ESTIMATING THE DATE OF EMERGENCE OF AN EPIDEMIC FROM DETECTION DATA: APPLICATIONS TO COVID-19

Jijón S[1], Czuppon P[2], Blanquart F[3] and Débarre F[1]

[1]Institute of ecology and environmental sciences of Paris (iEES-Paris, UMR 7618), Sorbonne Université, CNRS, UPEC, IRD, INRAE
[2]Institute for Evolution and Biodiversity, University of Münster
[3]Center for Interdisciplinary Research in Biology, CNRS, Collège de France, PSL Research University

Background

Estimates of dates of emergence for COVID-19

Dating attempts for the first human SARS-CoV-2 infection [1–2], as well as the emergence of SARS-CoV-2 variants of concern [3].

Main estimates of the date of emergence.
Abbreviations: CrI = Credibility interval; Ref. = Reference.

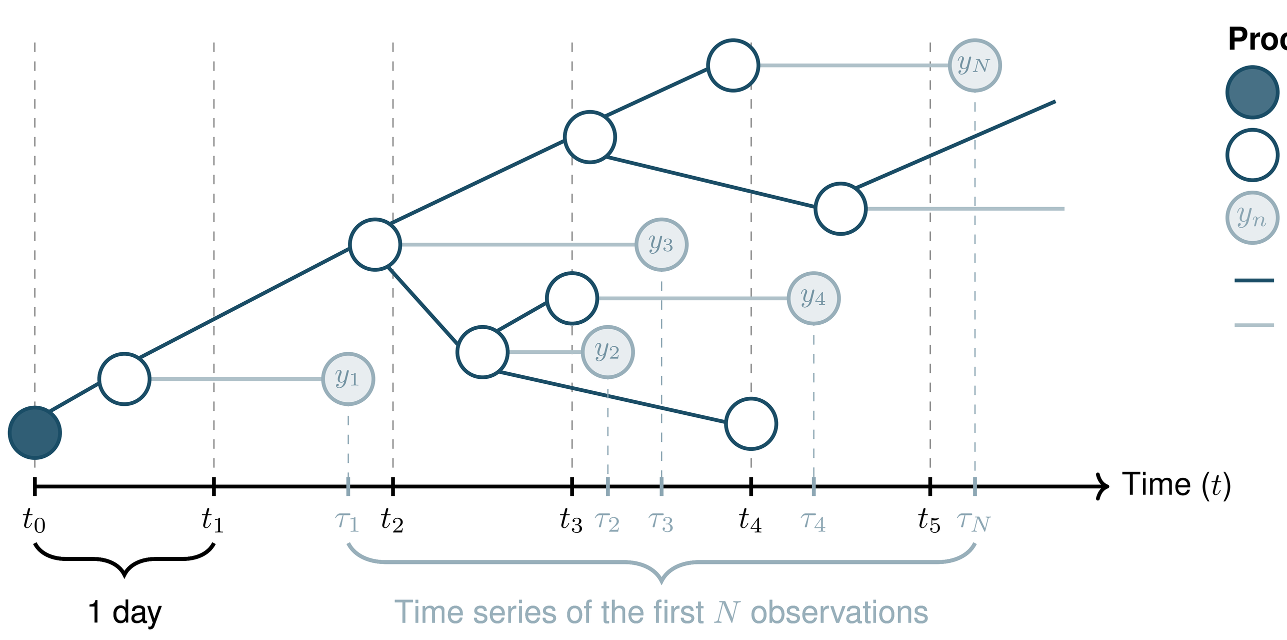
Epi context	Estimated date of emergence	Method	Ref.
COVID-19 in Wuhan	Mid-Oct to mid-Nov, 2019	Transmission model coupled with genomic data.	[1]
	Nov 18, 2019 (95%CrI: Oct 23 to Dec 8)	Estimates were recently updated.	[2]
B.1.1.7 ("Alpha") variant in the United Kingdom (UK)	Early August 2020	Stochastic model of early epidemics dynamics estimating the time of the first detection event.	[3]

Objective

To estimate the delay from emergence (i.e., first infection) to the N-th observed (i.e., detected) case, using available data.

Methods

Modelling infectious disease spread



We extended the model presented in [3], to model infectious disease spread from a single infectious individual (●) to N observed cases. The times and probabilities for pathogen transmission (○) and detection events (y_n, $n = 1, \ldots, N$) are drawn from the following distributions:

- Number of secondary cases $\sim NegBinom\left(\kappa, p = \frac{\kappa}{\kappa+R}\right)$, where $\kappa :=$ dispersion parameter and $R :=$ effective reproduction number.
- Number of detected infections $\sim Binom(n = I(t), p = p_{detect})$, where $I(t) :=$ Infected individuals at time t and $p_{detect} :=$ probability of detection.
- Times of infection (t_i) and detection (τ_n) $\sim Gamma(\kappa_x, \theta_x)$, with $x \in \{t, \tau\}$, respectively.

Model calibration

The numerical simulations (sim) are calibrated to reproduce

i) the observed (obs) time period between the first and the N-th case:

$$\left|(\tau_N^{\text{obs}} - \tau_1^{\text{obs}}) - (\tau_N^{\text{sim}} - \tau_1^{\text{sim}})\right| \leq \delta_{tol}\,(\tau_N^{\text{obs}} - \tau_1^{\text{obs}}),$$

and

ii) the observed daily number of cases:

$$|y_i^{\text{obs}} - y_i^{\text{sim}}| \leq \delta'_{tol}\,N, \qquad \forall i = 1, 2, \ldots, N$$

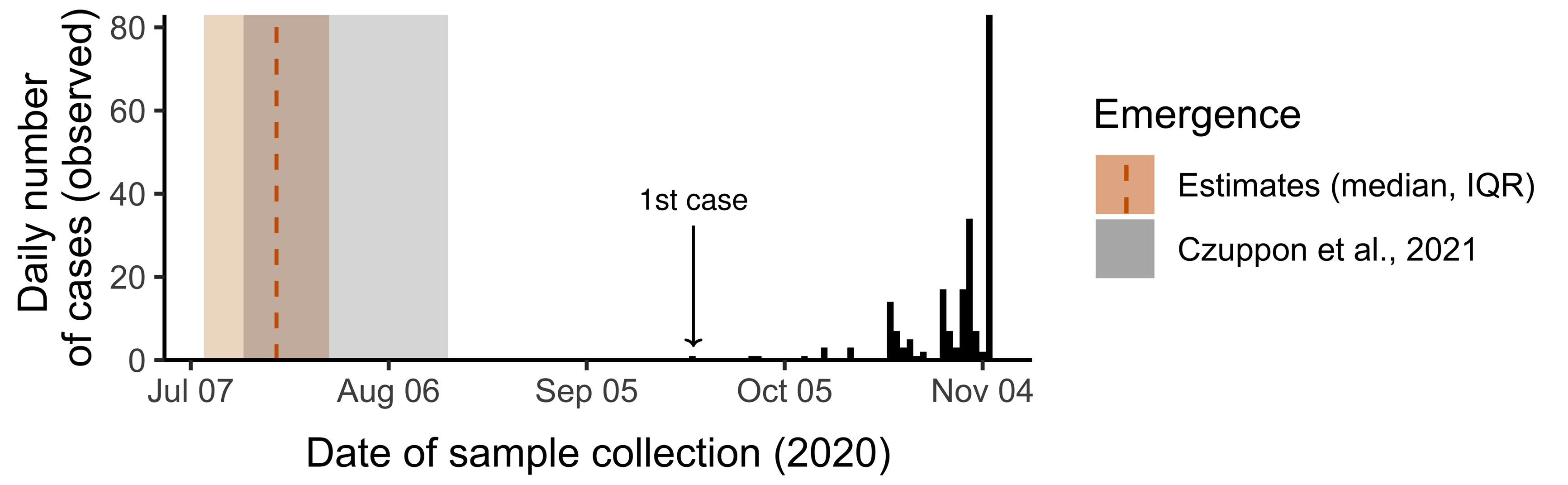
,

where δ_{tol} = 0.2 and δ'_{tol} = 0.9.

Findings

Estimating the date of the first infection with the "Alpha" variant in the UK

Data: N=406 detected cases sequenced by Nov 11, 2020 [4].



A-priori parameters. Input parameters as in [3].

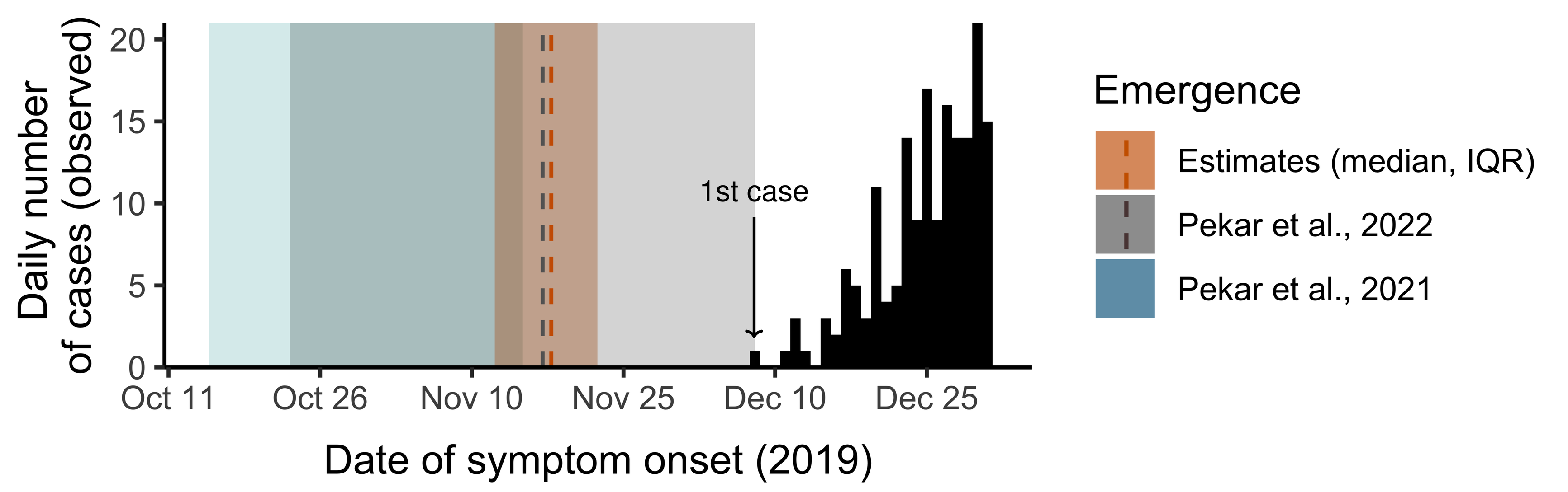
Param.	Value
R	1.50
κ	0.57
p_{detect}	0.01
κ_t	0.83
θ_t	6.60
κ_τ	0.58
θ_τ	12.0

Estimates of the emergence date and other epidemiological indicators. Median values (IQR), unless stated otherwise.

Indicator	Median (IQR)
Number of days from 1st infection to N-th case	114.9 (106.7–125.9)
Date of first Alpha infection	Jul 20 (09–28), 2020
Mean number of secondary infections (sd)	1.5 (2.3)
Epidemic size	1104 (433–2193)
Proportion of detected infections	62.3% (4.2%-81.0%)

Estimating the date of emergence of COVID-19 in Wuhan

Data: N=174 reported cases of COVID-19 with symptom onset up to Dec 31, 2019 [5].



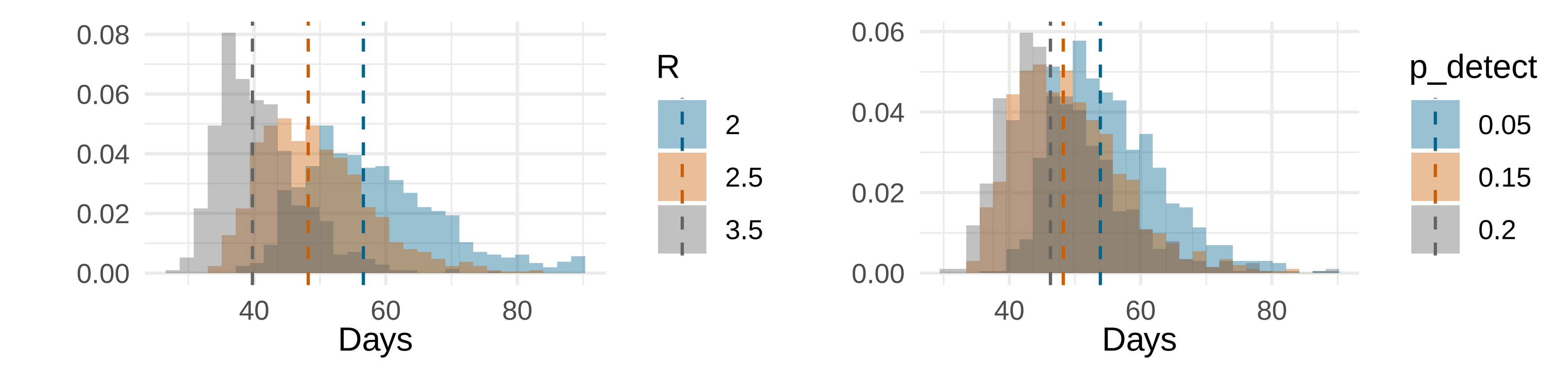
A-priori parameters.

Param.	Value	Ref.
R	2.50	[6]
κ	0.10	[7]
p_{detect}	0.15	[6]
κ_t	0.83	[3]
θ_t	6.60	[3]
κ_τ	1.04	[8]
θ_τ	6.25	[8]

Estimates of the emergence date and other epidemiological indicators.

Indicator	Median (IQR)
Number of days from 1st infection to N-th case	43.1 (38.6–48.7)
Date of first SARS-CoV-2 infection	Nov 11 (13–23), 2019
Mean number of secondary infections (sd)	2.6 (3.7)
Epidemic size	957 (567–1440)
Proportion of detected infections	18.6% (12.4%–31.4%)

Sensitivity analyses. Distributions of time from 1st infection to N-th reported case, varying the transmission (R) and detection (p_{detect}) parameters, without the calibration constraints. Median values are depicted by a dashed line.



Abbreviations: IQR = interquartile range, Param. = parameter, sd = standard deviation, Ref. = reference.

Conclusions

- We propose a generic and flexible modelling framework that can be applied to date epidemic outbreak emergence.
- Our results fall within the ranges previously estimated, by using different methods.

References

[1] Pekar JE et al. (2021) *Science*
[2] Pekar JE et al. (2022) *Science*
[3] Czuppon P et al. (2021) *J.R.Soc.Interface*
[4] Rambaut A et al. (2022) virological.org
[5] WHO-convened global study of origins of SARS-CoV-2: China Part (2021)
[6] Hao et al. (2020) *Nature*
[7] Endo et al. (2020) *Wellcome Open Res.*
[8] Backer et al (2020) *Eurosurveillance*

Contact information:
Sofía JIJON ALBAN
Mathematical epidemiology
sofia.jijon_alban@sorbonne-universite.fr
sjijon.github.io