The source of individual heterogeneity shapes infectious disease outbreaks

Baptiste Elie*, Christian Selinger, Samuel Alizon

MIVEGEC, Univ Montpellier, CNRS, IRD, Montpellier, France

* author for correspondance: baptiste.elie@ird.fr

Abstract

There is known heterogeneity between individuals in infectious disease transmission patterns and the source of this heterogeneity is thought to affect epidemiological dynamics. However, a potential bias is that studies tend not to control for the overall heterogeneity in the number of secondary cases caused by an infection. To explore the role of individual variation in infection duration and transmission rate on parasite emergence and spread, while controlling for this bias, we simulate stochastic outbreaks with and without parasite evolution. We find, as expected, that heterogeneity in the number of secondary cases decreases the probability of outbreak emergence, but we also find that the source of this heterogeneity matters. Furthermore, for epidemics that do emerge, assuming more realistic variances in infection duration distributions leads to faster outbreaks and higher epidemic peaks. When parasites require adaptive mutations to cause large epidemics, the impact of heterogeneity depends on the underlying evolutionary model. If within-host evolution matters, decreasing the infection duration variance decreases the probability of emergence. These results underline the importance of accounting for realistic distributions of infection duration to accurately anticipate the effect of individual heterogeneity on epidemiological dynamics.

Keywords: epidemiology, modelling, infection duration, superspreading, evolutionary rescue, emerging infectious diseases

The expected number of secondary cases produced by an infected individual in a naive population is a key concept in epidemiology [3, 23]. It is classically referred to as the basic reproduction number and denoted R_0. Only infections with $R_0 > 1$ can cause major outbreaks. However, this mean value does not reflect the impact of super-spreading events, where an individual causes an unusually large number of secondary cases [12, 19, 27, 31, 33, 43]. The more frequent these events are, the higher the variance in the number of secondary cases, and, therefore, the lower the probability of outbreak emergence and the faster the epidemic growth for outbreaks that do emerge [31].

Several biological processes can explain the heterogeneity in the number of secondary cases [38]. However, models investigating these processes tend only to vary one source of heterogeneity at a time. By doing so, they do not control for the (overall) heterogeneity in the number of secondary cases, which is known to have strong effects, independently of its source [31]. One of the few exceptions suggests that the biology matters since it finds, for instance, that heterogeneity in host susceptibility has a lesser impact on the probability of emergence than heterogeneity in transmission rate, which can be defined as the product between a contact rate and the probability of transmission given that there is a contact between two individuals [44].

We use a stochastic mechanistic model to explore whether heterogeneity in transmission rates or recovery rates have different effects on an epidemic spread. Based on earlier models, we hypothesise that a more homogeneous distribution of infectious period duration decreases the variability of population dynamics in the early outbreak, therefore increasing the probability of outbreak extinction [2], but also increasing epidemic growth as well as epidemic peak size [2, 32]. However, contrarily to us, these studies do not control for the distribution of the number of secondary cases, which could explain part of their results.

Even if initially maladapted (*i.e.* $R_0 < 1$), a parasite can evolve into a well-adapted strain before fading out and then cause a major outbreak, a phenomenon called evolutionary emergence or evolutionary rescue [5, 15]. Since, the final epidemic size increases with increasing heterogeneity in transmission rate [17], we hypothesise that the source of heterogeneity could affect evolutionary emergence.

Following earlier studies [17, 21, 31], we assume that the number of secondary infections caused by each individual follows a Negative-Binomial distribution $\mathscr{Z}$ with mean R_0 and dispersion parameter k. The smaller k, the more dispersed $\mathscr{Z}$. For example, in the case of the 2003 SARS outbreak in Singapore, where superspreading events were shown to be frequent, the dispersion was estimated to be $k = 0.16$ [31]. Furthermore, recent data from COVID-19 epidemics estimate k to be in the order of 0.3 [36].

We model individual transmission rates and infection duration values using gamma distributions, denoted respectively $\mathscr{B}$ and Γ. For a given distribution $\mathscr{Z}$, we vary the coefficient of variation (CV) of

the transmission rate (CV_B) and the CV of infection duration (CV_Γ) to compare heterogeneity originating from differences in infectious period duration and transmission rate.

We simulate outbreaks, without and with evolution, and measure key summary statistics to analyze the impact of different sources of heterogeneity on emerging outbreaks properties. We confirm that the dispersion of the distribution of the number of secondary infections ($\mathscr{Z}$) is the main driver of the frequency of emergence, but we also find that the source of this heterogeneity affects the risk of emergence, the properties of emerging epidemics, and even some aspects of evolutionary emergence.

Material and Methods

Model without evolution

We implement a non-Markovian version of the Susceptible-Infected-Recovered (SIR) epidemiological model [24], which means that not all rates are held constant throughout an infection. We assume that the host population is of fixed size N and that epidemics are initiated by a single infectious individual. At time t, each individual is characterized by its current state (susceptible, infectious, or removed), and, if infected, the time at which it will recover.

There are two sources of heterogeneity in the model:

i) Transmission rate. We take into account individual heterogeneity by drawing the *per capita* transmission rate β_i for each individual i from a Gamma distribution, denoted $\mathscr{B}$. For mathematical convenience, and without further qualitative impact, we set the mean of $\mathscr{B}$ such that $\mathbb{E}[B]S_0 \approx \mathbb{E}[B]N = 1$, with S_0 the initial number of susceptible individuals. The standard deviation of $\mathscr{B}$ is imposed by the choice of the coefficient of variation (CV_B) which is further detailed below. From a biological standpoint, this heterogeneity can have a behavioural (i.e. contact rates) or a biological (i.e. infectiousness) origin.

ii) Infection duration. We assume that individuals remain in the I compartment for a time drawn randomly from a Gamma distribution, noted Γ. By construction, the expectation of Γ is R_0 in our model and we vary its coefficient of variation between 0.2 and 1, which determines its shape, which equals to $1/CV_\Gamma^2$. Variations in infection duration main have a biological origin.

Coefficients of variation and $\mathscr{Z}$ dispersion

Heterogeneities in transmission rate and infection duration determine the distribution of the number of secondary infections ($\mathscr{Z}$). Little is known about the biological range of CV_B, whereas contact-tracing studies can provide some insights into the ranges of values of k and CV_Γ. Therefore, we fix the values of k and CV_Γ to compute the resulting value of CV_B.

More specifically, we look for the value of CV_B that results in the smallest Kolmogorov-Smirnov distance between the Negative Binomial distribution with mean R_0 and dispersion parameter k and the computed distribution of the secondary cases determined by CV_B and CV_Γ, simulated by repeating 10^6 times the following procedure. First, we draw the transmission rate β_i following the $\mathscr{B}$ distribution. Second, let γ_i be the infection duration of this individual. Since transmission events are assumed to be memoryless and independent, the number of secondary cases follows a homogeneous pure-birth/Poisson process, the rate of which is proportional to the infection duration γ_i, *i.e.* $\zeta_i \sim \text{Poisson}(\beta_i N \gamma_i)$. Since γ_i is drawn from a Gamma distribution with shape κ and mean R_0, the overall process follows a Negative Binomial distribution with mean $\beta_i R_0$ and size κ. Therefore, we can draw the number of secondary cases of this individual following this distribution.

Evolutionary emergence model

We introduce an additional class of individuals by distinguishing between I_r and I_m, which refer to individuals infected by the initial resident (resp. new mutant) parasite strain. The reproduction numbers of these strains are denoted R_0^r and R_0^m. We start with one infected individual of class I_r. Parasites can mutate during a transmission event or take over an infected host, both at a rate μ. We assume that the mutation increases their mean transmission rate without altering CV_B (therefore also increasing the standard deviation of $\mathscr{B}$). We further assume that the infectious period duration is not impacted by the mutation. Coinfections are neglected. Therefore, in the case of within-host mutation, we assume that the mutant instantaneously takes over the host.

Frequency of emergence

We use the total epidemic size to determine if an outbreak has emerged or not. Emergence is assumed to occur when the total epidemic size is greater than the herd immunity threshold, *i.e.* $1 - 1/R_0$ [3]. This implies any cross-reactive strain with the same R_0, would not be able to create an outbreak in the population. In the case of evolutionary emergence, we only consider the emergence of the mutant strain.

Numerical simulations

We simulate epidemics, i.e. the succession of infection and recovery events, using Gillespie's next reaction method [18] to generate non-Markovian distributions. The algorithm runs as follows:

1. *Initialize (i.e. set S, $I = 1$, $t = 0$)*
2. *In case of new infected individual i, draw β_i and γ_i.*

3. *Update the new force of infection S. $\sum_{i=1}^{I} \beta_i$ and draw the time to the next infection assuming an exponential distribution.*
4. *Look for the event with the closest time of occurrence, and update the compartments (S, I).*
5. *Update the time t to the time of the new event.*
6. *Go back to step 2.*

In the evolutionary emergence scenario with mutation during transmission, the model is similar but it includes two forces of infection, one for each class of infected host (I_r, I_m), and two additional events: infection by the mutant strain (assuming an exponential distribution with a rate $\sum_{i=1}^{I_r} \beta_i \mu + \sum_{j=1}^{I_m} \beta_j$), and recovery of an individual in the I_m class.

In the scenario where the mutant first takes over the host, we distinguish the event of infection by the mutant strain (assuming an exponential distribution with a rate $\sum_{j=1}^{I_m} \beta_j$) from the within-host mutation of a resident strain into a mutant strain (assuming an exponential distribution with a rate $I_r * \mu$).

The model was developed for this study and implemented in Java 11.0.7. We used parallel computation to decrease computing time.

We monitor the following metrics:

- The time until prevalence reaches 25, 50, 125, 250, 500, and 1,000 infected hosts. This enables us to estimate the growth rate during the exponential phase.
- The prevalence peak size and outbreak final size. In the evolutionary emergence simulations, we distinguish the resident final size from that of the mutant.
- When the pathogen can mutate into a more adapted strain, we register the time at which the mutation occurs, the number of infected hosts by the resident at this time, and the cumulative number of infected hosts by the resident at this time.

Simulation outputs were analyzed with R 3.6.3.

Parameters estimation for known outbreaks

To estimate CV_B and CV_Γ from observed outbreaks, we looked for epidemiological data in the literature with both serial interval and secondary cases available from the same outbreak. Since those two distributions are not independent, we looked for joint distributions but these were only available for Measles outbreaks [22]. For Ebola [13, 42], pneumonic plague [16], Smallpox [14, 35], and SARS outbreaks [29, 31] we had to assume that the two distributions were independent (see the Table S1 for further details about the data and parameters sources).

To obtain biologically relevant parameters from these empirical data, we infer parameters assuming a model with a latent period, the distribution of which we set using independent sources in the literature

[6, 26, 34, 42]. For simplicity, we assume that for a given parasite the distribution of the latent period does not vary between outbreak. We also use independent estimates of R_0 [22, 29, 42]. In the end, we estimate the following three free parameters: the mean infection period, CV_B and CV_Γ.

Assuming the pool of susceptible hosts remained constant during the outbreak, we simulate 5.10^6 secondary cases and serial intervals for a given mean Γ, CV_B and CV_Γ, in order to obtain their distribution. From these, we compute the likelihood of the observed data, jointly when possible, or independently otherwise. We then obtain the maximum likelihood parameters using a grid search optimization.

Code and data availability

The scripts used are available on the following GitLab repository: `https://gitlab.in2p3.fr/ete/heterogeneity-outbreak`.

Results

Emergence risk

In our model, the dispersion of the number of secondary cases k is entirely determined by the coefficients of variation in transmission rate (CV_B) and in infection duration (CV_Γ). This is shown in Figure 1A and further explained in the Methods.

Practically, it is more straightforward to estimate the values of k and CV_Γ, and then impose a value for CV_B because the data to infer the distributions of the number of secondary infections and the infection duration periods are easier to collect than that to infer distributions of transmission rates. We illustrate the feasibility of such inferences by indicating the parameter value for several well-studied outbreaks in Figure 1A. This also shows that the parameters ranges studied here are biologically realistic.

Figure 1B shows the proportion of outbreaks that emerge as a function of the source of heterogeneity and of R_0. For a given R_0 value, the probability of emergence appears to be mainly governed by the individual reproduction number dispersion k. However, for epidemics with many superspreading events (low k values), the frequency of emergence also increases with more homogeneous infection duration distributions (lower CV_Γ values). This effect is amplified for lower R_0 values.

In the next section, we analyze the properties of the outbreaks that do emerge, with a R_0 of 1.5.

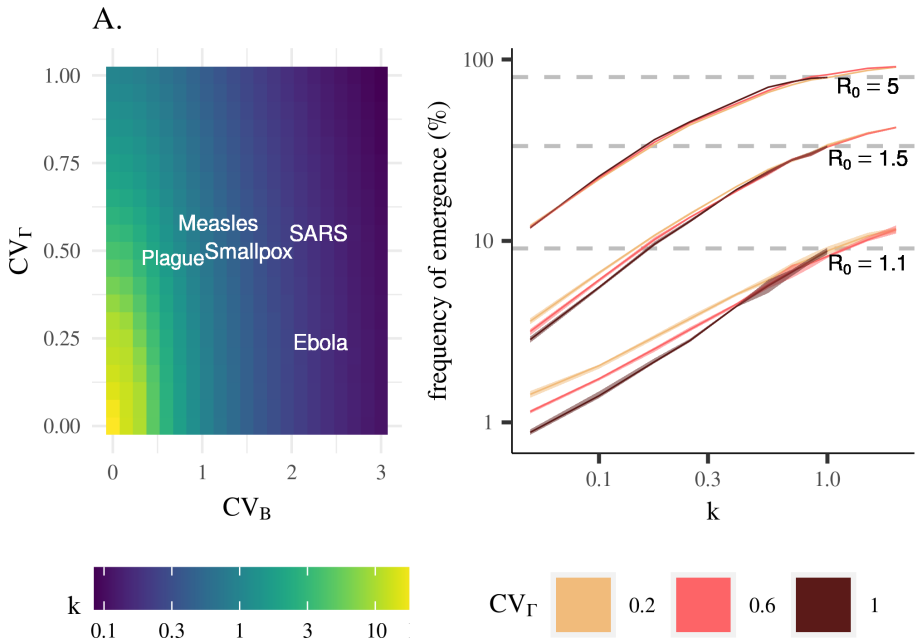

Figure 1 – **Relationship between k, CV_B, CV_Γ and the risk of outbreak emergence.** A) Numerical estimation of the secondary cases heterogeneity (k) as a function of CV_B and CV_Γ. Names in white show the range of values estimated using maximum likelihood methods from outbreak data. If k remains constant, increasing CV_Γ always decreases CV_B. B) Frequency of emergence of an outbreak starting from one infection as a function of model heterogeneity and R_0. The value of CV_B is determined by that of CV_Γ and k. The lines represent the frequency of emergence out of 20,000 runs. The dashed grey lines represent the theoretical frequency of emergence in a Markovian model (*i.e k=1)*. To illustrate the variability which can be expected from the numerical estimates, each frequency estimation is repeated 10 times, and the shaded area represent the two most extreme results. Note that when $k > 1$ it is mathematically impossible to have $CV_\Gamma = 1$.

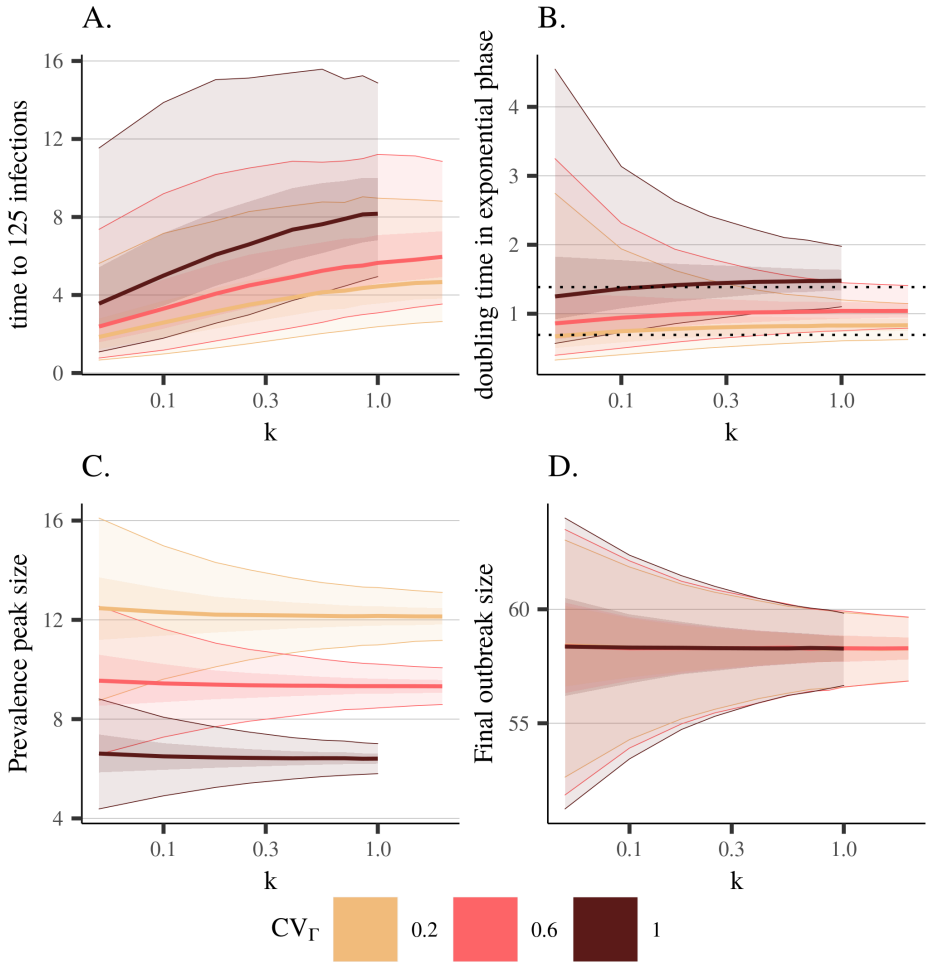

Figure 2 – **Properties of emerging outbreaks.** A) Time until prevalence reaches 125 infected hosts. B) Doubling time during the exponential phase (*i.e.* going from a prevalence of 250 to 500 infections). Top and bottom dotted lines indicate the expected doubling time for a memoryless model with $R_0 = 1.5$ or $R_0 = 2$, respectively. C) Prevalence peak size. D) Cumulative number of infections. Values are expressed as % of the total population. Lines represent the mean value out of the simulated outbreaks that emerged. The shaded areas represent the 95% confidence interval.

Dynamic of emerging outbreaks

Growth rates

We first study the initial phase of emerging outbreaks, when the number of events is low. Since the law of large numbers does not apply there, variations in prevalence are strongly affected by stochasticity (Fig. S2). We quantify the early growth during this stochastic phase by measuring the time until the prevalence reaches 125 infected individuals, which corresponds to the outbreak threshold with 99% of probability in the most dispersed secondary cases scenario, according to existing models [20]. As expected [31], we find that decreasing k leads to faster epidemic growth. For a given k, we further find that increasing the heterogeneity in infection duration (therefore decreasing that in transmission rate) also increases the early epidemic growth (Fig. 2A).

We then study the deterministic exponential growth phase, which follows the stochastic phase and ends when the depletion of susceptible host population cannot be neglected anymore [20] (Fig. S2). Figure 2B shows that, contrarily to the beginning of the outbreak, the growth rate during this exponential phase is mostly impacted by CV_Γ, with an extremely limited effect of k. When assuming an exponential distribution for the infection duration (*i.e.* $CV_\Gamma = 1$, black line in Figure 2B), CV_B and, therefore, k have little effect: the doubling time remains in the order of 1.4 days, which is expected with $R_0 = 1.5$. Decreasing CV_Γ leads to a strong decrease in doubling time. Therefore, ignoring the distribution of the infection duration, *e.g.* by implicitly assuming an exponential distribution, can lead to over-estimating R_0 by up to 33%.

Epidemic peak size and final size

The prevalence peak value is highly affected by the heterogeneity in the infection duration: its median increases from 6% to 12% of the population when the CV_Γ decreases from 1 to 0.2 (Fig. 2C). k has little effect on the mean epidemic peak size, but we do find a correlation between the variance in peak size and that of $\mathscr{Z}$.

Finally, none of our heterogeneity metric seems to affect the median final epidemic size, which is always close to 58% of the population (Fig. 2 D), corresponding to the expected value for $R_0 = 1.5$ according to classical theory [24]. As for the other metrics, the variance in the total epidemic size decreases with k.

Evolutionary emergence

Even if $R_0 < 1$, parasites can cause major outbreaks in a population if they acquire adaptive mutations. We now assume that the introduced 'resident' strain has a $R_0^r < 1$ and, therefore, will go extinct unless it evolves into a phenotypically different 'mutant' strain with $R_0^m > 1$. We also assume that the mutant strain can arise either by taking over a host infected by the resident strain or during a transmission event.

We first study the frequency of simulations having at least one mutation event occurring because this does not depend on the origin of the mutant (replacement in the host or via transmission event). In Figure 3A, we show that this frequency follows the same pattern as the frequency of (non-evolutionary) emergence (increasing CV_Γ decreases the probability of emergence, Fig. 1B). This makes sense since mutants cannot emerge if outbreaks go rapidly extinct.

However, the way the mutant emerges does affect the role of heterogeneity on the probability of outbreak (evolutionary) emergence. We explored two extreme cases: i) the mutant appears within the host, instantaneously replacing the resident strain, or ii) the mutant emerges during transmission, thanks

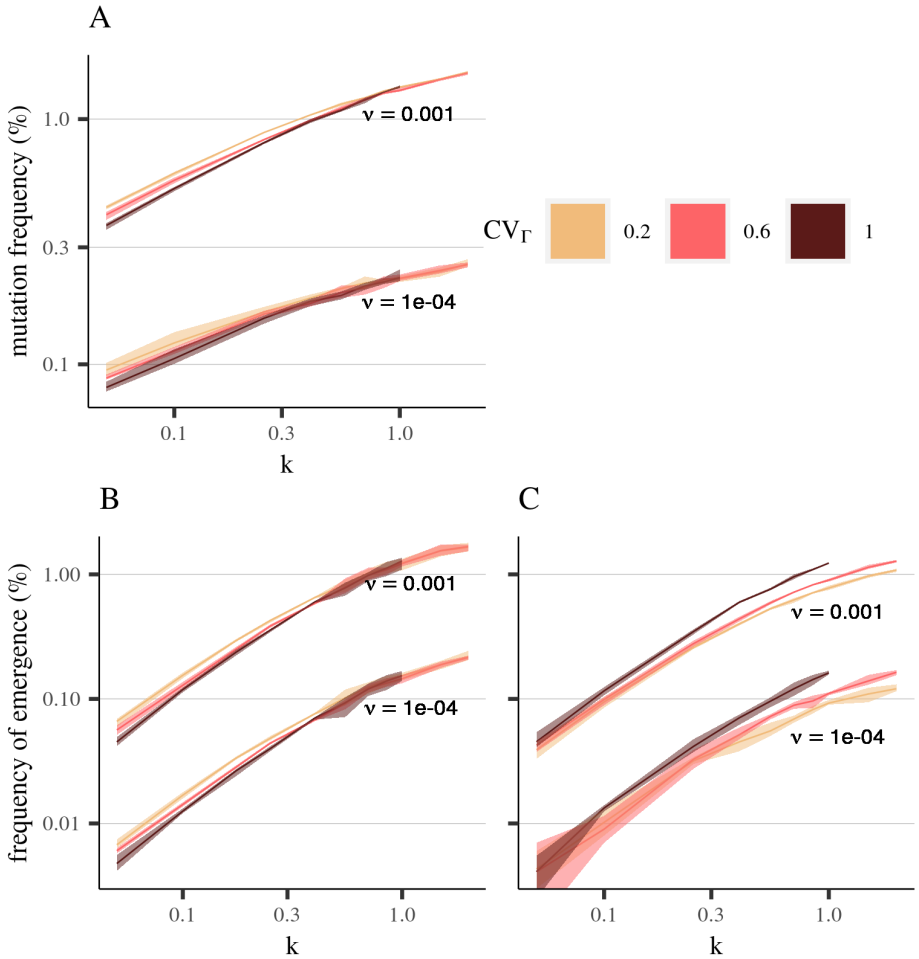

Figure 3 – **Individual heterogeneity and evolutionary emergence.** Results are presented as a function of secondary cases dispersion (k), infection duration coefficient of variation (CCV_Γ) and mutation rate (ν). A) Frequency of mutation (the origin of the mutation can be ignored), B) Frequency of emergence in the scenario when mutations occur during transmission, and C) when mutants take over the host. The lines show the frequency calculated on 20,000 runs. Each frequency estimation is repeated 10 times, and the shaded area represent the two most extreme results.

to the small inoculum size [4]. The frequency of emergence is similar in the two scenarios when $CV_\Gamma = 1$, but it diverges when the heterogeneity in infectious duration increases. When the mutation occurs during a transmission event (Fig. 3B), the probability of emergence is directly correlated to the frequency of mutation. Conversely, when the mutation occurs within the host, decreasing the heterogeneity of the infection duration leads to a decrease in the frequency of emergence, regardless of the individual secondary cases heterogeneity (Fig. 3C). This effect seems not to be impacted by the mutation rate ν.

Discussion

Individuals vary from their biology, their history, and their behaviour. In the context of an epidemic, all this variation can be aggregated into a single metric, the individual reproduction number, i.e. the

number of secondary infections caused by an individual. The distribution of these individual numbers shapes infectious disease outbreaks [31]. Several studies tried to be more specific to investigate the effect variations in a specific trait can have on epidemiological dynamics. However, the majority of studies overlook the fact that variations in an infection trait (e.g. the distribution of the duration of infectious periods) may also affect the distribution of individual reproduction number. Therefore, understanding the effect of biological variations of a given trait on the course of outbreak requires controlling for variations in the number of secondary cases. In this study, this is what we did to investigate the relative effects of variation in the infection duration (i.e. how long individuals are infectious for) and transmission rate (how contagious an infected personis).

Many models involving ordinary differential equations are memoryless, in that the probability of occurrence of an event does not depend on the history of individuals (but see Anderson & Watson [2], Lloyd [30], Malice & Kryscio [32]). While this may be true for transmission events, provided that the infection rate is constant, it is biologically unrealistic for recovery events since they often depend on the number of days since infection [11, 28]. When assuming a high, but biologically realistic [12, 31], heterogeneity in the number of secondary cases (low k), we find that the source of heterogeneity matters and that increased heterogeneity in the infection duration is associated with decreased frequencies of outbreak emergence. Importantly, this result is obtained while controlling for the distribution of secondary cases. Our interpretation is that decreasing the infection duration heterogeneity tends to decrease the stochastic phase duration (Fig.2A), thereby increasing the probability of emergence.

Earlier results find that in branching process models, increasing the heterogeneity in transmission rates leads to a faster increase in cases per generation among the outbreaks that do emerge [31]. We show that this effect does not translate into an increased growth rate after the epidemic evades the stochastic phase and its dynamic becomes deterministic. We can detect this effect because, contrarily to earlier studies, we use simulate the entire course of an epidemic. Note that an alternative approach could be to use recent developments of branching process theory in epidemiology to incorporate the depletion of susceptible hosts [7]. Being able to compare, among the epidemics that emerge, key characteristics such as the final cumulative incidence or the peak incidence size allows us to show, for instance, that even if the early growth rate is high, once the exponential growth phase is reached, superspreading events play a limited role in the epidemic growth.

Contrarily to superspreading events (parameter k), we show that the heterogeneity in infectious period duration plays an important role in the deterministic phase of the epidemic, by increasing the growth rate and, more strikingly, the prevalence peak size. While previous studies already reported a similar effect on both metrics [2, 40**?** , 41], our study further shows that this phenomenon is intrinsically

related to the infectious period duration, and not to the secondary cases heterogeneity (which was not kept constant in these previous studies). Indeed, as it has been earlier explained [40], a more heterogeneous infectious period leads to a longer generation time because transmission rely on long infections. This increases the doubling time and flattens the epidemic curve compared to an epidemic with a more homogeneous infectious period duration.

Finally, we show that infectious period duration heterogeneity can affect evolutionary emergence depending on the process that generates the mutant infection [15]. The impact of the mutational pathway and evolutionary scenario have been already pointed out by several studies [1, 44]. First, as expected, we find no difference between the two mutation scenarios if the process is memoryless -that is the duration of the infections is assumed to be exponentially distributed ($CV_\Gamma = 1$). This further underlines the importance of questioning this biologically-unrealistic assumption [2, 9, 30, 32]. When assuming more realistic infection duration distributions, we find that if mutations appear during transmission events, then the evolutionary emergence trends are identical to the case without evolution. However, when the mutation appears within a host during the infection, infection duration heterogeneity increases the frequency of emergence. An interpretation is that longer infectious periods represent more opportunities to transmit for the infection, thereby increasing the probability that the mutation causes an outbreak [4].

Our effort to maintain a simple and tractable model of outbreak emergence naturally leads to several limitations. In particular, there is an identifiability issue regarding the biological bases of the transmission rate heterogeneity, which could originate from variations in transmission rate or in host susceptibility. On this issue, see the results from Yates *et al.* [44]. It could also be interesting to enrich the model by considering a latent period during which exposed hosts are not yet infectious. This has been shown to affect R_0 estimates but in a deterministic model that did not take into account superspreading events [41]. More generally, investigating other sources of heterogeneity of the number of secondary infections may help uncover potential biases.

We assumed that the population have no spatial structure, which is more realistic for directly transmitted diseases, such as SARS or measles, than for sexually transmitted infections for which contact networks impose strong constraints [17]. Furthermore, at the beginning of an epidemic, spatial structure appears to have little effect on outbreak metrics, especially R_0 [37]. However, it is known that heterogeneity in host susceptibility and spatial structure decrease the final epidemic size, *i.e.* the total proportion of the population infected throughout the epidemic [8, 39]. We also do not include host demography and limit our analysis to a single epidemic wave.

Finally, this analysis relies on numerical results. This enables us to explore the role of stochasticity, which is particularly important to consider in the context of outbreak emergence from a mathematical

modelling [10] and a statistical inference [25] point of view. However, it limits our analysis to the area of punctual parameters that we selected as being biologically relevant.

These theoretical results have implications for outbreak monitoring. In particular, we show that making simplifying but biologically unrealistic assumptions about the distributions of infection duration can lead to underestimating the risk of emergence, the epidemic doubling time, and the prevalence peak size. Given the risk of saturation of healthcare systems, accurately anticipating these values is a major issue. This stresses the importance of collecting detailed biological data to better inform epidemiological models.

Acknowledgements

The authors thank the CNRS, the IRD, and acknowledge the itrop HPC (South Green Platform) at IRD Montpellier for providing HPC resources that have contributed to the research results reported within this study (https://bioinfo.ird.fr/).

References

[1] Alexander, H. K. & Day, T., 2010 Risk factors for the evolutionary emergence of pathogens. *Journal of The Royal Society Interface* **7**, 1455–1474. (doi: 10.1098/rsif.2010.0123).

[2] Anderson, D. & Watson, R., 1980 On the spread of a disease with gamma distributed latent and infectious periods. *Biometrika* **67**, 191–198. (doi: 10.1093/biomet/67.1.191).

[3] Anderson, R. M. & May, R. M., 1992 *Infectious Diseases of Humans: Dynamics and Control*. United Kingdom: Oxford Univ. Press, original edn.

[4] André, J.-B. & Day, T., 2005 The effect of disease life history on the evolutionary emergence of novel pathogens. *Proceedings of the Royal Society B: Biological Sciences* **272**, 1949–1956. (doi: 10.1098/rspb.2005.3170).

[5] Antia, R., Regoes, R. R., Koella, J. C. & Bergstrom, C. T., 2003 The role of evolution in the emergence of infectious diseases. *Nature* **426**, 658–661. (doi: 10.1038/nature02104).

[6] 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).

[7] Barbour, A. & Reinert, G., 2013 Approximating the epidemic curve. *Electronic Journal of Probability* **18**. (doi: 10.1214/EJP.v18-2557).

[8] Becker, N. & Marschner, I., 1990 The effect of heterogeneity on the spread of disease. In *Stochastic Processes in Epidemic Theory* (eds. J.-P. Gabriel, C. Lefèvre & P. Picard), pp. 90–103. Berlin, Heidelberg: Springer Berlin Heidelberg.

[9] Britton, T. & Lindenstrand, D., 2008 Epidemic modelling: Aspects where stochasticity matters. *arXiv:0812.3505 [math, q-bio]* .

[10] Britton, T. & Pardoux, E., 2019 Stochastic epidemics in a homogeneous community. *arXiv:1808.05350 [math]* **2255**. (doi: 10.1007/978-3-030-30900-8).

[11] Chan, M. & Johansson, M. A., 2012 The Incubation Periods of Dengue Viruses. *PLoS ONE* **7**, e50972. (doi: 10.1371/journal.pone.0050972).

[12] Endo, A., Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Abbott, S., Kucharski, A. J. & Funk, S., 2020 Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Research* **5**, 67. (doi: 10.12688/wellcomeopenres.15842.1).

[13] 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).

[14] Fenner, F., Henderson, D. A., Arita, I., Jezek, Z., Ladnyi, I. D. & Organization, W. H., 1988 *Smallpox and Its Eradication*. World Health Organization.

[15] Gandon, S., Hochberg, M. E., Holt, R. D. & Day, T., 2013 What limits the evolutionary emergence of pathogens? *Philosophical Transactions of the Royal Society B: Biological Sciences* **368**, 20120086. (doi: 10.1098/rstb.2012.0086).

[16] Gani, R. & Leach, S., 2004 Epidemiologic Determinants for Modeling Pneumonic Plague Outbreaks. *Emerging Infectious Diseases* **10**, 608–614. (doi: 10.3201/eid1004.030509).

[17] Garske, T. & Rhodes, C., 2008 The effect of superspreading on epidemic outbreak size distributions. *Journal of Theoretical Biology* **253**, 228–237. (doi: 10.1016/j.jtbi.2008.02.038).

[18] Gibson, M. A. & Bruck, J., 2000 Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels. *The Journal of Physical Chemistry A* **104**, 1876–1889. (doi: 10.1021/jp993732q).

[19] Gomes, M. G. M., Águas, R., Lopes, J. S., Nunes, M. C., Rebelo, C., Rodrigues, P. & Struchiner, C. J., 2012 How host heterogeneity governs tuberculosis reinfection? *Proceedings of the Royal Society B: Biological Sciences* **279**, 2473–2478. (doi: 10.1098/rspb.2011.2712).

[20] Hartfield, M. & Alizon, S., 2014 Epidemiological Feedbacks Affect Evolutionary Emergence of Pathogens. *The American Naturalist* **183**, E105–E117. (doi: 10.1086/674795).

[21] Hellewell, J., Abbott, S., Gimma, A., Bosse, N. I., Jarvis, C. I., Russell, T. W., Munday, J. D., Kucharski, A. J., Edmunds, W. J., Funk, S., Eggo, R. M., Sun, F., Flasche, S., Quilty, B. J., Davies, N., Liu, Y., Clifford, S., Klepac, P., Jit, M., Diamond, C., Gibbs, H. & van Zandvoort, K., 2020 Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health* **8**, e488–e496. (doi: 10.1016/S2214-109X(20)30074-7).

[22] Jombart, T., Frost, S., Nouvellet, P., Campbell, F. & Sudre, B., 2020 Outbreaks: A Collection of Disease Outbreak Data.

[23] Keeling, M. J. & Rohani, P., 2008 *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press.

[24] Kermack, W. O. & McKendrick, A. G., 1927 A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A* **115**, 700–721. (doi: 10.1098/rspa.1927.0118).

[25] King, A. A., Domenech de Cellès, M., Magpantay, F. M. G. & Rohani, P., 2015 Avoidable errors in the modelling of outbreaks of emerging pathogens, with special reference to Ebola. *Proceedings of the Royal Society B: Biological Sciences* **282**, 20150347. (doi: 10.1098/rspb.2015.0347).

[26] 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).

[27] Lemieux, J. E., Siddle, K. J., Shaw, B. M., Loreth, C., Schaffner, S. F., Gladden-Young, A., Adams, G., Fink, T., Tomkins-Tinch, C. H., Krasilnikova, L. A., DeRuff, K. C., Rudy, M., Bauer, M. R., Lagerborg, K. A., Normandin, E., Chapman, S. B., Reilly, S. K., Anahtar, M. N., Lin, A. E., Carter, A., Myhrvold, C., Kemball, M. E., Chaluvadi, S., Cusick, C., Flowers, K., Neumann, A., Cerrato, F., Farhat, M., Slater, D., Harris, J. B., Branda, J. A., Hooper, D., Gaeta, J. M., Baggett, T. P., O'Connell, J., Gnirke, A., Lieberman, T. D., Philippakis, A., Burns, M., Brown, C. M., Luban, J., Ryan, E. T., Turbett, S. E., LaRocque, R. C., Hanage, W. P., Gallagher, G. R., Madoff, L. C., Smole, S., Pierce, V. M., Rosenberg, E., Sabeti, P. C., Park, D. J. & MacInnis, B. L., 2021 Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of superspreading events. *Science* **371**, eabe3261. (doi: 10.1126/science.abe3261).

[28] Lessler, J., Reich, N. G., Brookmeyer, R., Perl, T. M., Nelson, K. E. & Cummings, D. A., 2009 Incubation periods of acute respiratory viral infections: A systematic review. *The Lancet Infectious Diseases* **9**, 291–300. (doi: 10.1016/S1473-3099(09)70069-6).

[29] 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.

[30] Lloyd, A. L., 2001 Realistic Distributions of Infectious Periods in Epidemic Models: Changing Patterns of Persistence and Dynamics. *Theoretical Population Biology* **60**, 59–71. (doi: 10.1006/tpbi.2001.1525).

[31] 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).

[32] Malice, M.-P. & Kryscio, R. J., 1989 On the Role of Variable Incubation Periods in Simple Epidemic Models. *Mathematical Medicine and Biology* **6**, 233–242. (doi: 10.1093/imammb/6.4.233).

[33] Marm Kilpatrick, A., Daszak, P., Jones, M. J., Marra, P. P. & Kramer, L. D., 2006 Host heterogeneity dominates West Nile virus transmission. *Proceedings of the Royal Society B: Biological Sciences* **273**, 2327–2333. (doi: 10.1098/rspb.2006.3575).

[34] 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).

[35] 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).

[36] Sun, K., Wang, W., Gao, L., Wang, Y., Luo, K., Ren, L., Zhan, Z., Chen, X., Zhao, S., Huang, Y., Sun, Q., Liu, Z., Litvinova, M., Vespignani, A., Ajelli, M., Viboud, C. & Yu, H., 2021 Transmission heterogeneities, kinetics, and controllability of SARS-CoV-2. *Science* **371**. (doi: 10.1126/science.abe2424).

[37] Trapman, P., Ball, F., Dhersin, J.-S., Tran, V. C., Wallinga, J. & Britton, T., 2016 Inferring R_0 in emerging epidemics—the effect of common population structure is small. *Journal of The Royal Society Interface* **13**, 20160288. (doi: 10.1098/rsif.2016.0288).

[38] VanderWaal, K. L. & Ezenwa, V. O., 2016 Heterogeneity in pathogen transmission: Mechanisms and methodology. *Functional Ecology* **30**, 1606–1622. (doi: 10.1111/1365-2435.12645).

[39] Volz, E., 2008 SIR dynamics in random networks with heterogeneous connectivity. *Journal of Mathematical Biology* **56**, 293–310. (doi: 10.1007/s00285-007-0116-4).

[40] Wallinga, J. & Lipsitch, M., 2007 How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences* **274**, 599–604.

[41] Wearing, H. J., Rohani, P. & Keeling, M. J., 2005 Appropriate Models for the Management of Infectious Diseases. *PLoS Medicine* **2**, e174. (doi: 10.1371/journal.pmed.0020174).

[42] 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).

[43] Woolhouse, M. E. J., Dye, C., Etard, J.-F., Smith, T., Charlwood, J. D., Garnett, G. P., Hagan, P., Hii, J. L. K., Ndhlovu, P. D., Quinnell, R. J., Watts, C. H., Chandiwana, S. K. & Anderson, R. M., 1997 Heterogeneities in the transmission of infectious agents: Implications for the design of control programs. *Proceedings of the National Academy of Sciences* **94**, 338–342. (doi: 10.1073/pnas.94.1.338).

[44] Yates, A., Antia, R. & Regoes, R. R., 2006 How do pathogen evolution and host heterogeneity interact in disease emergence? *Proceedings of the Royal Society B: Biological Sciences* **273**, 3075–3083. (doi: 10.1098/rspb.2006.3681).