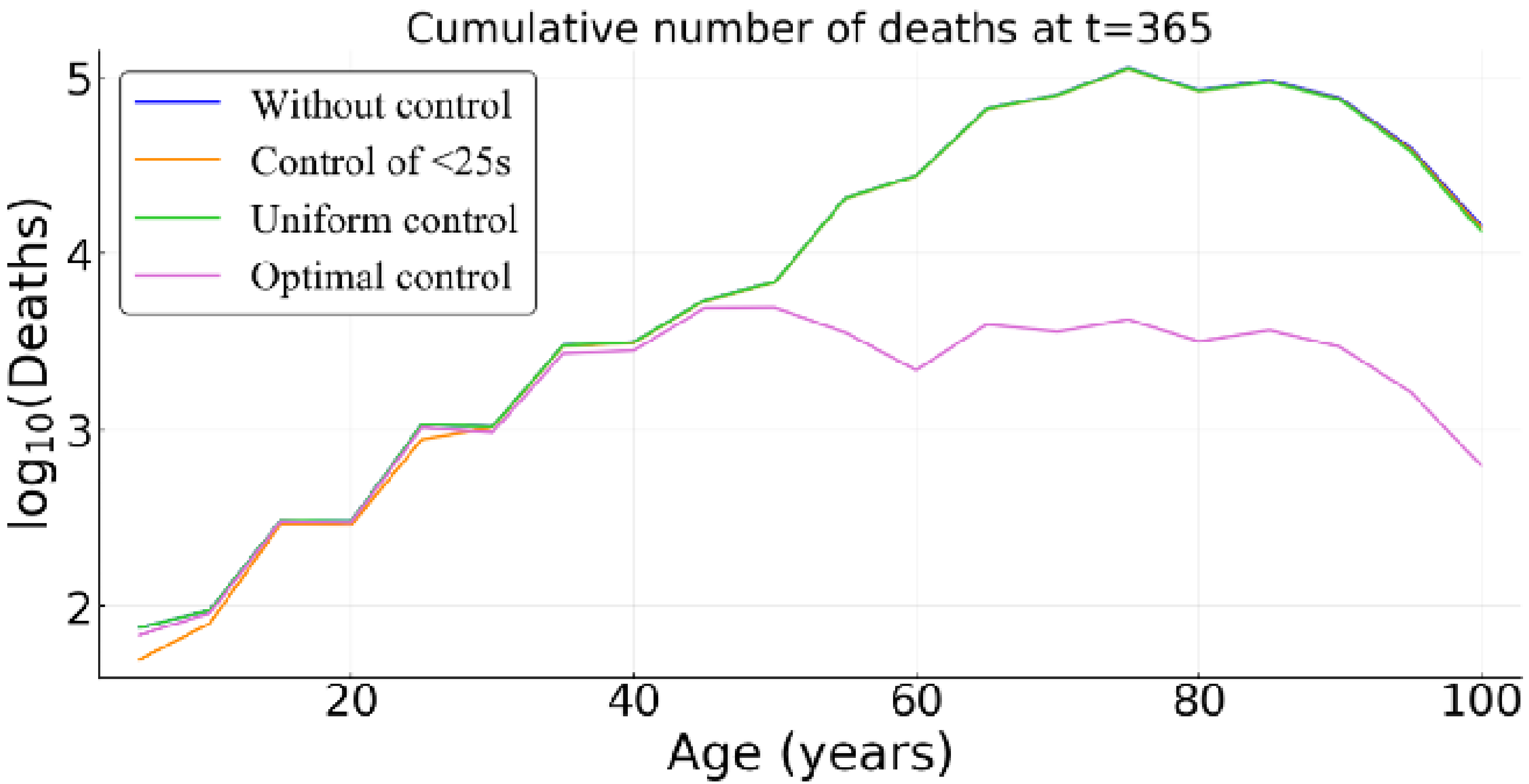
Richard *et alii* (2021, *PLoS Comput Biol*)

Comparing control scenarios



Richard *et alii* (2021, *PLoS Comput Biol*)

Comparing control scenarios



Richard *et alii* (2021, *PLoS Comput Biol*)

French COVID modelling

2. ICU modelling

3. Phylodynamics

4. EPIDEMAP

5. Optimal age-based control

6. Within-host kinetics

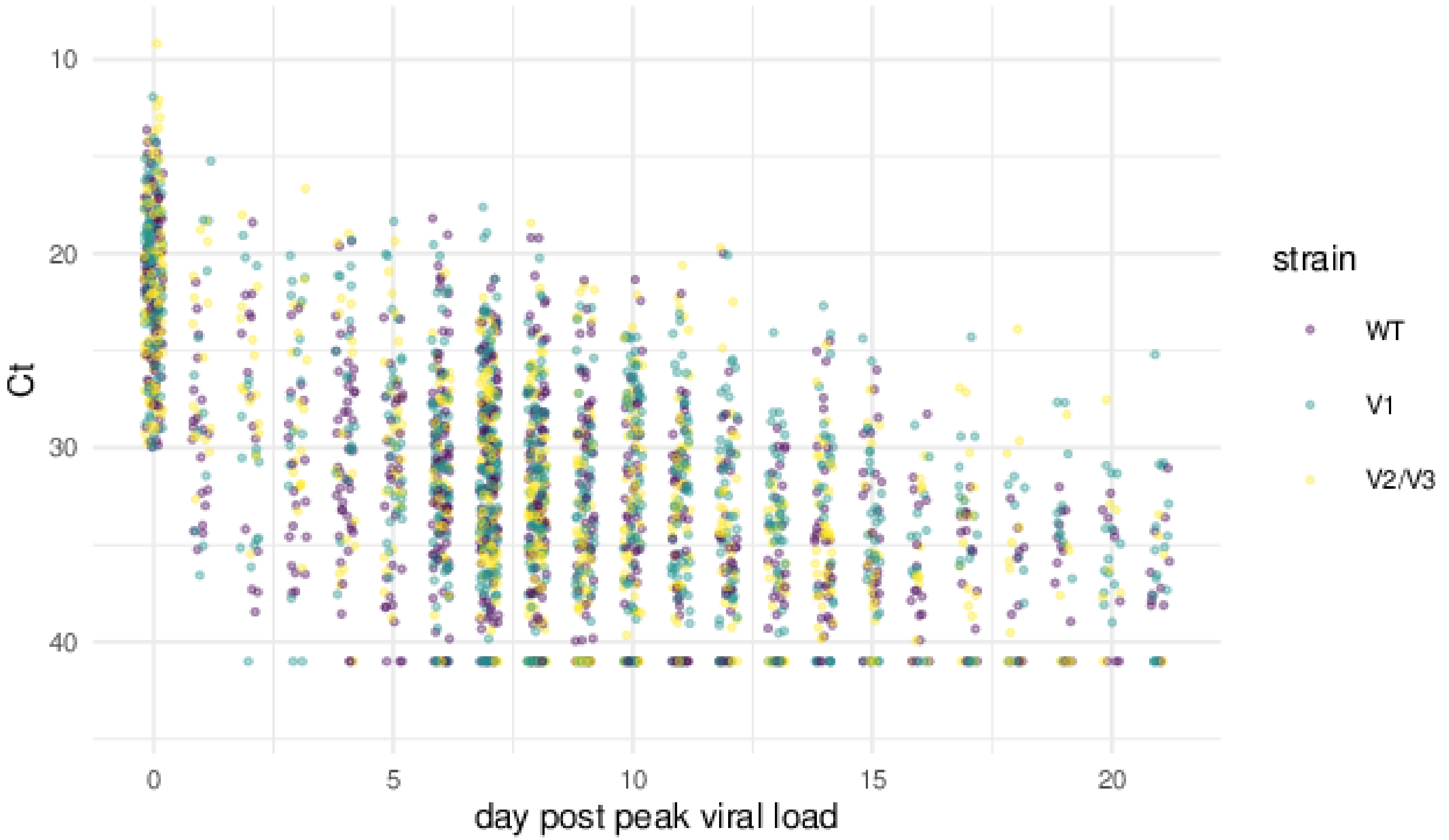
7. Variant transmission advantage

Elie, Lecorche, Sofonea, Trombert-Paolantoni,
Foulongne, Guedj, Haim-Boukobza, Roquebert, Alizon (2021, medRxiv)

Data analyzed (N=8,006)

Variable	Value	Median (IQR or n %)
sampling context	*general population (ref)*	7510 (94%)
	hospital	496 (6%)
age		42 [26-59] years
follow up duration		8 [6-12] days
date of first swab (in 2021)		Mar 7 [Feb 19-Mar 20]
lineage	*wildtype (ref)*	1643 (21%)
	B.1.1.7 (V1)	5874 (73%)
	B.1.351 or P.1 (V2/V3)	489 (6%)
samples per individual	*2*	7091 (88%)
	3	774 (10%)
	>3	141 (2%)
sampling region	*Ile-de-France (ref.)*	5124 (64%)
	other	2882 (36%)
infection phase observed	*viral load increase*	734 (9%)
	viral load decrease	7363 (92%)
	both	91 (1%)

SARS-CoV-2 kinetics



- Collaboration with CERBA and CHU de Montpellier
- >40,000 variant-specific RT-PCR performed between Feb 8 and Mar 20, 2021
- 8,006 individuals with longitudinal follow-up

Analysis from Baptiste Elie

Linear mixed models

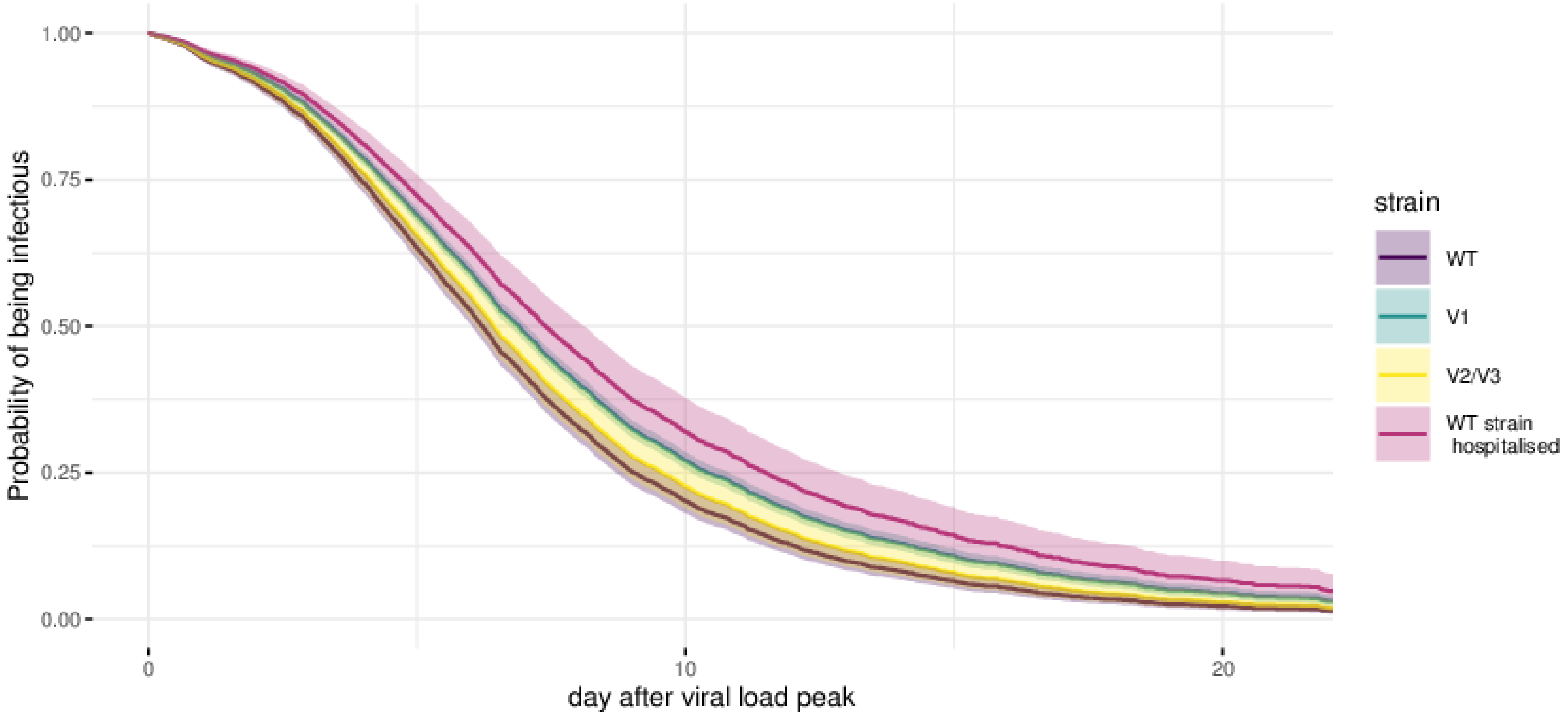
- We consider main effects (age, strain, day, date, region, hospital),
- interactions (strain:age, strain:hospital, day:strain, day:age, day:hospital),
- a random effect on the intercept (lmer function in the lme4 R package),
- and perform model selection on all the combinations (BIC criterion).

Analysis from Baptiste Elie, data from CERBA

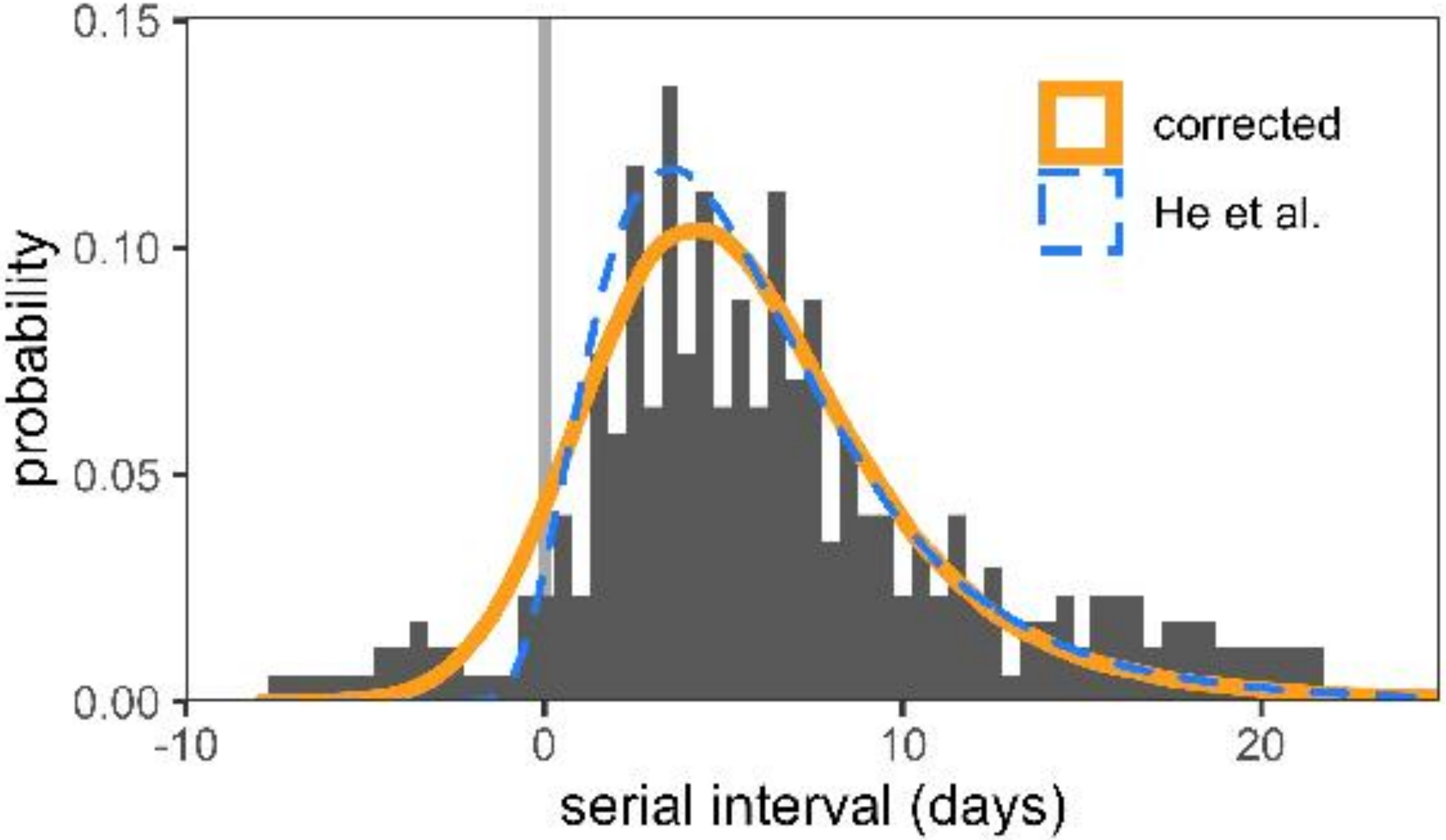
Best model

Predictor		Estimate	95% CI
Intercept		**23.4**	**(22.9,23.9)**
day		**1.19**	**(1.15,1.23)**
strain	*wildtype*	ref	—
	V1	-0.452	(-0.974,0.0702)
	V2/V3	**-1.27**	**(-2.21,-0.332)**
hospital	*no*	ref	—
	yes	**-1.01**	**(-1.46,-0.551)**
age		-0.00575	(-0.0143,0.00284)
region	*Ile-de-France*	ref	—
	Normandie	**-1.01**	**(-1.3,-0.713)**
date		**0.0066**	**(0.000851,0.0124)**
age:strain	***V1***	**-0.014**	**(-0.0234,-0.00456)**
	V2/V3	0.0123	(-0.00626,0.0308)
day:strain	***V1***	**-0.0481**	**(-0.0795,-0.0166)**
	V2/V3	0.0246	(-0.033,0.0823)
day:hospital		**-0.0725**	**(-0.125,-0.0201)**
day:age		**-0.00358**	**(-0.00415,-0.003)**

Survival analysis approach

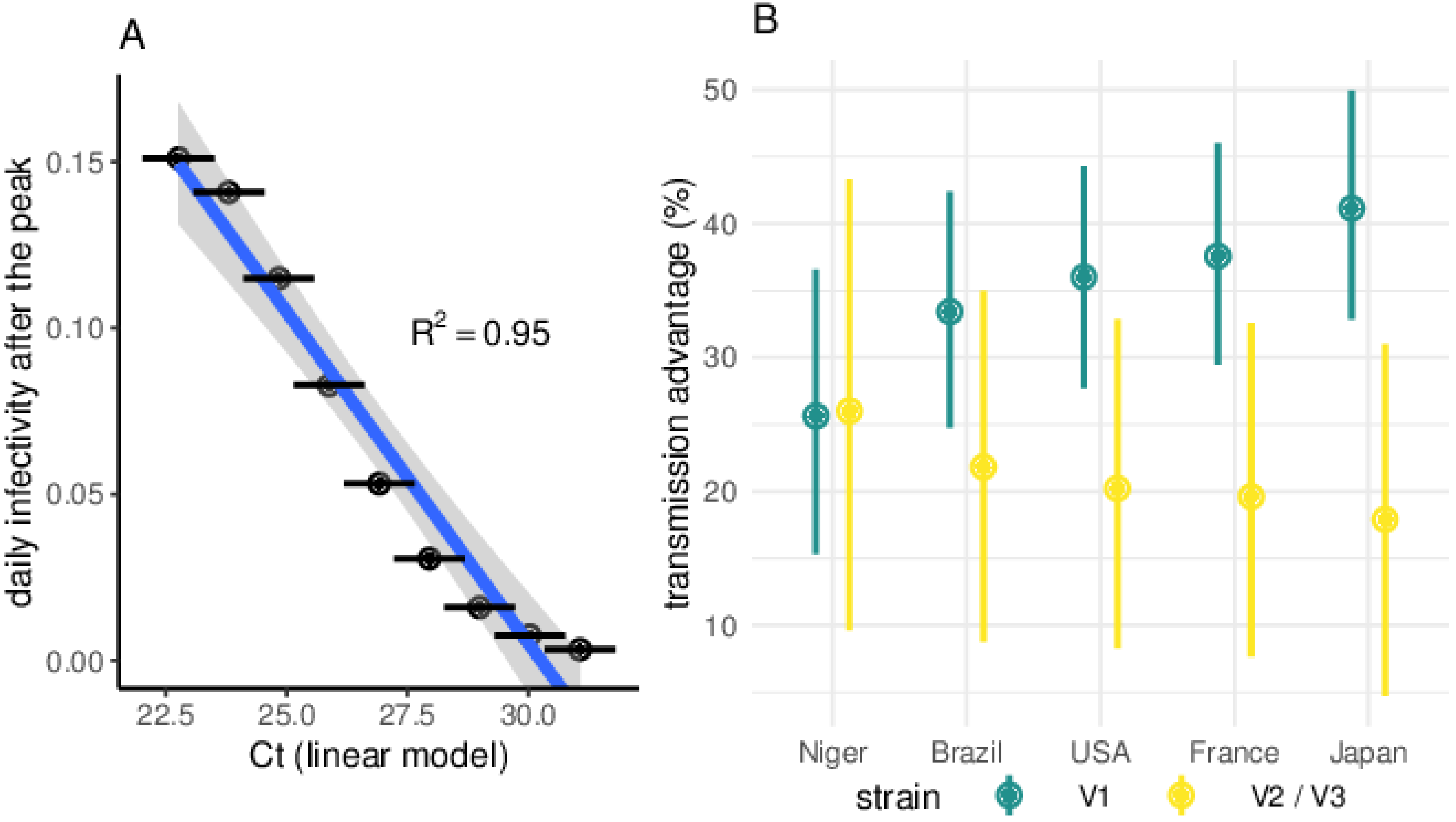


Epidemiological impact



Ashcroft *et alii* (2020, *Swiss Med Wkly*) correcting He *et alii* (2020, *Nat Med*)

Epidemiological impact



Variant-specific kinetics

- V1 and V2/V3 have a higher virus load peak than WT (with a correlation with age for V1)
- V1 causes longer infections than WT
- V1's transmission advantage could depend on population demography

Elie *et alii* (2021, *medRxiv*)

French COVID modelling

2. ICU modelling

3. Phylodynamics

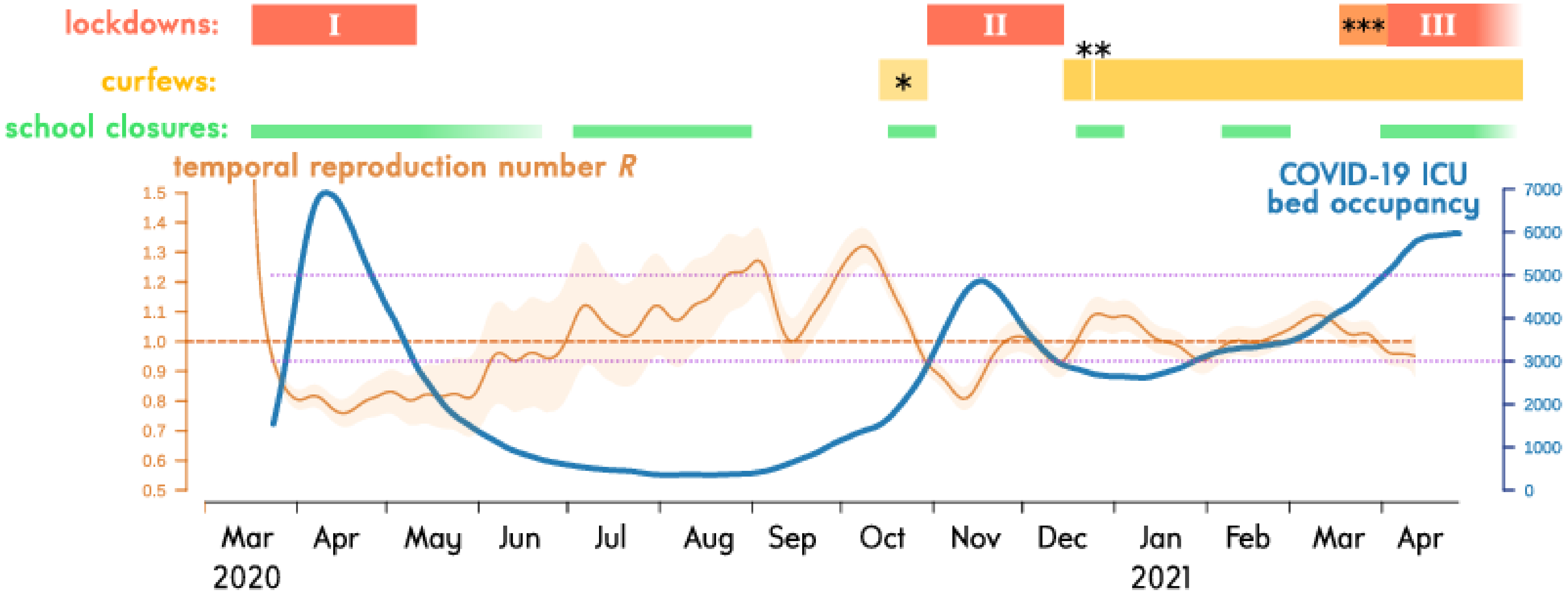
4. EPIDEMAP

5. Optimal age-based control

6. Within-host kinetics

7. Variant transmission advantage

Two waves and a high tide



Sofonea *et alii* (2021, *ACCPM*)

Transmission advantage

The selection advantage s of a variant in frequency p in the population is

$$s = \frac{\mathrm{d}}{\mathrm{d}t} \log\left(\frac{p(t)}{1 - p(t)}\right)$$

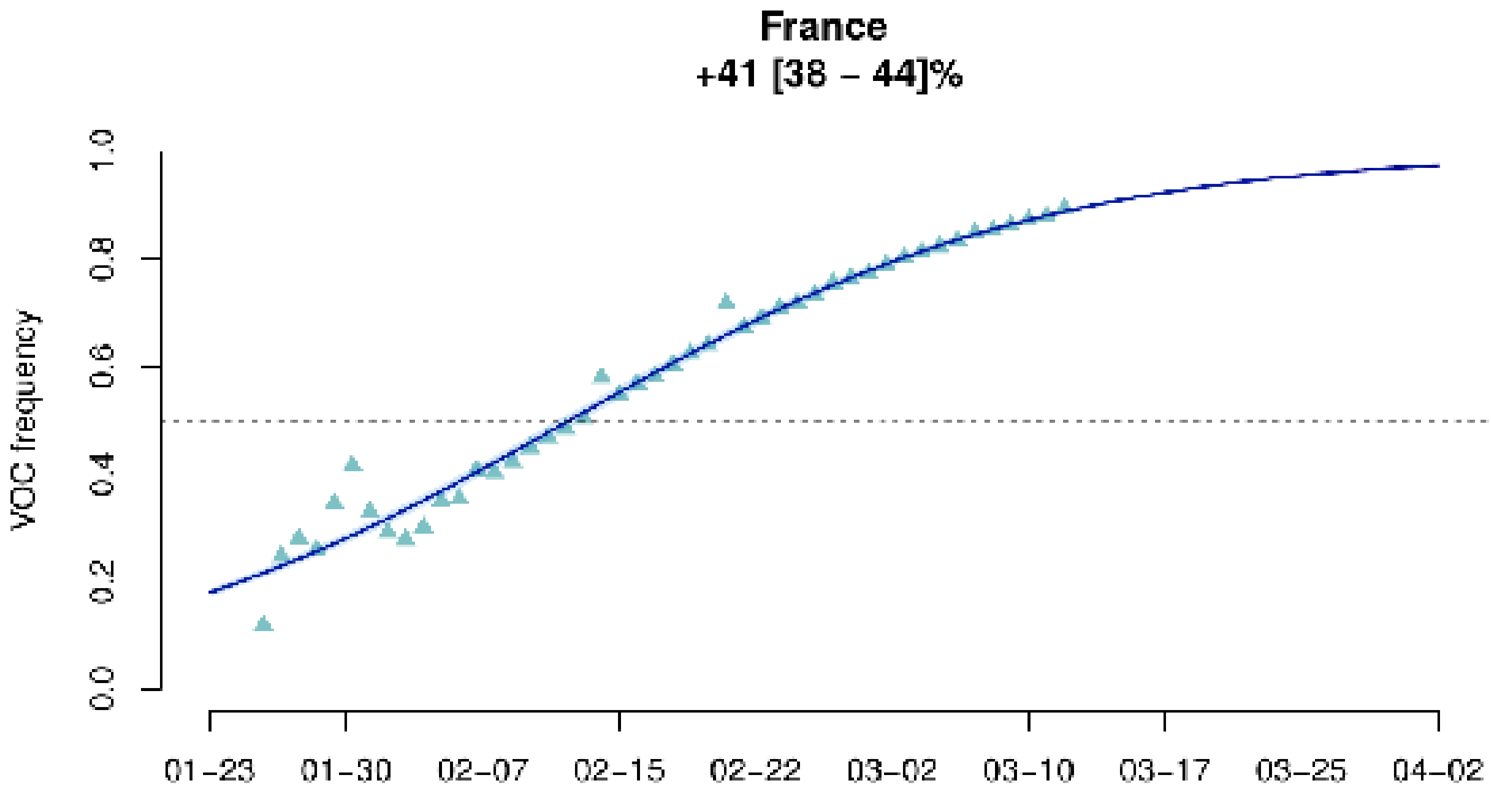
If $V(t)$ is the number of variants and $W(t)$ the number of wild types, then

$$\begin{aligned} s &= \frac{\mathrm{d}\ \log(V)}{\mathrm{d}t} - \frac{\mathrm{d}\ \log(W)}{\mathrm{d}t} \\ &= r_V - r_W \\ &= \Delta r \end{aligned}$$

Finally, if T is the generation time we get $s_T = s \times T$

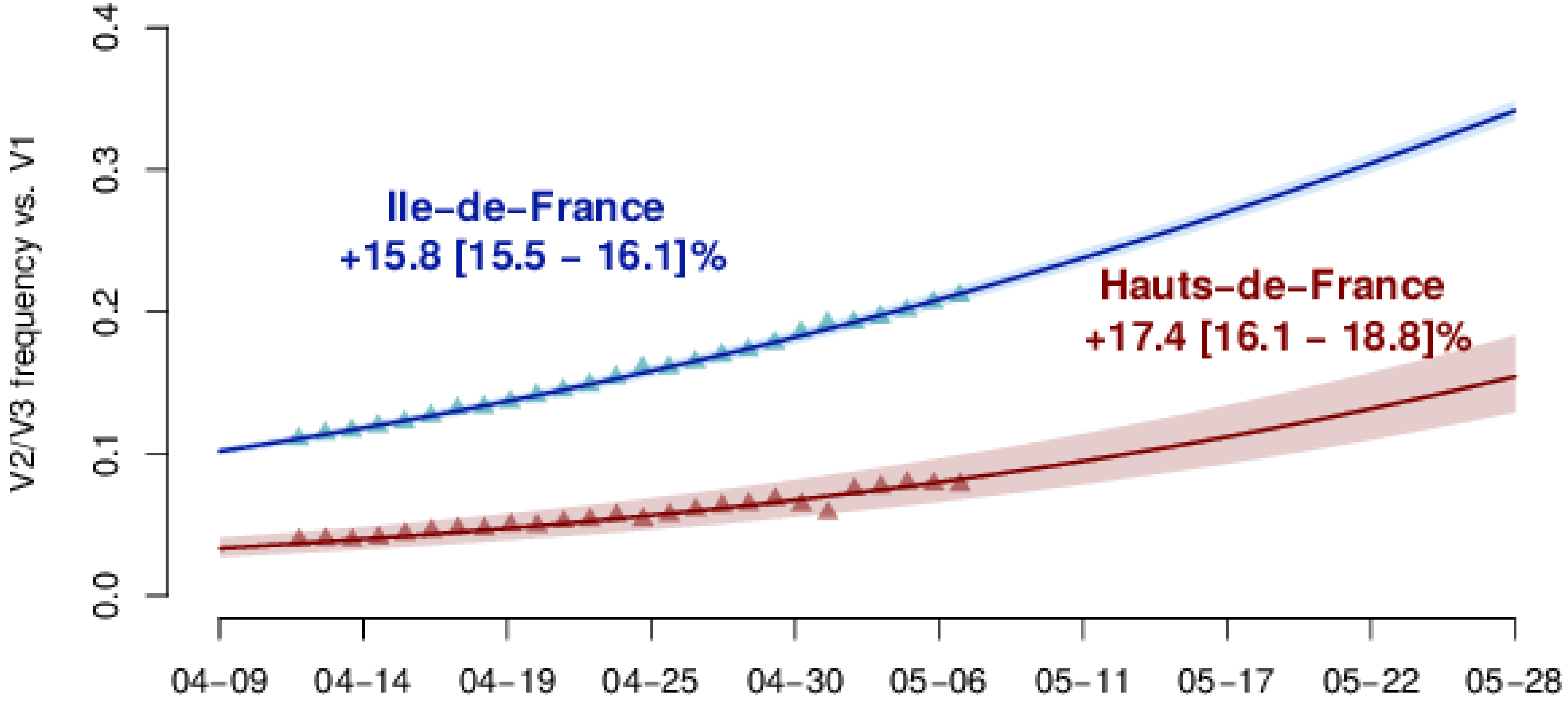
Chevin *et alii* (2011, *Biol. Lett.*)

Variant advantage



Haim-Boukobza *et alii* (2021, *Emerg. Infect. Dis.*)

Variant replacement



Roquebert *et alii* (2021, *Eurosurveillance*)