

Different forms of superspreading lead to different outcomes: heterogeneity in infectiousness and contact behavior relevant for the case of SARS-CoV-2

Elise J. Kuylen^{1,2*}, Andrea Torneri¹, Lander Willem¹, Pieter J. K. Libin^{2,3,4}, Steven Abrams^{2,5}, Pietro Coletti², Nicolas Franco^{2,6}, Frederik Verelst¹, Philippe Beutels^{1,7}, Jori Liesenborgs⁸, Niel Hens^{1,2}

1 Centre for Health Economic Research and Modeling Infectious Diseases, University of Antwerp, Antwerp, Belgium

2 Data Science Institute, I-BioStat, Hasselt University, Hasselt, Belgium

3 Artificial Intelligence Lab, Vrije Universiteit Brussel, Brussels, Belgium

4 Rega Institute for Medical Research, Clinical and Epidemiological Virology, University of Leuven, Leuven, Belgium

5 Global Health Institute, University of Antwerp, Antwerp, Belgium

6 Namur Institute for Complex Systems, Department of Mathematics, University of Namur, Namur, Belgium

7 School of Public Health and Community Medicine, The University of New South Wales, Sydney, NSW, Australia

8 Expertise Centre for Digital Media, Hasselt University - transnational University Limburg, Hasselt, Belgium

* elise.kuylen@uantwerp.be

Abstract

Superspreading events play an important role in the spread of SARS-CoV-2 and several other pathogens. Hence, while the basic reproduction number of the original Wuhan SARS-CoV-2 is estimated to be about 3 for Belgium, there is substantial inter-individual variation in the number of secondary cases each infected individual causes. Multiple factors contribute to the occurrence of superspreading events: heterogeneity in infectiousness and susceptibility, variations in contact behavior, and the environment in which transmission takes place. While superspreading has been included in several infectious disease transmission models, our understanding of the effect that these different forms of superspreading have on the spread of pathogens and the effectiveness of control measures remains limited. To disentangle the effects of infectiousness-related heterogeneity on the one hand and contact-related heterogeneity on the other, we implemented both forms of superspreading in an individual-based model describing the transmission and spread of SARS-CoV-2 in the Belgian population. We considered its impact on viral spread as well as on the effectiveness of social distancing. We found that the effects of superspreading driven by heterogeneity in infectiousness are very different from the effects of superspreading driven by heterogeneity in contact behavior. On the one hand, a higher level of infectiousness-related heterogeneity results in less outbreaks occurring following the introduction of one infected individual. Outbreaks were also slower, with a lower peak which occurred at a later point in time, and a lower herd immunity threshold. Finally, the risk of resurgence of an outbreak following a period of lockdown decreased. On the other hand, when contact-related heterogeneity was high, this also led to smaller final sizes, but caused outbreaks to be more explosive in regard

to other aspects (such as higher peaks which occurred earlier, and a higher herd immunity threshold). Finally, the risk of resurgence of an outbreak following a period of lockdown increased. Determining the contribution of both source of heterogeneity is therefore important but left to be explored further.

Author summary

To investigate the effect of different sources of superspreading on disease dynamics, we implemented superspreading driven by heterogeneity in infectiousness and heterogeneity in contact behavior into an individual-based model for the transmission of SARS-CoV-2 in the Belgian population. We compared the impact of both forms of superspreading in a scenario without interventions as well as in a scenario in which a period of strict social distancing (i.e. a lockdown) is followed by a period of partial release. We found that both forms of superspreading have very different effects. On the one hand, increasing the level of infectiousness-related heterogeneity led to less outbreaks being observed following the introduction of one infected individual in the population. Furthermore, final outbreak sizes decreased, and outbreaks became slower, with lower and later peaks, and a lower herd immunity threshold. Finally, the risk for resurgence of an outbreak following a period of lockdown also decreased. On the other hand, when contact-related heterogeneity was high, this also led to smaller final sizes, but caused outbreaks to be more explosive regarding other aspects (such as higher peaks that occurred earlier). The herd immunity threshold also increased, as did the risk of resurgence of outbreaks.

Introduction

As of December 2021, the SARS-CoV-2 pandemic has led to over 300 million confirmed cases and more than 5 million confirmed deaths worldwide [1]. Mathematical modeling has been instrumental in understanding transmission dynamics, as well as in evaluating the impact of both pharmaceutical and non-pharmaceutical interventions [2–9]. A large number of different models were developed to account for the multitude of factors that were found to be important for the spread and control of SARS-CoV-2, including age [10, 11], seasonality [12, 13], and superspreading [14–20].

Through the analysis of contact tracing data and the reconstruction of transmission clusters, superspreading events have been shown to be a driving factor in the spread of several pathogens, among which are SARS-CoV-1, MERS, and, more recently, SARS-CoV-2 [21–27]. This means that the number of secondary cases caused by a single infectious individual is subject to substantial inter-individual variation. In other words, a small number of infected persons generates the majority of new infections, while most infected individuals cause only very few to no secondary cases.

In 2005, Lloyd-Smith et al. [28] proposed a framework to study superspreading dynamics in which the expected number of secondary cases caused by an infected individual, i.e., the individual reproduction number, is represented by a random variable, following a distribution – Lloyd-Smith et al. use a Negative Binomial distribution – on the positive real axis [28, 29]. As such, superspreading events can be characterized as occurrences in the right-hand tail of the aforementioned distribution. When such a superspreading event occurs, different factors, such as for example heterogeneity in infectiousness or heterogeneity in contact behavior, are at play [30]. Increased infectiousness – defined here as the likelihood that a social contact between an infectious and a susceptible individual leads to transmission – either due to behavioral, physical or biological reasons, may play a role. For example, more virus particles are shed when talking loudly or singing [31]. Furthermore, given the importance of aerosol transmission

for the spread of SARS-CoV-2, differences in immune response among individuals may explain why some persons experience a more virulent infection than others, thereby excreting more virus particles and producing more secondary cases [32–34]. Additionally, timing is also important, i.e., individuals infected with SARS-CoV-2 are most infectious during a short interval [35], a period which is not necessarily accompanied by COVID-19 symptoms and associated behavioural change [36].

Some individuals have a higher number of contacts, and thus more opportunity to infect others, which is especially important during the short interval in which they are most infectious [37]. Some persons might not have a higher number of total contacts, but meet more susceptible individuals than others – such as for example in nursing homes [38]. Finally, the environment also plays an important role in the genesis of superspreading events: the risk of transmission is much higher in enclosed spaces than it is outside, and ventilation works well to limit transmission indoors [38, 39].

Some transmission models for SARS-CoV-2 [14–20, 40], as well as for other pathogens like SARS-CoV-1 [41–43] and MERS [44], have taken superspreading into account. These approaches include compartmental models [15, 16], that divide the population in different and exclusive subpopulations based on realistic disease states, and branching process models [14, 28], as well as network and individual-based models [17–20, 40–44], in which each individual in the population is represented as a separate entity. Some models include the superspreading potential of individuals as a general factor [14, 17, 28], while others implement superspreading events as the result of heterogeneity in infectiousness [16], contact behavior [18, 19, 43], or a combination of both [15, 20, 41, 42, 44].

However, while some of these models do demonstrate that both factors contribute to the occurrence of superspreading events [20], it is still poorly understood how exactly these different forms of superspreading impact the spread of disease and the effectiveness of control measures. To further investigate these questions in detail, we tested the effect of different forms of superspreading on the spread of SARS-CoV-2 in the Belgian population. To this end, we implemented both infectiousness-related and contact-related heterogeneity into STRIDE, an individual-based model for the transmission of SARS-CoV-2 [9, 45].

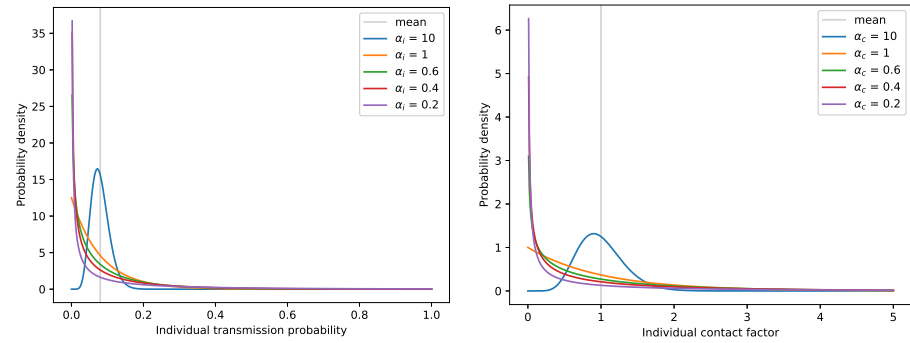
In an individual-based model, each individual is represented by a separate entity with a unique set of characteristics, such as age, health status, and behavioral traits. As such, individual-based models are particularly suited to model superspreading, as they allow for a direct integration of different sources of heterogeneity at the individual level.

We investigated the effects of superspreading caused by variation in infectiousness versus heterogeneity in contact behaviour, in the absence of intervention measures. Additionally, we looked at the impact of different modes of superspreading on the effectiveness of social distancing – a non-pharmaceutical intervention which continues to play an important role in the control of SARS-CoV-2.

Materials and methods

Implementation

We used STRIDE [45], an individual-based, stochastic model, which was recently adapted to encompass the disease-specific features related to the transmission of SARS-CoV-2 in the Belgian population [9]. In this model, individuals are assigned to social contact pools representing their household, a workplace or school (depending on the individual’s age and employment status), and more general communities representing an individual’s leisure contacts, contacts on public transportation and contacts in other locations. These latter communities consist of on average 500 persons



(a) Right-truncated gamma distribution considered for the individual transmission probability, with different values for the shape parameter α_i and mean = 0.08. (b) Gamma distribution considered for the individual contact factor, with different values for the shape parameter α_c and mean = 1.

Fig 1. Probability density function of the Gamma distributions considered for the individual transmission probability (panel a) and the individual contact factor (panel b).

and differ for each person between week- and weekend days. 77

Simulations move forward in discrete time-steps of one day. Each day, individuals can access the pools of which they are a member, depending on the day of the week and their health status – e.g., children do not visit their school pool during the weekend, and symptomatic individuals primarily remain within their household pool. The number of contacts an individual makes each day in each of the above locations is based on their age and the type of location (household, school, workplace, or community). More specifically, daily contact rates are based on a social contact survey conducted in Belgium in 2010-2011 [46]. 78
79
80
81
82
83
84
85

We adapted the model to represent both infectiousness-related and contact-related heterogeneity. Infectiousness-related heterogeneity was implemented as follows. When a person becomes infected, an ‘individual transmission probability’ is assigned to this person. This probability determines whether an actual transmission event takes place whenever a contact occurs between the infected individual and a susceptible individual. We assume that this probability remains the same for the entire duration of an individual’s infectious period – one of the limitations we elaborate on in the discussion. 86
87
88
89
90
91
92

Following the work of Lloyd-Smith et al. [28], we chose to represent inter-individual variation by using a Gamma distribution. As such, each individual transmission probability is drawn from a (right-)truncated Gamma distribution on the interval (0, 1]. The truncated distribution is characterized by a shape parameter α_i , which determines the level of overdispersion of the distribution, and a mean – hereafter referred to as the mean transmission probability. A lower value of α_i entails more variation in the individual transmission probability, and thus a higher level of superspreading, as shown in Figure 1a. 93
94
95
96
97
98
99
100

To account for contact-related heterogeneity, we multiply an individual’s contact rate in community and workplace pools by a factor, which is unique for each individual. This factor is drawn from a Gamma distribution with shape parameter α_c and mean 1, upon creation of the individual. Again, a lower value of α_c implies more variation in contact rates and thus a higher level of superspreading, which is shown in Figure 1b. 101
102
103
104
105

The code of STRIDE is open-source and available at <https://github.com/lwillem/stride> and <https://github.com/elisekaa/stride>. All code used for the research in this paper will be made available on Zenodo upon acceptance of the paper. 106
107
108
109

Simulations

For an exhaustive overview of the population and parameters used in the simulations described below, we refer to Supplementary Material S1 and earlier work conducted on SARS-CoV-2 transmission using the STRIDE simulator [9]. As there is no standard procedure to estimate the number of realisations necessary for the type of simulations we conduct, we chose to run 200 stochastic simulations for each scenario described below, which led to stable results.

Verification

We verified our implementation of superspreading in STRIDE in a number of ways. First, we confirmed that the mean R_0 was consistent between the different scenarios regarding infectiousness-related and contacts-related heterogeneity.

We also checked that the baseline case – in which both the individual transmission probability and the individual contact factor are constant – corresponds to a scenario in which either the individual transmission probability or the individual contact probability follows a Gamma distribution with respectively $\alpha_i = 10$ and $\alpha_c = 10$. This should be the case, since, as the overdispersion parameter approaches infinity, the distribution becomes more centered around the mean.

Furthermore, we calculated P_{80} : the minimal proportion of infected individuals that is responsible for 80% of cases – a measure commonly used in the superspreading literature [22, 28, 29]. We calculated P_{80} for each simulation by ordering individuals that were infectious during the simulation according to the number of secondary cases they caused, in decreasing order. Next, we compute the proportion of individuals that are responsible for 80% of the total number of infections caused over the entire simulation run. We used this measure to verify that as we decreased either α_i or α_c , this resulted in more heterogeneity in transmission – leading in turn to a lower P_{80} .

Finally, we constructed a theoretical description for the transmission process in STRIDE, which allowed us to calculate the expected number of secondary cases per infected individual, as well as the variance of this quantity. We then compared the results of this theoretical description with simulation results for different forms and levels of superspreading.

For more details on the way in which these checks were conducted, and their results, see Supplementary Material S2.

Superspreading effects in the absence of interventions

First, we investigated the effect of different modes of superspreading on the unmitigated spread of SARS-CoV-2. To do this, we introduced one infected individual in a completely susceptible population at the beginning of a simulation and tracked transmission events over a period of 200 days. Typically, after this period, no more new infections were recorded for all considered scenarios, so we assumed we had observed most, if not all, of the epidemic curve. This is apparent from the plots depicting the evolution of the number of new cases and the cumulative number of cases per day, as shown in Supplementary Figures S23-S26.

To compare the effect of infectiousness-related heterogeneity to the effect of contact-related heterogeneity, we varied the distribution of the individual transmission probability and the individual contact factor as described in Table 1. The values for α_i and α_c were chosen as follows: the value 10 represents a distribution in line with the baseline scenario (i.e., a distribution which approaches a degenerate one), with low levels of heterogeneity for either the individual transmission probability or the individual contact factor, a value of 1 was chosen as an intermediate scenario, and

Scenario	Individual transmission probability	Individual contact factor
Baseline	0.08	1.00
A	Trunc. Gamma ($\alpha_i = 10.0$, mean = 0.08)	1.00
B	Trunc. Gamma ($\alpha_i = 1.0$, mean = 0.08)	1.00
C	Trunc. Gamma ($\alpha_i = 0.6$, mean = 0.08)	1.00
D	Trunc. Gamma ($\alpha_i = 0.4$, mean = 0.08)	1.00
E	Trunc. Gamma ($\alpha_i = 0.2$, mean = 0.08)	1.00
F	0.08	Gamma ($\alpha_c = 10.0$, mean = 1.0)
G	0.08	Gamma ($\alpha_c = 1.0$, mean = 1.0)
H	0.08	Gamma ($\alpha_c = 0.6$, mean = 1.0)
I	0.08	Gamma ($\alpha_c = 0.4$, mean = 1.0)
J	0.08	Gamma ($\alpha_c = 0.2$, mean = 1.0)

Table 1. Overview of scenarios based on the distribution considered for the individual transmission probability and the individual contact factor.

values 0.6, 0.4 and 0.2 were chosen to represent a high level of heterogeneity, in line with what we would expect to see in the transmission of SARS-CoV-2 [29]. The mean of the truncated Gamma distribution considered for the individual transmission probability (i.e., 0.08) was chosen so that it corresponds to an R_0 value of about 3 in the baseline scenario. This is in line with initial estimates for R_0 for SARS-CoV-2 [3, 9, 47, 48].

We then compared epidemiological metrics between the different scenarios. First, we looked at the probability of extinction, calculated as the fraction of simulation runs that produce no more new cases after only a few (or no) initial secondary infections [49]. We define what is regarded as extinction by looking at the final outbreak sizes after 200 simulation days, resulting from all simulations over all scenarios. After ordering these in descending order, a sharp drop-off can be observed, separating runs with persistent outbreaks from runs in which extinction occurs. As we found that outbreaks starting from one initial case either remain below 15 cases, or grow much larger, we set the threshold below which we will consider an outbreak to have gone extinct at 20 cases. In Supplementary Figure S22, a histogram is shown depicting the frequency of final outbreak sizes per scenario.

We also compared the attack rate of outbreaks, and looked at the peak size and the timing of the peak. Furthermore, we investigated the evolution of the daily effective reproduction number R_t . We approximated R_t by calculating the mean number of secondary cases caused by individuals that contracted infection on day t . However, since at the end of a simulation only a few infected individuals remain, thereby making inference about R_t prone to substantial oscillations in daily estimates, we applied a LOWESS smoothing approach for the time-varying R_t -values [50].

To further gauge the effect of different forms of superspreading on epidemic spread, we estimated the herd immunity threshold, the day this threshold is reached, and the day on which the last infection is recorded. In the absence of a standard procedure to define the herd immunity threshold in an individual-based model, we estimated this measure as follows. We looked at the proportion of individuals that were no longer susceptible (i.e., recovered or currently infected) on the last day for which the smoothed $R_t \geq 1$.

Finally, we kept track of the type of location (household, school, workplace or community) in which infections occurred.

Superspreading effects in the presence of social distancing

We considered a scenario to investigate the impact of different forms of superspreading on the effectiveness of a period with strong social distancing followed by a period of

mild relaxations. 193

The scenario was implemented as follows: after the introduction of a single infected 194
individual in the population we simulated 30 days without interventions, after which a 195
'lockdown' period began, in which schools were closed (primary, secondary, and tertiary 196
education), contacts at the workplace were reduced by 94.51% and contacts in the 197
communities were reduced by 88.74%. These contact reductions were inferred based on 198
social contact data collected for Belgium from April to mid-May 2020 in the CoMix 199
study [51]. 200

After 60 days of lockdown (i.e. on day 90 of the simulation), a partial release of the 201
lockdown followed. Schools were re-opened and contacts in the workplace increased to 202
25.09% of pre-pandemic levels, while contacts in the community increased to 28.55% of 203
pre-pandemic levels. Again, these contact reductions were estimated based on data 204
collected in the CoMix study, from mid-May to August 2020. More information on how 205
these and the above estimates were obtained can be found in Supplementary 206
Material S1.2. 207

We tested the same distributions for the individual contact probability and for the 208
individual contact factor as listed in Table 1. We ran each simulation for 600 days (30 209
days pre-lockdown, 60 days lockdown and 510 days partial release), after which we 210
typically did not observe any new infections. This can be seen in the plots in 211
Supplementary Figures S35–S38, showing the number of new cases and the number of 212
cumulative cases per day. 213

We compared the number of cases before, during, and after the lockdown between 214
the different scenarios, as well as the attack rate over the entire simulation. We also 215
looked at the evolution of the number of new cases, the cumulative number of cases and 216
the effective R_t per day. Additionally, we calculated a 'resurgence probability', to 217
represent the chance that a lockdown followed by a partial release would not be effective 218
in stopping the epidemic in a particular scenario. This resurgence probability was 219
calculated as follows. As the effectiveness of a lockdown can only be gauged when the 220
epidemic is still ongoing, and we did not consider the importation of new cases, we only 221
took into account those runs in which new cases were still being observed when the 222
lockdown period started. As such, we excluded simulation runs in which 0 cases were 223
recorded during the entire lockdown period. Then, to define resurgence, we looked at 224
the distribution of the number of cases during the period of partial release. 225

For all scenarios, we observed that either at most 309 cases occurred during the 226
period of partial release, or a much larger number. As such we set the 'resurgence 227
threshold' at 500 cases. In Supplementary Figure S33, a histogram is shown depicting 228
the frequency of numbers of cases during the release period per scenario. 229

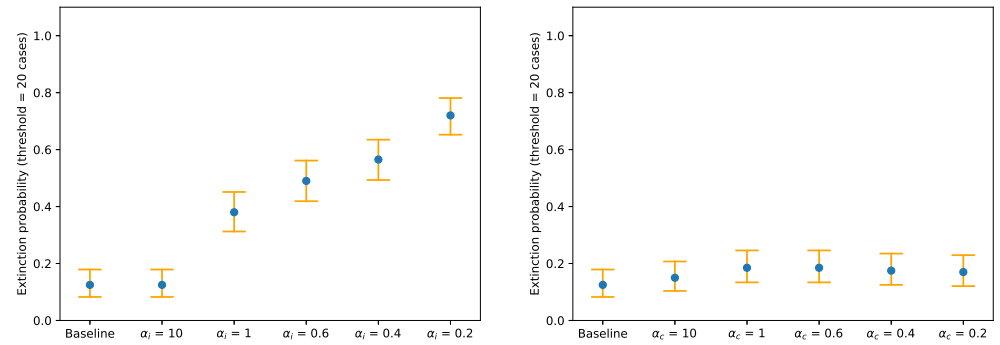
Sensitivity analysis 230

We conducted a sensitivity analysis regarding the mean transmission probability and 231
the number of infected cases introduced at the beginning of the simulation. More 232
information on this can be found in Supplementary Material S3. 233

Results 234

Superspreading effects in the absence of interventions 235

We investigated the spread of SARS-CoV-2 throughout the Belgian population, 236
following the introduction of one infected individual. First, we looked at the probability 237
of extinction. Using the extinction threshold of 20 cases, we calculated, for each 238
scenario, the proportion of simulation runs that lead to extinction. We observe that as 239



(a) Varying α_i for the Truncated Gamma distribution considered for the individual transmission probability.

(b) Varying α_c for the Gamma distribution considered for the individual contact factor.

Fig 2. Extinction probabilities for scenarios investigating infectious-related heterogeneity (panel a) and contact-related heterogeneity (panel b), defined as the proportion of simulation runs ($n = 200$) that produces less than 20 cases. Orange error bars represent 95% (Clopper-Pearson) confidence intervals.

infectiousness-related heterogeneity increases, the extinction probability also increases (see Figure 2a). In the baseline scenario, where the transmission probability is the same for all individuals, extinction occurs in 12.5% of all simulation runs. As α_i decreases to 0.2, however, we see that as much as 72% of the simulation runs produce less than 20 cases.

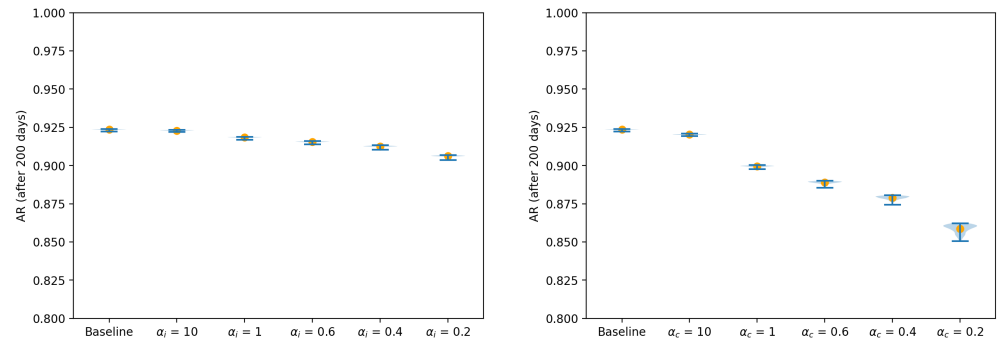
A different trend can be observed when varying contact-related heterogeneity (see Figure 2b). Here, varying α_c has only little effect on the fraction of runs that lead to extinction. While initially, with an intermediate value for α_c of 1, the extinction probability shows a limited increase, a slight downward trend can be noticed when α_c decreases from 0.6 to 0.2.

We also looked at the attack rate of outbreaks. In the baseline scenario, around 92.5% of the population is infected 200 days after the introduction of an infected individual in the population, whenever extinction did not occur. When increasing infectiousness-related heterogeneity as well as when increasing contact-related heterogeneity, the final size of outbreaks lowers. However, this decrease is much more pronounced in the case of contact heterogeneity. This is shown in Figure 3.

When looking at the size and timing of the peak of the epidemic curve of outbreaks when no extinction occurs (see Figure 4), opposite trends can be observed for increasing heterogeneity in infectiousness and contacts, respectively. When infectiousness-related heterogeneity increases, the mean size of the peak decreases, while no effect is seen on the time when the peak occurs. However, when contact-related heterogeneity increases, the mean peak size becomes higher and also occurs earlier. Furthermore, contact-related heterogeneity leads to more variability in peak size.

These trends can also be observed when looking at the evolution of the number of new cases, the number of cumulative cases and the smoothed daily R_t values. The number of new cases per day, the cumulative number of cases per day, and the smoothed R_t per day for the different scenarios is shown in Supplementary Figures S23–S28.

We estimated the impact of different modes of superspreading on the herd immunity threshold (see Figure 5). We observed that when infectiousness-related heterogeneity increases as α_i decreases from 10 to 0.2, the average herd immunity threshold decreases from 58.97% to 32.66%. This strong decline is in part due to the increase in the number of runs in which extinction occurs, but is still present when these runs are excluded, as



(a) Varying α_i for the Truncated Gamma distribution considered for the individual transmission probability.

(b) Varying α_c for the Gamma distribution considered for the individual contact factor.

Fig 3. Violin plots for the attack rate over 200 days for scenarios investigating the infectiousness-related heterogeneity (panel a) and contact-related heterogeneity (panel b). The orange dots represent the mean attack rate across the simulation runs without extinction, i.e., simulation runs in which extinction occurs (< 20 cases) were excluded.

can be seen in Figure 5c.

However, when contact-related heterogeneity increases, an opposite trend can be observed: the herd immunity threshold increases to 64.28% when α_c decreases to 0.2.

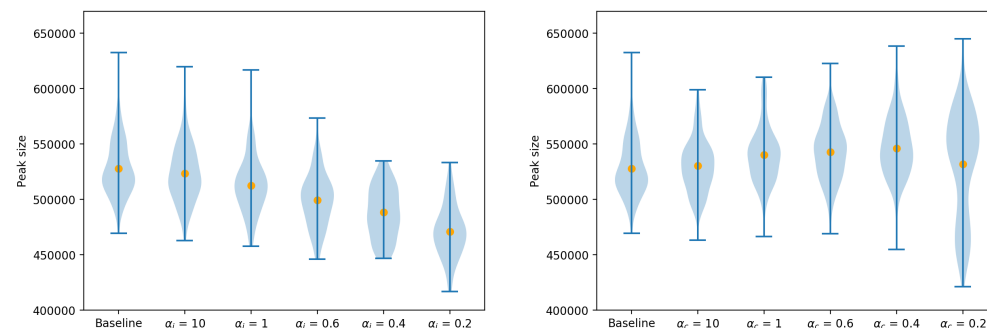
We also looked at the day on which the herd immunity threshold is reached (see Supplementary Figure S29), and the day on which the last transmission event is observed (see Supplementary Figure S30). We observed that, with higher infectiousness-related heterogeneity, the epidemic slows down, as the herd immunity threshold is reached later – even though it is lower – and the last infections are observed at a later time compared to the baseline scenario. Conversely, with higher contact-related heterogeneity, the herd immunity threshold is reached faster and infections stop occurring at an earlier point in the simulation.

Finally, we looked at the type of locations in which transmissions occurred (see Supplementary Figures S31–S32). We found that neither increasing heterogeneity in infectiousness nor increasing heterogeneity in contact behavior changed the locations where most infections took place, which for all scenarios were communities and households.

Superspreading effects in the presence of social distancing

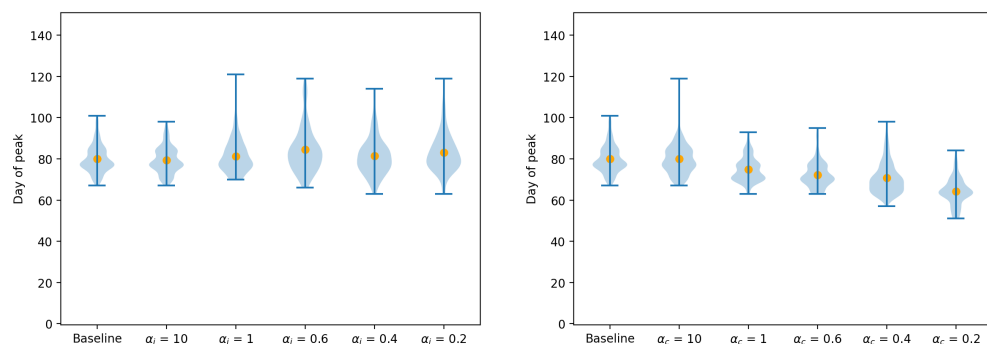
We investigated the effect of different forms of superspreading on the outcome of a social distancing scenario. We compared the number of cases before lockdown (day 0 – day 30, see Figure 6a–b), during lockdown (day 30 – day 90, see Figure 6c–d) and during the period of partial release (day 90 – day 600, see Figure 6e–f) between the different scenarios.

The number of cases observed before lockdown confirmed what we observed when examining unmitigated transmission: when contact-related heterogeneity is high, the start of outbreaks is more explosive. On the contrary, when infectiousness-related heterogeneity is high, many outbreaks stop after only a few cases. During lockdown, we also observe more cases when there is higher contact-related heterogeneity. However, as infectiousness-related heterogeneity increases, the average number of cases observed during lockdown slightly decreases. Finally, the number of cases during the partial release phase decreases sharply as infectiousness-related heterogeneity increases, while



(a) Size of peak when varying α_i for the Truncated Gamma distribution considered for the individual transmission probability.

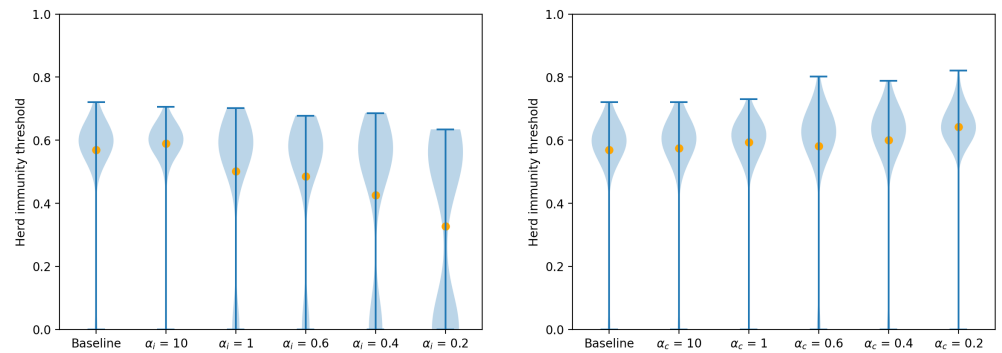
(b) Size of peak when varying α_c for the Gamma distribution considered for the individual contact factor.



(c) Day of peak when varying α_i for the Truncated Gamma distribution considered for the individual transmission probability.

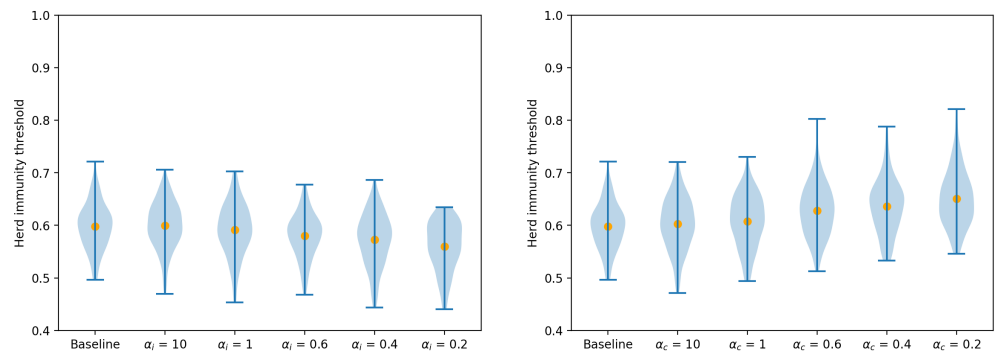
(d) Day of peak when varying α_c for the Gamma distribution considered for the individual contact factor.

Fig 4. Violin plots for the size and day of peak across the different simulation runs for scenarios investigating the infectiousness-related heterogeneity (panels a and c, respectively) and contact-related heterogeneity (panels b and d, respectively). The orange dots represent the mean of the simulated peak sizes and days at which the peak is reached. Simulation runs in which extinction occurs (< 20 cases) were excluded.



(a) Varying α_i for the Truncated Gamma distribution considered for the individual transmission probability.

(b) Varying α_c for the Gamma distribution considered for the individual contact factor.



(c) Varying α_i for the Truncated Gamma distribution considered for the individual transmission probability. Runs in which extinction occurred (< 20 cases) were excluded.

(d) Varying α_c for the Gamma distribution applied to the individual contact factor. Runs in which extinction occurred (< 20 cases) were excluded.

Fig 5. Violin plots of the estimated herd immunity threshold values over the different simulation runs including or excluding runs with extinction for scenarios investigating the infectiousness-related heterogeneity (panels a and c, respectively) and contact-related heterogeneity (panels b and d, respectively). The orange dots represent the mean of the simulated values.

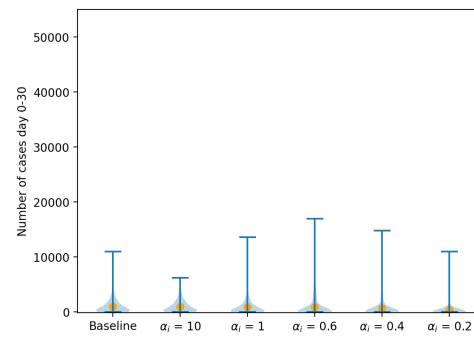
changes in contact-related heterogeneity seem to have little impact on the number of cases during this period. The same trends can be observed when looking at the attack rate over the entire simulation period (see Supplementary Figure S34).

Furthermore, we looked at the evolution of the number of new cases per day, the cumulative number of cases and the smoothed daily R_t values (see Supplementary Figures S35–S40). We observed that with higher infectiousness-related heterogeneity, outbreaks were slower to fade out during the period of partial release. When contact-related heterogeneity was high outbreaks were more explosive and took off again faster once social distancing measures were relaxed. However, what is remarkable here as well is that, as contact-related heterogeneity increases, distinct waves can be observed during the period of partial relaxations (see for example Supplementary Figure XX).

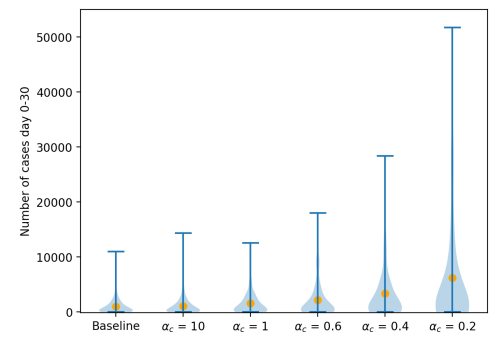
We calculated the resurgence probability for different forms and levels of superspreading (see Figure 7). As infectiousness-related heterogeneity increases, the resurgence probability decreases, meaning that it is less likely that the epidemic grows large in size again when relaxing measures after a period of lockdown. However, resurgence probabilities increase slightly as contact-related heterogeneity increases.

We assessed how the herd immunity threshold changes when social distancing is applied (see Figure 8). For the baseline case, the estimated herd immunity threshold lowered from 56.89% without interventions to 18.18% on average in the social distancing scenario. When increasing infectiousness-related heterogeneity, the herd immunity threshold decreased even further: to 2.28% when $\alpha_i = 0.2$. However, this is not the case when contact-related heterogeneity is high. Even though the average herd immunity threshold initially lowers from 16.94% to 13.89% when α_c decreases from 10 to 0.6, it again increases to 22.43% when α_c is further decreased to 0.2.

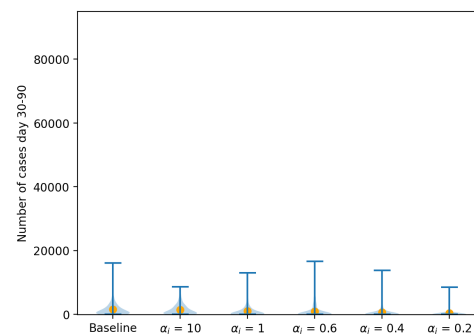
Finally, we looked at the locations in which transmissions occurred (see Supplementary Figures S41–S42). We found that social distancing had little effect on where transmission happened: even though there were less transmissions overall, most transmissions still occurred in households and communities for all scenarios.



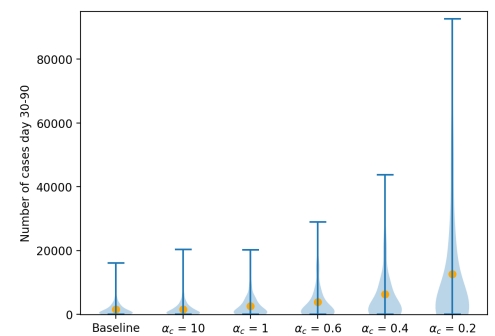
(a) Number of cases before lockdown (day 30), when varying α_i for the Truncated Gamma distribution considered for the individual transmission probability.



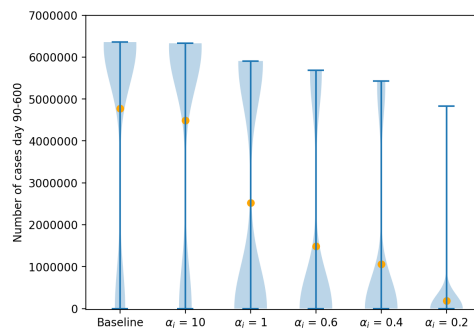
(b) Number of cases before lockdown (day 30), when varying α_c for the Gamma distribution considered for the individual contact factor.



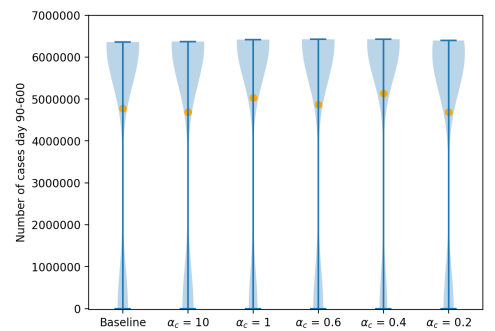
(c) Number of cases during lockdown (day 30–day 90), when varying α_i for the Truncated Gamma distribution considered for the individual transmission probability.



(d) Number of cases during lockdown (day 30–day 90), when varying α_c for the Gamma distribution considered for the individual contact factor.

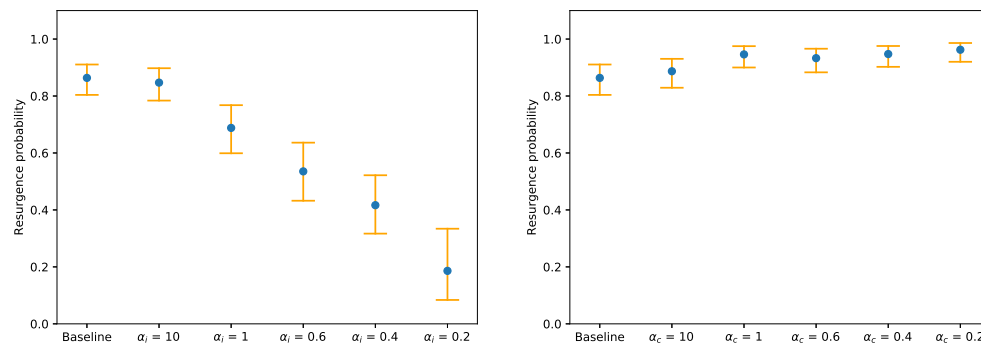


(e) Number of cases during partial release phase (day 90–day 600), when varying α_i for the Truncated Gamma distribution considered for the individual transmission probability.



(f) Number of cases during partial release phase (day 90–day 600), when varying α_c for the Gamma distribution considered for the individual contact factor.

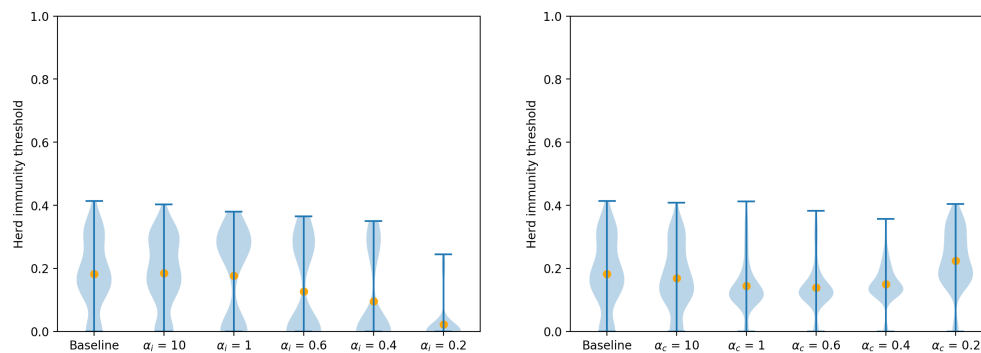
Fig 6. Violin plots for the number of cases before lockdown (day 0–day 30), during lockdown (day 30–day 90) and during partial release phase (day 90–day 600) for the different scenarios tested. The orange dots represent the mean of the simulated values.



(a) Varying α_i for the Truncated Gamma distribution considered for the individual transmission probability.

(b) Varying α_c for the Gamma distribution considered for the individual contact factor.

Fig 7. The proportion of simulation runs, in which the number of cases during lockdown is greater than 0, that produces more than 500 cases during the partial release phase. Orange error bars represent 95% (Clopper-Pearson) confidence intervals.



(a) Varying α_i for the Truncated Gamma distribution considered for the individual transmission probability.

(b) Varying α_c for the Gamma distribution considered for the individual contact factor.

Fig 8. Violin plots for the herd immunity threshold as estimated for the social distancing scenario. The orange dots represent the mean of the simulated values.

Discussion

Conclusion

We investigated different forms of superspreading in an individual-based model by considering both infectiousness-related heterogeneity and contact-related heterogeneity. We found that these two types of superspreading have very different effects, both on the unmitigated spread of SARS-CoV-2 as well as on the effectiveness of social distancing measures. In the absence of containment measures, we observed that with high infectiousness-related heterogeneity, the introduction of an infected individual in the population led to large outbreaks less frequently. Furthermore, peak sizes were smaller and occurred at a later time. The estimated herd immunity threshold also decreased. On the other hand, when contact-related heterogeneity is high, we observed that the introduction of an infected individual led to outbreaks slightly more often compared to the baseline scenario in which the mean transmission probability is 0.08 for all individuals and the individual contact factor is set to 1. Outbreaks that did not go extinct after only a few secondary cases were also more explosive, with higher peaks that occurred earlier. The herd immunity threshold also increased slightly.

The difference between the effects caused by these two forms of superspreading might be explained by the fact that when infectiousness-based heterogeneity is high, superspreaders have to infect other superspreaders to keep the epidemic going – leading to more extinctions and a drawn-out, but less explosive, outbreak whenever extinction does not occur. Furthermore, not all infected individuals with a very high individual transmission probability will have a lot of contacts and thus the opportunity to realize their ‘transmission potential’. However, when a person has a large number of contacts, they not only have more opportunity to infect others once they are infected, but they also have a disproportionately high chance to be exposed to infection themselves [37], leading to faster, more explosive outbreaks.

We also observed that the total attack rate of outbreaks decreased much faster when contact-related heterogeneity increased than when infectiousness-related heterogeneity increased. This seems to be in contrast with the results described above, but can be explained due to a faster depletion of superspreaders (and susceptible individuals in general). This is enhanced by the fact that STRIDE uses a structured population, in which individuals are limited in the contacts they can make by the constraints of the contact pools (household, school, workplace, communities) they belong to. As such, when a large fraction of the persons in their contact pools has already been infected, an infected individual cannot cause a large number of secondary cases, no matter how infectious they are, or how many contacts they have [52]. Furthermore, as contact-related heterogeneity increases, more individuals will have no or very few contacts, making them ‘unreachable’, thereby decreasing the maximum number of cases that can be generated during an outbreak.

We should also note here that the same α_i and α_c parameters lead to different levels of overdispersion in transmissions (as shown in Supplementary Figures S1–S2) – which might explain some of the differences between both forms of superspreading.

Other models [16, 42, 44] that implement both heterogeneity in infectiousness and heterogeneity in contacts also conclude that superspreading driven by contact-related heterogeneity leads to more explosive outbreaks. These models (and others that only consider infectiousness-related heterogeneity), however, do not capture the slowdown of outbreaks due to infectiousness-related heterogeneity, since they do not implement a structured population, and hence every highly infectious person is able to realize their ‘transmission potential’.

The current SARS-CoV-2 pandemic seems to exhibit characteristics of both these modes of superspreading. It is likely that a combination of factors is responsible for the

occurrence of SARS-CoV-2 superspreading events, such as a moderate heterogeneity in infectiousness combined with a high heterogeneity in contact behavior [15, 20].

We also investigated the impact of different forms of superspreading on the effectiveness of social distancing. We found that, when superspreading is driven by heterogeneity in infectiousness, a period of strict social distancing, followed by a partial release is most effective and extinguishes almost all outbreaks. This is in line with the conclusion of other models that also assume superspreading events are driven by heterogeneity in infectiousness [18, 19].

However, when superspreading is driven by heterogeneity in contact behavior, we found that while social distancing measures might limit cases during a strict lockdown, the chance of resurgence of the epidemic following relaxations increases. Furthermore, with high infectiousness-related heterogeneity, few outbreaks grow large in size after a period of lockdown, but outbreaks that do not go extinct linger for a long time before completely disappearing.

Limitations

Some limitations need to be taken into account when interpreting these results.

Each individual that is infected over the course of the simulation, eventually recovers and gains immunity for the remainder of the simulation. As such, neither deaths nor re-infections occur. We also did not model the waning of immunity, nor vaccination, making these results more representative of the initial wave of the SARS-CoV-2 pandemic. Additionally, we only introduce infected individuals at the beginning of the simulation, and do not take into account the continuous importation of infectious individuals into the population. Furthermore, we focus on the initial course of the infections, in which individuals can transmit the disease. We did not model hospitalizations with case isolation, since we assume the transmission dynamics are mostly shaped by the pre-symptomatic and mild symptomatic stages and individuals exhibit adapted social contact patterns once they become symptomatic. To fully capture the effect of superspreading on the spread of SARS-CoV-2, these factors would, however, need to be taken into account.

We also did not explicitly model mass events, which may be an important factor in the occurrence of superspreading events [25]. We did however model two different community pools (for weekdays and weekend days, respectively), where individuals could potentially contact a larger number (up to 500) of people, which could serve as a proxy for attending a large event. However, large events also have a temporal aspect that is not modeled through community contact pools. As such, explicitly modeling mass gatherings might lead to new insights regarding the effect of different types of superspreading on disease spread and on the effectiveness of interventions.

To represent heterogeneity in infectiousness and contact behavior, we used a right-truncated and untruncated Gamma distribution respectively. However, other distributions could be used to obtain similar levels of heterogeneity, which we did not test in the current study [29].

Finally, next to infectiousness-related heterogeneity and contact-related heterogeneity, other factors, such as heterogeneity in susceptibility and the environment (such as indoor venues) in which contacts occur, also contribute to the occurrence of superspreading events [53–55]. Additionally, infectiousness may vary over time for the same individual, while, in our model, it remains the same for the entire infectious period of an individual [35].

Future work

Several extensions of this work could be useful in furthering our understanding of superspreading both within the context of the current COVID-19 pandemic and within a broader context.

Further research is needed to disentangle effects of both forms of superspreading. As such, it would be possible to estimate the level of both infectiousness-related and contact-related heterogeneity at play in the current SARS-CoV-2 pandemic. We plan to use data collected during the CoMix study [51] to estimate the level of contact-related heterogeneity during different phases of the SARS-CoV-2 pandemic. Subsequently, assuming that the two forms of heterogeneity investigated in this paper are most important for the occurrence of superspreading events, it is possible to estimate the level of infectiousness-related heterogeneity from the overall observed heterogeneity in transmissions. Further insight could also be gained by comparing the importance of these two types of superspreading for the spread of different pathogens, as contact-related heterogeneity presumably remains largely the same for a population under the spread of different pathogens.

We implemented heterogeneity in contact behavior by applying a distribution to the daily contact rate of individuals. There is however, in the context of social distancing, another possible type of contact heterogeneity, namely non-compliance to social distancing measures. It would be relevant to investigate how different forms of superspreading modulate the negative effect of this phenomenon on the effectiveness of social distancing measures.

Furthermore, while we investigated the impact of superspreading on the effectiveness of social distancing, it is conceivable that superspreading also has an impact on other interventions, both pharmaceutical and non-pharmaceutical, such as vaccination, contact tracing and universal testing [14, 56].

Finally, we did not take into account person characteristics that might make an individual more or less likely to be a superspreader, due to the lack of data about this subject. Instead, the individual transmission probability and individual contact factor were drawn at random from a distribution. However, some characteristics, such as age, profession, or perception of the severity of COVID-19, might have an impact on how likely an individual is to transmit the disease [17]. Taking such characteristics into account when distributing transmission potential in the population would provide a more accurate model of superspreading.

References

1. Dong E, Du H, L G. Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*. 2020;20(5):533–534.
2. Abrams S, Wambua J, Santermans E, Willem L, Kuylen E, Coletti P, et al. Modelling the early phase of the Belgian COVID-19 epidemic using a stochastic compartmental model and studying its implied future trajectories. *Epidemics*. 2021;35:100449.
3. Coletti P, Libin P, Petrof O, Willem L, Abrams S, Herzog SA, et al. A data-driven metapopulation model for the Belgian COVID-19 epidemic: assessing the impact of lockdown and exit strategies. *BMC Infectious Diseases*. 2021;21(1):503.
4. Franco N. COVID-19 Belgium: Extended SEIR-QD model with nursing homes and long-term scenarios-based forecasts. *Epidemics*. 2021;37:100490.

5. Kerr CC, Stuart RM, Mistry D, Abeyseriya RG, Rosenfeld K, Hart GR, et al. Covasim: An agent-based model of COVID-19 dynamics and interventions. *PLOS Computational Biology*. 2021;17(7):1–32.
6. Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The Lancet Infectious Diseases*. 2020;20(5):553–558.
7. Moore S, Hill EM, Dyson L, Tildesley MJ, Keeling MJ. Modelling optimal vaccination strategy for SARS-CoV-2 in the UK. *PLoS computational biology*. 2021;17(5):e1008849.
8. Torneri A, Libin P, Vanderlocht J, Vandamme AM, Neyts J, Hens N. A prospect on the use of antiviral drugs to control local outbreaks of COVID-19. *BMC medicine*. 2020;18(1):1–9.
9. Willem L, Abrams S, Libin PJ, Coletti P, Kuylen E, Petrof O, et al. The impact of contact tracing and household bubbles on deconfinement strategies for COVID-19. *Nature Communications*. 2021;12(1):1–9.
10. Britton T, Ball F, Trapman P. A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. *Science*. 2020;369(6505):846–849.
11. Davies NG, Klepac P, Liu Y, Prem K, Jit M, working group CC, et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nature Medicine*. 2020;26:1205–1211.
12. Baker RE, Yang W, Vecchi GA, Metcalf CJE, Grenfell BT. Susceptible supply limits the role of climate in the early SARS-CoV-2 pandemic. *Science*. 2020;369(6501):315–319.
13. Merow C, Urban MC. Seasonality and uncertainty in global COVID-19 growth rates. *Proceedings of the National Academy of Sciences*. 2020;117(44):27456–27464.
14. Endo A, for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group C, Leclerc QJ, Knight GM, Medley GF, Atkins KE, et al. Implication of backward contact tracing in the presence of overdispersed transmission in COVID-19 outbreaks. *Wellcome Open Research*. 2020;5(239).
15. Goyal A, Reeves DB, Cardozo-Ojeda EF, Schiffer JT, Mayer BT. Viral load and contact heterogeneity predict SARS-CoV-2 transmission and super-spreading events. *eLife*. 2021;10:e63537.
16. Kochanczyk M, Grabowski F, Lipniacki T. Super-spreading events initiated the exponential growth phase of COVID-19 with R_0 higher than initially estimated. *Royal Society Open Science*. 2020;6(9):200786.
17. Lau MSY, Grenfell B, Thomas M, Bryan M, Nelson K, Lopman B. Characterizing superspreading events and age-specific infectiousness of SARS-CoV-2 transmission in Georgia, USA. *Proceedings of the National Academy of Sciences*. 2020;117(36):22430–22435.
18. Nielsen BF, Simonsen L, Sneppen K. COVID-19 superspreading suggests mitigation by social network modulation. *Physical Review Letters*. 2021;126(11):118301.

19. Sneppen K, Nielsen BF, Taylor RJ, Simonsen L. Overdispersion in COVID-19 increases the effectiveness of limiting nonrepetitive contacts for transmission control. *Proceedings of the National Academy of Sciences*. 2021;118(14).
20. Susswein Z, Bansal S. Characterizing superspreading of SARS-CoV-2 : from mechanism to measurement. *medRxiv*. 2020;.
21. Adam DC, Wu P, Wong JY, Lau EH, Tsang TK, Cauchemez S, et al. Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nature Medicine*. 2020;26(11):1714–1719.
22. Endo A, for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group C, S A, Kucharski A, Funk S. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China [version 3; peer review: 2 approved]. *Wellcome Open Res*. 2020;5(67).
23. Kucharski AJ, Althaus CL. The role of superspreading in Middle East respiratory syndrome coronavirus (MERS-CoV) transmission. *Euro Surveillance*. 2015;20(25):14–18.
24. Laxminarayan R, Wahl B, Dudala SR, Gopal K, B CM, Neelima S, et al. Epidemiology and transmission dynamics of COVID-19 in two Indian states. *Science*. 2020;370(6517):691–697.
25. Lemieux JE, Siddle KJ, Shaw BM, Loreth C, Schaffner SF, Gladden-Young A, et al. Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of superspreading events. *Science*. 2021;371(6529).
26. Nishiura H, Endo A, Saitoh M, Kinoshita R, Ueno R, Nakaoka S, et al. Identifying determinants of heterogeneous transmission dynamics of the Middle East respiratory syndrome (MERS) outbreak in the Republic of Korea, 2015: a retrospective epidemiological analysis. *BMJ Open*. 2016;6(2).
27. Shen Z, Ning F, Zhou W, He X, Lin C, Chin DP, et al. Superspreading SARS events, Beijing, 2003. *Emerging infectious diseases*. 2004;10(2):256–260.
28. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. *Nature*. 2005;438(17):355–359.
29. Kremer C, Torneri A, Boesmans S, Meuwissen H, Verdonchot S, Vanden Driessche K, et al. Quantifying superspreading for COVID-19 using Poisson mixture distributions. *Scientific reports*. 2021;11(1):1–11.
30. Althouse BM, Wenger EA, Miller JC, Scarpino SV, Allard A, Hebert-Dufresne L, et al. Superspreading events in the transmission dynamics of SARS-CoV-2: Opportunities for interventions and control. *PLoS Biology*. 2020;18(11):e3000897.
31. Asadi S, Wexler AS, Cappa CD, Barreda S, Bovier NM, Ristenpart WD. Aerosol emission and superemission during human speech increase with voice loudness. *Scientific Reports*. 2019;9:2348.
32. Buonanno G, Stabile L, Morawska L. Estimation of airborne viral emission: Quanta emission rate of SARS-CoV-2 for infection risk assessment. *Environment International*. 2020;141:105794.
33. Chen PZ, Bobrovitz N, Premji Z, Koopmans M, Fisman DN, Gu FX. Heterogeneity in transmissibility and shedding SARS-CoV-2 via droplets and aerosol. *eLife*. 2021;10:e65774.

34. Edwards DA, Ausiello D, Salzman J, Devlin T, Langer R, Beddingfield BJ, et al. Exhaled aerosol increases with COVID-19 infection, age, and obesity. *Proceedings of the National Academy of Sciences*. 2021;118(8):e2021830118.
35. He X, Lau EHY, Wu P, Xilong D, Wang J, Hao X, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine*. 2020;26.
36. Van Kerckhove K, Hens N, Edmunds JW, Eames KTD. The impact of illness on social networks: Implications for transmission and control of influenza. *American Journal of Epidemiology*. 2013;178(11):1655–1662.
37. Galvani A, May R. Dimensions of superspreading. *Nature*. 2005;438:293–295.
38. Leclerc QJ, Fuller NM, Knight LE. What settings have been linked to SARS-CoV-2 transmission clusters? *Wellcome Open Research*. 2020;5(83).
39. Azuma K, Yanagi U, Kagi N, Kim H, Ogata M, Hayashi M. Environmental factors involved in SARS-CoV-2 transmission: effect and role of indoor environmental quality in the strategy for COVID-19 infection control. *Environmental Health and Preventive Medicine*. 2020;25:66.
40. Torneri A, Libin P, Scalia Tomba G, Faes C, Wood JG, Hens N. On realized serial and generation intervals given control measures: The COVID-19 pandemic case. *PLoS Computational Biology*. 2021;17(3).
41. Duan W, Qiu X, Cao Z, Zheng X, Cui K. Heterogeneous and Stochastic Agent-Based Models for Analyzing Infectious Diseases' Super Spreaders. *IEEE Intelligent Systems*. 2013;28(4):18–25.
42. Fujie R, Odagaki T. Effects of superspreaders in spread of epidemic. *Physica A: Statistical Mechanics and its Applications*. 2007;374(2):843–852.
43. Meyers LA, Pourbohloul B, Newman ME, Skowronski DM, Brunham RC. Network theory and SARS: predicting outbreak diversity. *Journal of Theoretical Biology*. 2005;232(1):71–81.
44. Kim Y, Ryu H, Lee S. Agent-Based Modeling for Super-Spreading Events: A Case Study of MERS-CoV Transmission Dynamics in the Republic of Korea. *Int J Environ Res Public Health*. 2018;15(11):2369.
45. Kuylen E, Stijven S, Broeckhove J, Willem L. Social contact patterns in an individual-based simulator for the transmission of infectious diseases (stride). *Procedia Computer Science*. 2017;108:2438–2442.
46. Van Hoang T, Coletti P, Kifle YW, Van Kerckhove K, Vercruyssen S, Willem L, et al. Close contact infection dynamics over time: insights from a second large-scale social contact survey in Flanders, Belgium, in 2010–2011. *BMC infectious diseases*. 2021;21(1):1–15.
47. Alimohamadi Y, Taghdir M, Sepandi M. Estimate of the basic reproduction number for COVID-19: a systematic review and meta-analysis. *Journal of Preventive Medicine and Public Health*. 2020;53(3):151.
48. Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *International Journal of Infectious Diseases*. 2020;92:214–217.

49. Andersson H, Britton T. Stochastic epidemic models and their statistical analysis. vol. 151. Springer Science & Business Media; 2012.
50. Seabold S, Perktold J. Statsmodels: Econometric and statistical modeling with python. In: Proceedings of the 9th Python in Science Conference. vol. 57; 2010. p. 61.
51. Coletti P, Wambua J, Gimma A, Willem L, Vercruyse S, Vanhoutte B, et al. CoMix: comparing mixing patterns in the Belgian population during and after lockdown. *Scientific Reports*. 2020;10(1):21885.
52. Thurner S, Klimek P, Hanel R. A network-based explanation of why most COVID-19 infection curves are linear. *Proc Natl Acad Sci U S A*. 2020;117(37):22684–22689.
53. Cooper L, Kang SY, Bisanzio D, Maxwell K, Rodriguez-Barraquer I, Greenhouse B, et al. Pareto rules for malaria super-spreaders and super-spreading. *Nature Communications*. 2019;10:3939.
54. Gomes MGM, Aguas R, King JG, Langwig KE, Souto-Maior C, Carneiro J, et al. Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold. *medRxiv*. 2020;.
55. James A, Pitchford JW, Plank MJ. An event-based model of superspreading in epidemics. *Proceedings of the Royal Society B*. 2007;274:741–747.
56. Libin PJK, Willem L, Verstraeten T, Torneri A, Vanderlocht J, Hens N. Assessing the feasibility and effectiveness of household-pooled universal testing to control COVID-19 epidemics. *PLoS Computational Biology*. 2021;17(3):e1008688.