

Appendix: Estimating dates of origin and end of COVID-19 epidemics: the importance of stochastic dynamics and memory effects

January 19, 2021

Contents

S1 Models parameters	2
S2 Markovian SEAIRHD model	3
S3 DS model sensitivity analysis	5
S3.1 Effect of the initial number of imported cases	6
S4 Contact rate variation during and post lock-down period	8
S5 Eradication and rebound risk with superspreading events	8
S6 Eradication and lock-down initiation date	10

S1 Models parameters

Table S1 – Models parameters and constraints.

Notation	Signification	Model	Constraint/value	Reference
\mathcal{R}_0	Basic reproduction number	all	3.02	(1)
p	Infection fatality ratio	all	0.874 %	(1)
$(\omega_i)_{i \in \mathbb{N}}$	Serial interval distribution	stochastic	discretised <i>Weibull</i> (mean: 4.8 d., sd: 2.3 d.)	(2)
θ	Infection to death delay	stochastic	<i>LogNormal</i> (mean: 23.3 d., sd: 9.7 d.)	(1)
k	Shape parameter (Gamma distribution)	stochastic	0.16	(3)
t_0	Time of epidemic wave initiation	SEAIRHD	2020/01/22	inferred
β	<i>Per capita</i> infectious rate	SEAIRHD	$\frac{\mathcal{R}_0 \gamma \sigma}{S_0 (\gamma + \sigma)}$	calculated
η_{lockdown}	Lockdown effect	SEAIRHD	0.232	inferred
ϵ	Rate of end of latency	SEAIRHD	0.415	inferred
σ	Rate of symptoms onset	SEAIRHD	$\frac{1}{5.6 - 1/\epsilon}$	(4)
γ	End of infectivity rate	SEAIRHD	0.653	inferred
α	Death rate	SEAIRHD	$\frac{1}{17.7 - 1/\gamma}$	(1)
\mathcal{N}	total population	SEAIRHD	64171900	INSEE

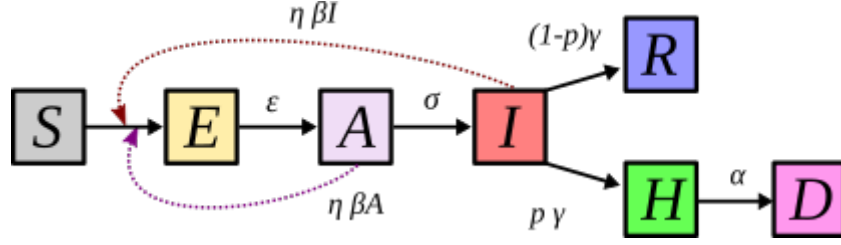


Figure S1 – Parameter notations are shown in the main text.

S2 Markovian SEAIRHD model

The model is captured by the following set of differential equations:

$$\frac{dS}{dt} = -\eta_n \beta (A + I) S \quad (\text{S1a})$$

$$\frac{dE}{dt} = \eta_n \beta (A + I) S - \epsilon E \quad (\text{S1b})$$

$$\frac{dA}{dt} = \epsilon E - \sigma A \quad (\text{S1c})$$

$$\frac{dI}{dt} = \sigma A - \gamma I \quad (\text{S1d})$$

$$\frac{dR}{dt} = (1 - p) \gamma I \quad (\text{S1e})$$

$$\frac{dH}{dt} = p \gamma I - \alpha H \quad (\text{S1f})$$

$$\frac{dD}{dt} = \alpha H \quad (\text{S1g})$$

where η_n measures the public health intervention impacts on the disease spread at day n , β is the per capita transmission rate of asymptomatic and symptomatic hosts, ϵ is the rate of end of latency, σ is the rate of symptoms onset, γ is the recovery rate, p is the infection fatality ratio (IFR), and α is the rate at which hospitalised patients die. Our goal is to capture the key features of the infection life-history, especially the incubation period, the asymptomatic transmission, and the delay to hospitalized deaths, but not to fit the epidemic in details.

Deterministic implementation: The set of ODE shown in the previous paragraph is solved using 'odeint' function from Numpy on Python 3.8.3. We then applied a moving average, with a window of 7 days, as done with the real data. We estimated the following parameters for the SEAIRHD model using a maximum likelihood procedure: t_0 , γ , σ and $\eta_{lockdown}$. We computed the likelihood of our model using the data of daily hospital mortality in France from January 1st to May 11 (end of national lockdown), on which a moving average of 7 days is applied to avoid "week-end effects".

We compared this model to the discrete time non-markovian model, and a SEAIRH4D model

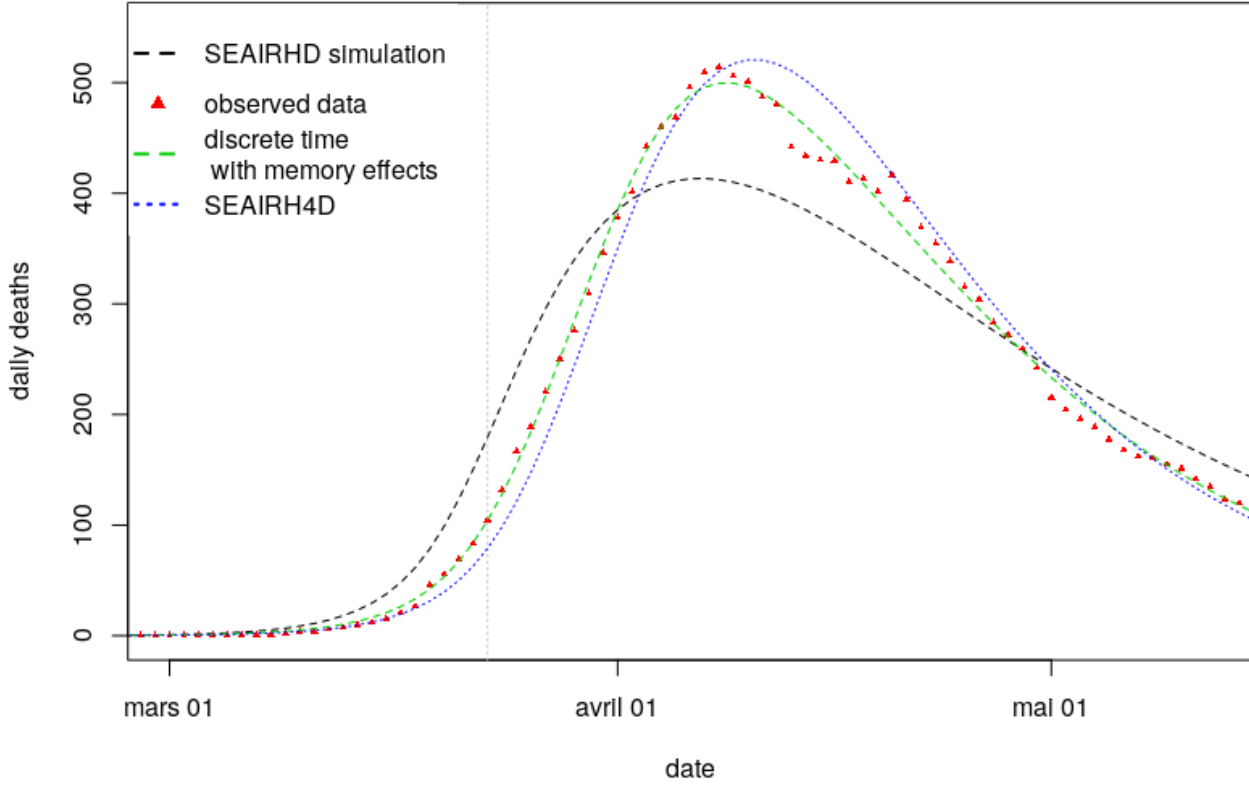


Figure S2 – Best fitted SEAIRDH model number of daily deaths compared to the actual death data in France (AIC: 1324). The SEAIRH4D is the case where memory is introduced on the hospitalisation duration, following an Erlang with shape parameter 4 and the same mean (AIC: 730). The discrete time model with memory effects provides a better fit (AIC: 461).

in which memory in the delay from hospitalization to death is implemented (Fig. S2).

Stochastic implementation: Using the same parameters, we simulated 1,000 times a stochastic version of this model, using a Gillespie algorithm with the package *TiPS* (5) on R v.3.6.3 (6). We stored the time to 100 daily deaths, after applying a moving average with a window of 7 days on the simulations.

S3 DS model sensitivity analysis

Effect of the serial interval distribution

The time elapsed between infection events in an infector-infectee pair is called the generation time. It is a key epidemiological parameter in our model, though almost impossible to measure directly. However, the serial interval distribution - i.e. the time between symptom onsets in an infector-infectee pair- has theoretically the same expectation, but a higher variance.

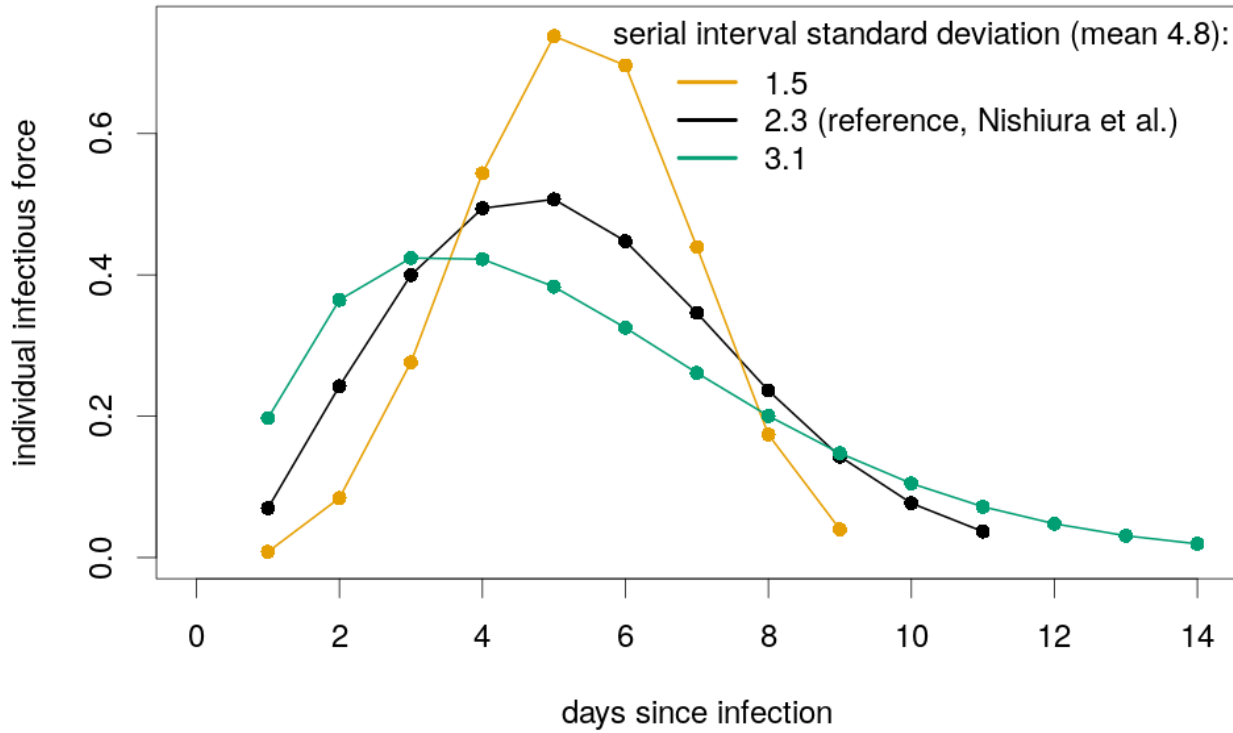


Figure S3 – Serial interval distributions used hereafter to analyze the sensitivity to the serial interval standard deviation. In black is shown the reference from (7). The mean is kept constant, but the standard deviation is increased or decreased.

To our knowledge, the only available data to estimate this parameter come from Asia (7, 8). This estimate could change, e.g. through behavioral on contact structure between the countries where the data come from and France. Intuitively, the epidemic growth is very sensitive to the generation time expectation, hence the epidemic starting date would be shifted significantly. Here, we focus on the generation time variance to see to what extent the generation time distribution can affect the epidemiological dynamic (fig. S3).

We observe that the higher the standard deviation, the later the starting date is inferred (Fig. S4). The median value is shifted from January 11th to 19th when the standard deviation is doubled, under the homogeneous infectivity model. However, this variation remains within the interval containing 95 % of the variation, using the serial interval determined by (7).

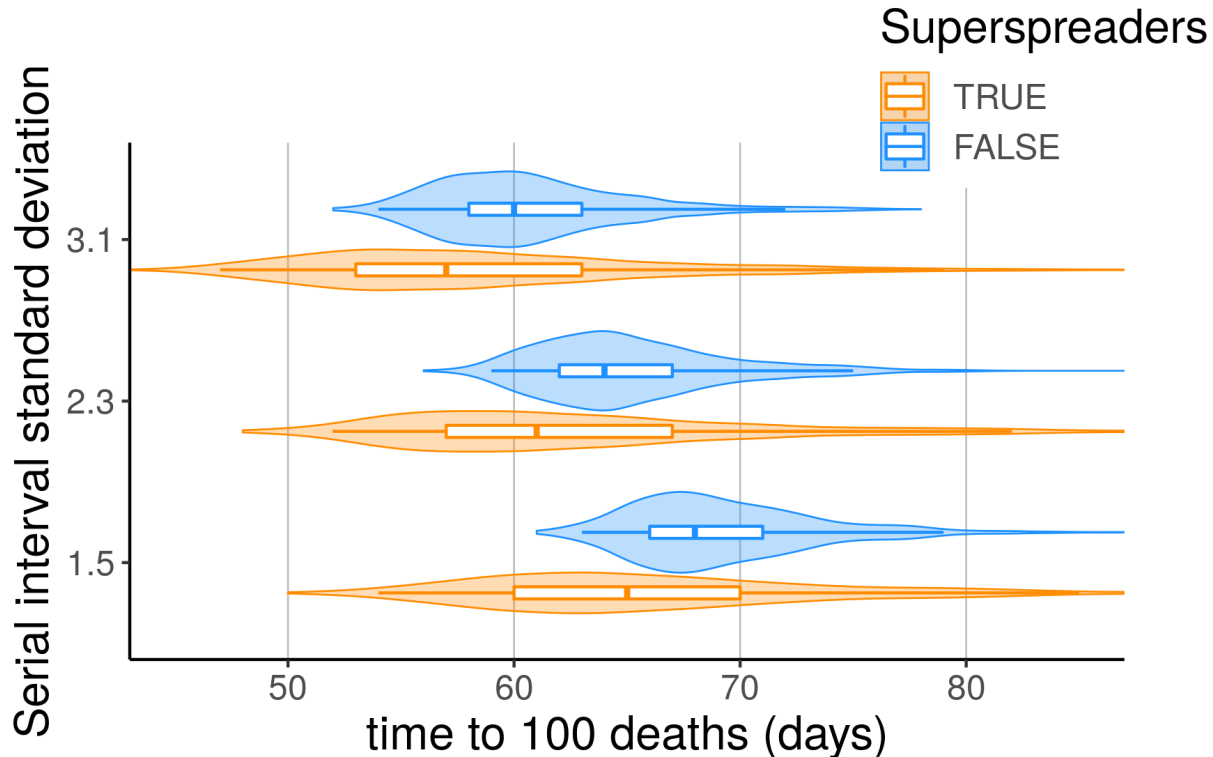


Figure S4 – Generation time standard deviation impact on the starting date inference. The mean generation time is kept constant, following (7). The boxplots show the following quantiles: 2.5%, 25%, 50 %, 75%, 97.5 %.

S3.1 Effect of the initial number of imported cases

We made the assumption that only one imported case was responsible for the whole outbreak wave. To note, this is not incompatible with the occurrence of earlier cases, such as the one that occurred on December 27th in France, because in case of individual R_0 heterogeneity, most of those early cases could have died out without contributing to the main outbreak wave.

However, it can be argued that several imported cases may have contributed to the outbreak wave, and consequently may have accelerated the dynamic. This would imply that the outbreak would have started later.

The total number and intensity of imported cases responsible for the outbreak wave in France affects the starting date of the wave in the order of 4 to 9 days (fig. S5), which is similar to the

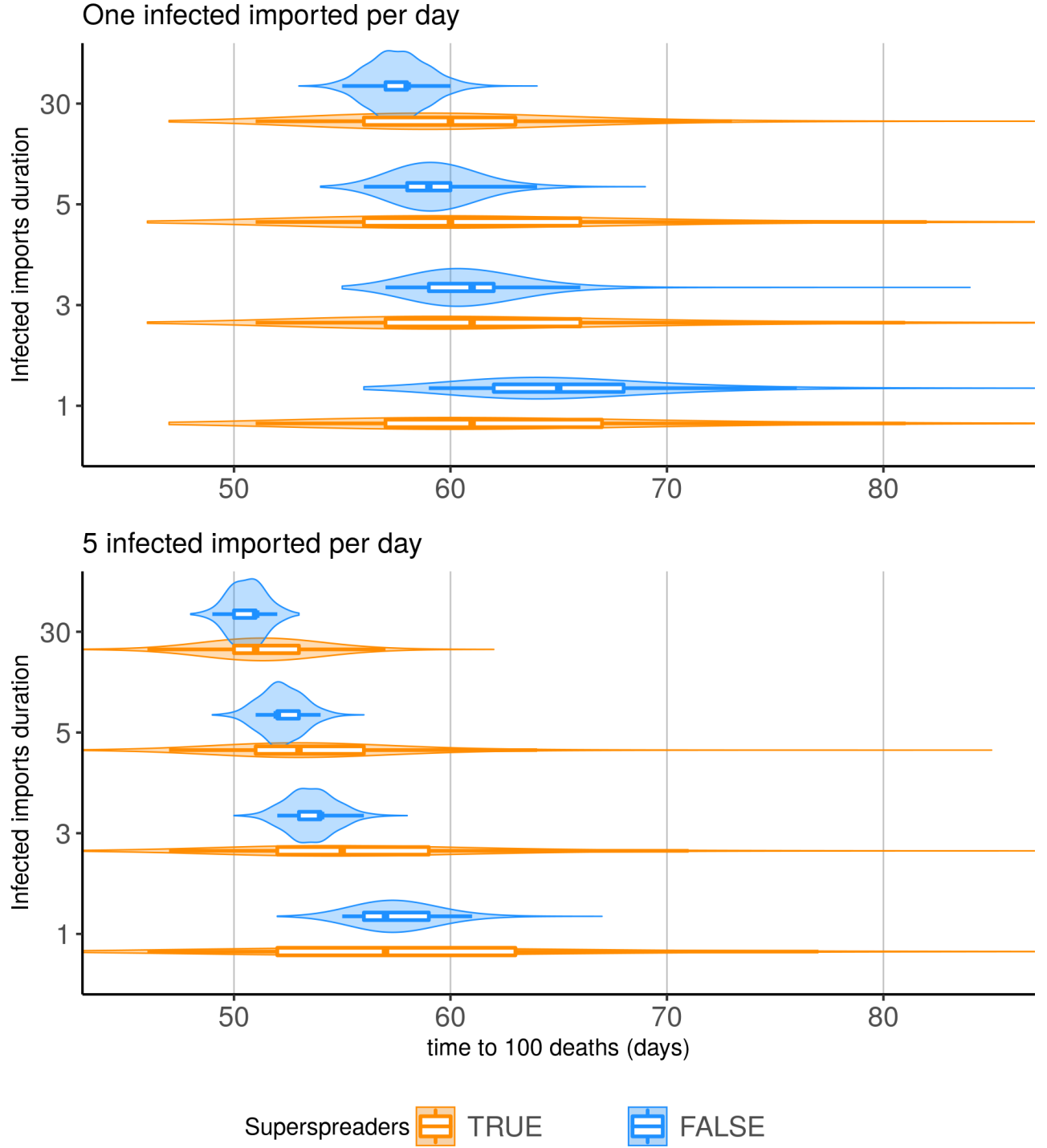


Figure S5 – Duration and intensity of the infectious individuals importation impacts on the starting date inference. The boxplots show the following quantiles: 2.5%, 25%, 50 %, 75%, 97.5 %.

sensitivity to the individual infectivity heterogeneity. We can see that only the importation of new infected individuals during the first days has an impact on the epidemic.

S4 Contact rate variation during and post lock-down period

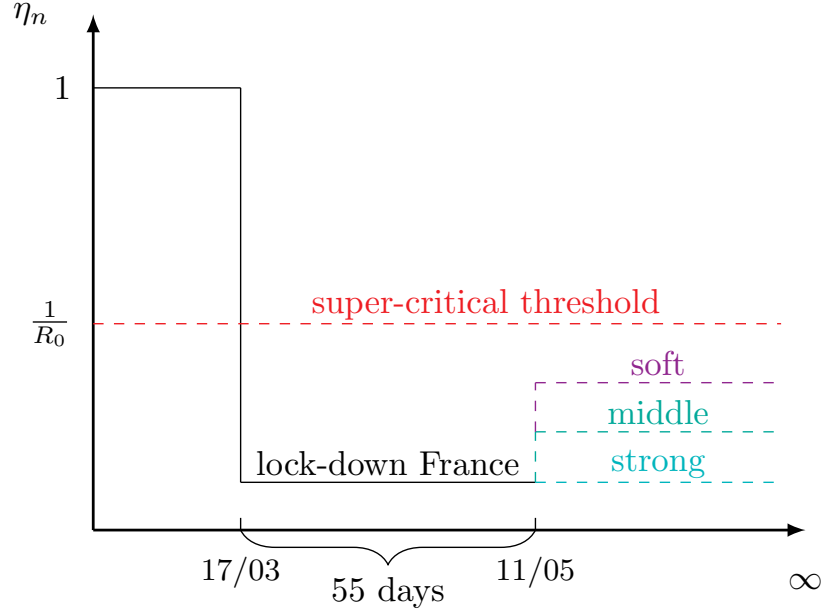


Figure S6 – **Variation of η_t for the estimation of τ :** On Mar 17 a lock-down was instated in France. We estimated its efficacy to be around $1-\eta_t = 0.76$ (1). The lock-down officially ended on May 11, for a total of 55 days under full lock-down. We evaluated the time to reach extinction under various extension constraints by assuming an infinite lock-down prolongation with fixed intensity, potentially different from the value estimated during the Mar 17 and May 11. The intensity of the lock-down/extension is inversely proportional to η_t .

S5 Eradication and rebound risk with superspreading events

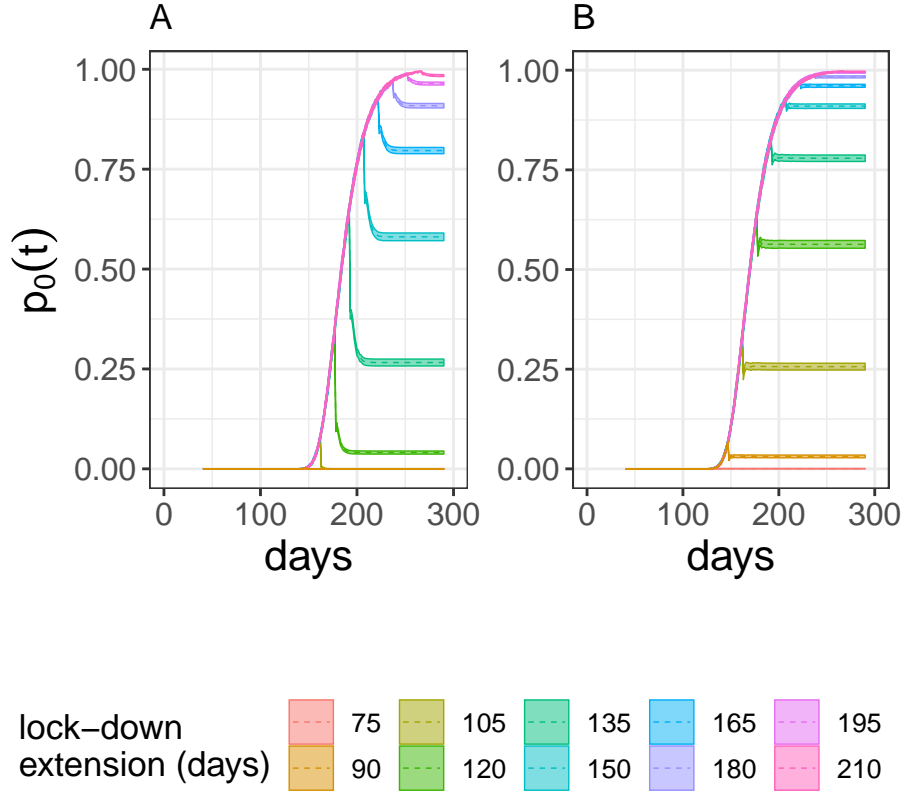


Figure S7 – **Estimated variation of $p_0(t)$ with finite lock-down extension post 11 May:** Here the x-axis corresponds to the number of days since the start of the lock-down (Mar 17) and the y-axis to estimation of $p_0(t)$. For the first 55 days $\eta_t = 0.24$, during the extension of the lock-down we set $\eta_t = 0.25$, and for the rest of the simulation we fixed $\eta_t = 1$. On the left panel (panel A) we displayed the case without transmission heterogeneity and on the right panel (panel B) with superspreaders. We plotted the estimation of $p_0(t)$ as dashed lines and the confidence interval as solid lines.

S6 Eradication and lock-down initiation date

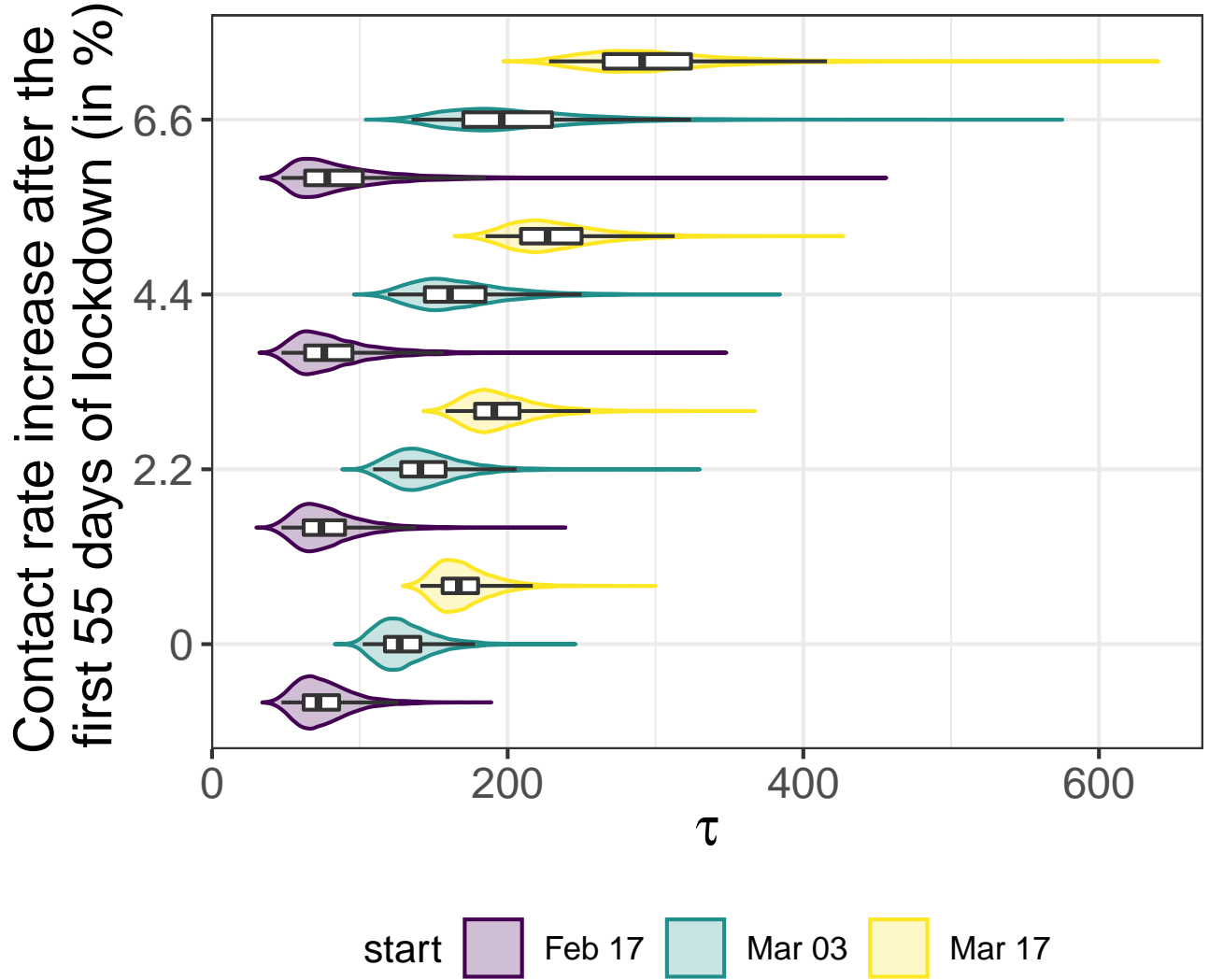


Figure S8 – **Effect of the lock-down intensity, stochasticity, and initiation date on the time to extinction (τ) under individual spreading heterogeneity assumption.** The distributions of τ (number of days since the start of the lock-down on Mar 17) for several lock-down intensities increase after the first 55 days are plotted on the Y-axis (ζ_t) using violin plots and boxplots. In this graph we assume individual spreading heterogeneity. The colors indicates the different initiation date of the lock-down: in purple it starts on Feb 17, green Mar 03 and yellow on Mar 17 (official start). The box extends from the lower to upper quartiles of the data. The whiskers expand from the 2.5% to the 97.5% quantiles.

References

- [1] Sofonea MT, Reyné B, Elie B, Djidjou-Demasse R, Selinger C, Michalakis Y, et al. Epidemiological monitoring and control perspectives: application of a parsimonious modelling framework to the COVID-19 dynamics in France. medRxiv. 2020;p. 2020.05.22.20110593.
- [2] Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. International Journal of Infectious Diseases. 2020 Apr;93:284–286. Available from: <https://doi.org/10.1016/j.ijid.2020.02.060>.
- [3] Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. Nature. 2005;438(7066):355–9.
- [4] Linton NM, Kobayashi T, Yang Y, Hayashi K, Akhmetzhanov AR, Jung Sm, et al. Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data. Journal of Clinical Medicine. 2020 Feb;9(2):538. Available from: <https://www.mdpi.com/2077-0383/9/2/538>.
- [5] Danesh G, Saulnier E, Gascuel O, Choisy M, Alizon S. Simulating trajectories and phylogenies from population dynamics models with TiPS [preprint]; 2020.
- [6] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2020. Available from: <https://www.R-project.org/>.
- [7] Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. International Journal of Infectious Diseases. 2020 Mar;0(0). Available from: [https://www.ijidonline.com/article/S1201-9712\(20\)30119-3/abstract](https://www.ijidonline.com/article/S1201-9712(20)30119-3/abstract).
- [8] Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. New England Journal of Medicine. 2020 Jan;Available from: <https://www.nejm.org/doi/10.1056/NEJMoa2001316>.