

Supplementary Information:
The source of individual heterogeneity
shapes infectious disease outbreaks

Baptiste Elie*, Christian Selinger, Samuel Alizon

February 4, 2022

Parameters estimation for known outbreaks

A limitation of the estimations of CV_I and CV_B from real epidemics is that we could only find joint distributions of secondary cases and serial interval in a measles outbreak. For the others, we used independent distributions for the other outbreaks. However, although this assumption does increase the confidence interval, we do not expect it to bias our results because any potential correlation is expected to be minimal. For instance, in the measles outbreak we analysed, the Spearman correlation coefficient between serial interval and secondary cases generated by the infector was 0.20.

Moreover, assuming constant infectiousness during the infectious period and absence of infection during the latent period, we infer the mean infection duration as $\bar{\Gamma} = 2(\bar{s} - \bar{e})$, where $\bar{s}$ is the mean serial interval and $\bar{e}$ is the mean latent period.

Outbreak	Secondary cases	Serial interval	R_0	Latent period	CV_B (95% CI)	CV_T (95% CI)	mean Γ (days)	Note
SARS Singapore	data from 57 cases [9]	Data from 180 cases [8]	1.63 [9]	Gamma (mean 5.2d, sd 2.5d) [7]	5.93 (3.08 - 9.82)	0.36 (0.03 - 0.89)		Serial interval and incubation period distribution cover the whole outbreak / secondary cases represent only the early phase before public health interventions.
Measles Hagelloch	Data from 66 cases	[6]	2.2 (from secondary cases data)	Gamma (mean 8.58d, sd 1.33 d) [1]	4.4 (1.90 - 8.3)	0.69 (0.12 - 1.71)	3.64	Outbreak in a German village in 1865. Data were restricted to the first third of the transmission events. Incubation period estimated from an independent outbreak (264 families from Providence, Rhode Island).
Ebola Guinea	Data from 152 individuals [2]	Data from 192 individuals [13]	1.71 [13]	Data from 155 individuals [13]	4.10 (1.83-7.26)	1.04 (0.45 - 2.00)	10.1	Transmission chains come from Guinea, whereas the other data come from the same outbreak but in the whole region of West Africa. The whole outbreak is considered: heterogeneity in infection duration might be affected by public health measures.
Smallpox in Europe	32 independent importations - only the first indigenous generation [3]	5 outbreaks in England and India [11] (n=223)	3.19 (from secondary cases data)	131 cases [10]	3.3 (1.45 - 6.62)	0.71 (0.52-0.96)	6.93	The dataset for latent period, secondary cases, and serial intervals come from different outbreaks.
Pneumonic Plague	Data from 74 cases (6 outbreaks)	[4]	Inferred: 1.32 [4]		0.67	0.48	2.5	In this case, we directly used the infection period duration coefficient of variation estimated in the study and retrieved CV_B from the secondary cases heterogeneity. No confidence interval is available with this computation.
Monkeypox	147 cases in Zaire 1980-1984 [5]	62 cases in Zaire 1980-1984 [5]	0.32 (from secondary cases data)	28 cases, 2013 outbreak in the Dem. Rep. of Congo [12]	1.83 (0.72-3.80)	12.7	0.12 (0.02-0.32)	The data for incubation period and secondary cases come from the same geographic region, but 30 year apart.

Table S1: Parameters, data sources for each studied outbreak.

References

- [1] Bailey, N. T. J., 1956 On Estimating the Latent and Infectious Periods of Measles: I. Families with Two Susceptibles Only. *Biometrika* **43**, 15. (doi: 10.2307/2333574).
- [2] Faye, O., Boëlle, P.-Y., Heleze, E., Faye, O., Loucoubar, C., Magassouba, N., Soropogui, B., Keita, S., Gakou, T., Bah, E. H. I., Koivogui, L., Sall, A. A. & Cauchemez, S., 2015 Chains of transmission and control of Ebola virus disease in Conakry, Guinea, in 2014: An observational study. *The Lancet Infectious Diseases* **15**, 320–326. (doi: 10.1016/S1473-3099(14)71075-8).
- [3] Fenner, F., Henderson, D. A., Arita, I., Jezek, Z., Ladnyi, I. D. & Organization, W. H., 1988 *Smallpox and Its Eradication*. World Health Organization.
- [4] Gani, R. & Leach, S., 2004 Epidemiologic Determinants for Modeling Pneumonic Plague Outbreaks. *Emerging Infectious Diseases* **10**, 608–614. (doi: 10.3201/eid1004.030509).
- [5] Jezek, Z., Grab, B. & Dixon, H., 1987 Stochastic model for interhuman spread of Monkeypox. *American Journal of Epidemiology* **126**, 1082–1092. (doi: 10.1093/oxfordjournals.aje.a114747).
- [6] Jombart, T., Frost, S., Nouvellet, P., Campbell, F. & Sudre, B., 2020 Outbreaks: A Collection of Disease Outbreak Data.
- [7] Kuk, A. Y. C. & Ma, S., 2005 The estimation of SARS incubation distribution from serial interval data using a convolution likelihood. *Statistics in Medicine* **24**, 2525–2537. (doi: 10.1002/sim.2123).
- [8] Lipsitch, M., Cohen, T., Cooper, B., Robins, J. M., Ma, S., James, L., Gopalakrishna, G., Chew, S. K., Tan, C. C., Samore, M. H., Fisman, D. & Murray, M., 2003 Transmission Dynamics and Control of Severe Acute Respiratory Syndrome. *Science* **300**, 6.
- [9] Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E. & Getz, W. M., 2005 Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359. (doi: 10.1038/nature04153).
- [10] Nishiura, H., 2009 Determination of the appropriate quarantine period following smallpox exposure: An objective approach using the incubation period distribution. *International Journal of Hygiene and Environmental Health* **212**, 97–104. (doi: 10.1016/j.ijheh.2007.10.003).
- [11] Nishiura, H. & Eichner, M., 2007 Infectiousness of smallpox relative to disease age: Estimates based on transmission network and incubation period. *Epidemiology and Infection* **135**, 1145–1150. (doi: 10.1017/S0950268806007618).

- [12] Nolen, L. D., Osadebe, L., Katomba, J., Likofata, J., Mukadi, D., Monroe, B., Doty, J., Hughes, C. M., Kabamba, J., Malekani, J., Bomponda, P. L., Lokota, J. I., Balilo, M. P., Likafi, T., Lushima, R. S., Ilunga, B. K., Nkawa, F., Pukuta, E., Karhemere, S., Tamfum, J.-J. M., Nguete, B., Wemakoy, E. O., McCollum, A. M. & Reynolds, M. G., 2016 Extended Human-to-Human Transmission during a Monkeypox Outbreak in the Democratic Republic of the Congo. *Emerging Infectious Diseases* **22**, 1014–1021. (doi: 10.3201/eid2206.150579).
- [13] WHO Ebola Response Team, 2014 Ebola Virus Disease in West Africa — The First 9 Months of the Epidemic and Forward Projections. *New England Journal of Medicine* **371**, 1481–1495. (doi: 10.1056/NEJMoa1411100).

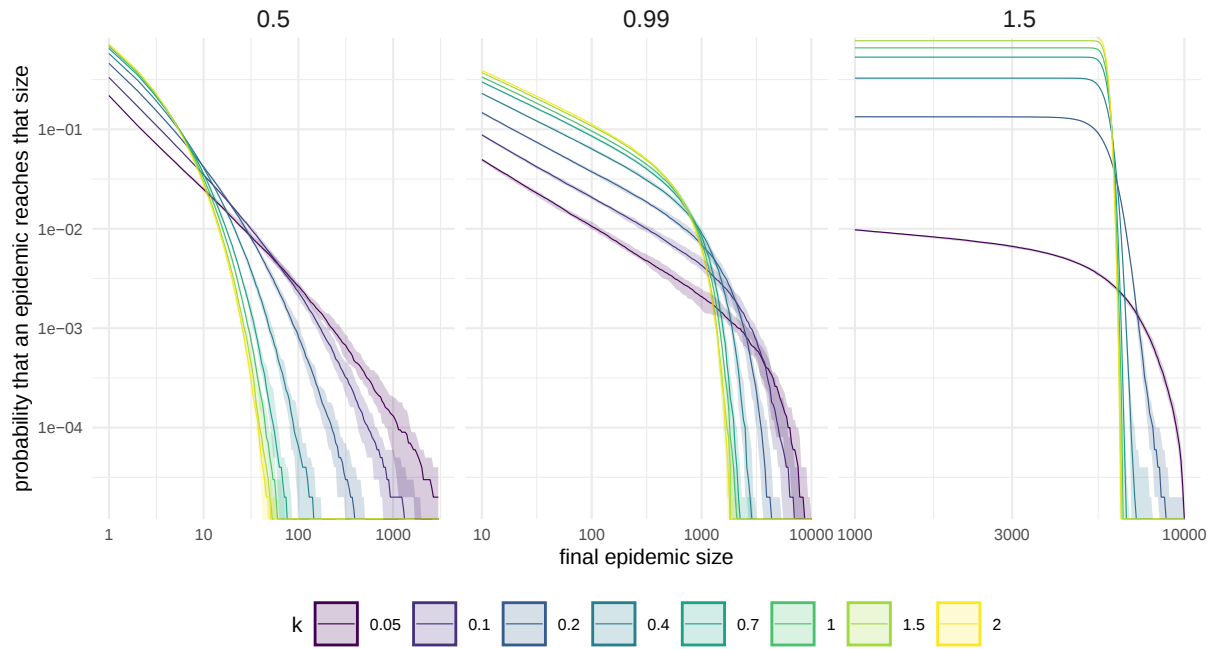


Figure S1: **Final epidemic size cumulative distribution, as a function of secondary cases heterogeneity.** Results are presented in the case without evolution, and with three different R_0 : 0.5, 0.99, 1.5. No difference was observed between the sources of heterogeneity, and the results are presented combining all ranges of simulated CV_T .

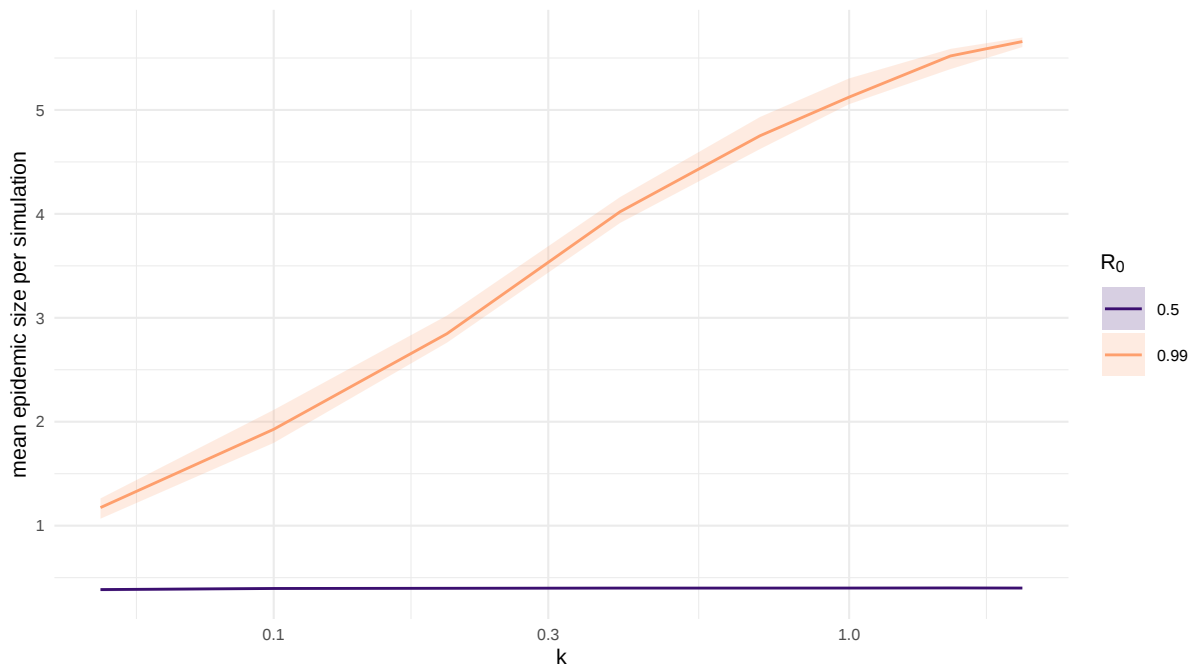


Figure S2: **Mean number of infected individuals during an epidemic, as a function of the secondary cases heterogeneity and R_0 .**

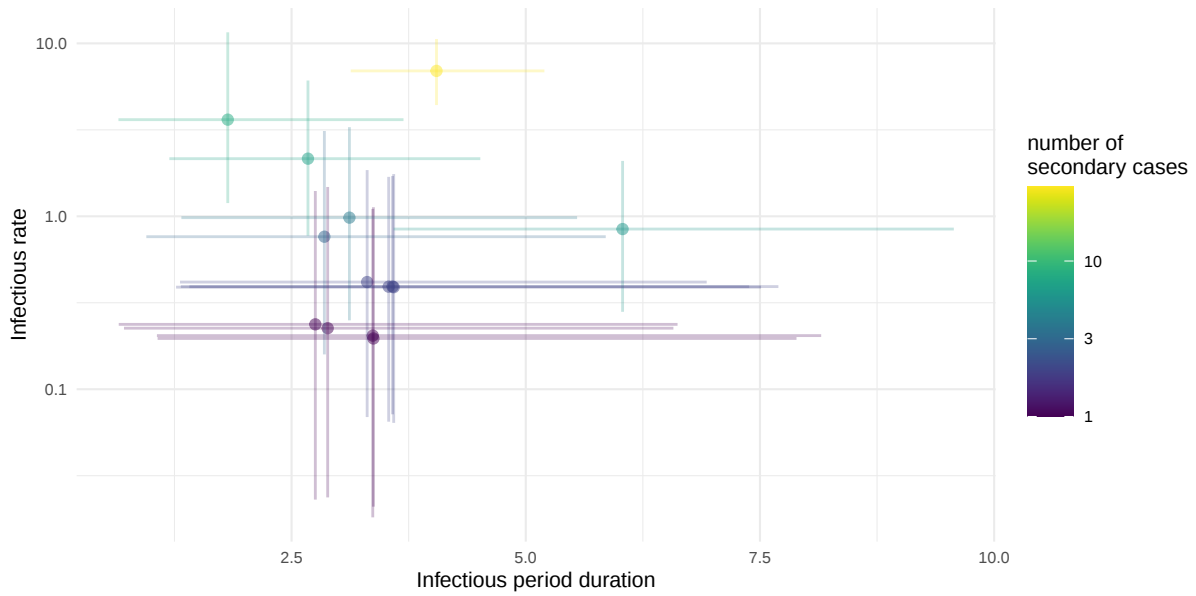


Figure S3: **Relationship between the estimated transmission rate and infection duration for Measles outbreak in Hagelloch.** The parameters could be jointly estimated thanks to patient line data [6] in our Bayesian model. No significant correlation was found between the two metrics, even when removing the superspreading event (Spearman's rank correlation $\rho = -0.34$, $p = 0.25$). Line ranges represent the 95% credible interval.

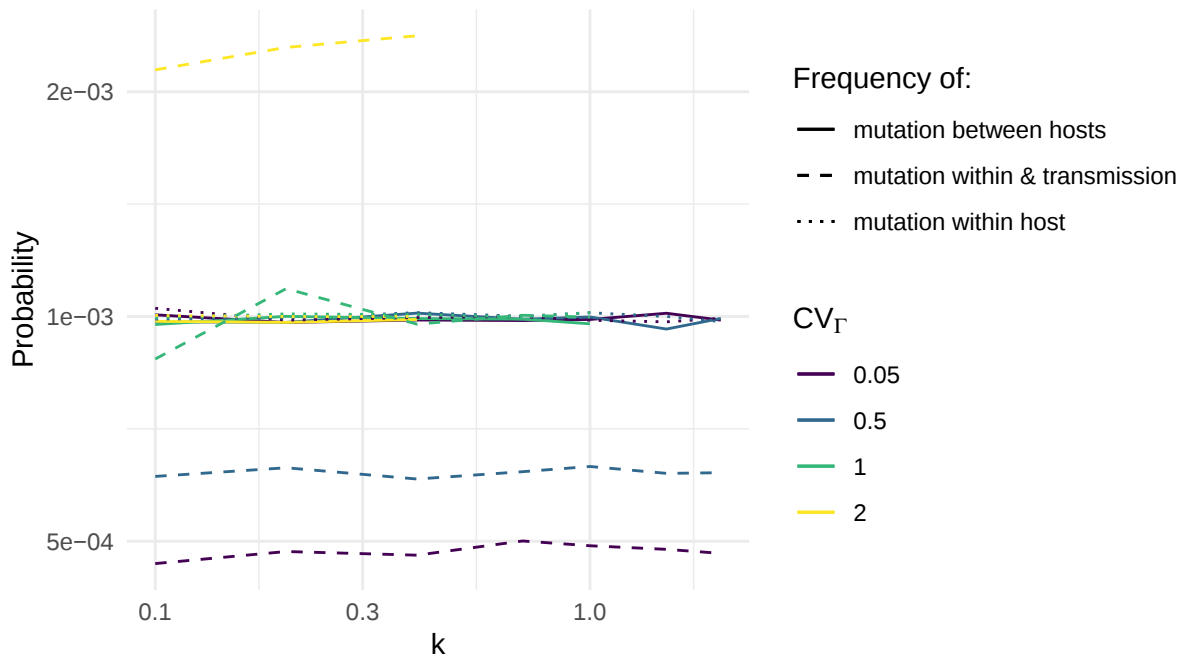


Figure S4: **Probability of mutation as a function of the mutation scenario, k and CV_T .**

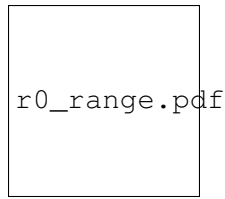


Figure S5: **Sensitivity analysis: the impact of R_0 on the metrics without evolution.** In order to simplify the reading, results are shown for only two values of CV_T . Results should be read similarly to figure ???. For each R_0 value, relative results are shown with $k = 1$, $CV_T = 1$ and the corresponding R_0 as a reference. A. Frequency of emergence as a function of k and R_0 . B. Stochastic phase relative duration. C. Relative doubling time. D. Prevalence peak size. E. Final outbreak size.