medRxiv preprint doi: https://doi.org/10.1101/2020.05.17.20104745; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Impact of Superspreaders on dissemination and mitigation of COVID-19

Kim Sneppen[1,*] **and Lone Simonsen**[2]

[1]University of Copenhagen, Niels Bohr Institute, 2100 København Ø, Denmark
[2]University of Roskilde, Department of Science and Environment, 4000 Roskilde, Denmark
[*]sneppen@nbi.ku.dk

ABSTRACT

Background: The emerging SARS-CoV-2 virus has caused a global pandemic characterized by "superspreader" events. Heterogeneity in transmission risk is a known phenomenon in infectious diseases and was also seen in SARS and MERS outbreaks since 2003. The pandemic has led countries to control the pandemic spread using unprecedented severe mitigation strategies that include quarantine and lock-down. This has has been highly successful in terms of halting the spread, but had enormous socioeconomic cost. We set out to theoretically explore whether a model that includes the phenomenon of “superspreaders” may guide us towards a cost-effective mitigation strategy.

Methods: We developed an agent-based model that includes persons that spread the disease far more readily than other persons. This allow us to investigate effect of containment strategies in both a random SEIR like scenario, and in a transmission model that work with structured society. Our model considered even contact patterns in three settings: home, school/work and a category of "other" settings representing diffuse social contacts. We introduced superspreaders either as a fixed proportion of the population, or as a broad spectrum of infectivity. For each choice, the model was then calibrated to an overall realistic daily growth rate of 23%. As sensitivity analyses we varied the fraction of superspreaders as well the distribution of their infectivity. To compare the simulation findings for each mitigation scenario we considered the maximum daily ICU utility.

Results: As expected, without mitigation imposed, the inclusion of "superspreaders" does not meaningfully change the epidemic trajectory. However when introducing mitigation strategies imposed on each of the three settings, we find that the presence of "superspreaders" made a substantial difference. The simulations demonstrate that the best strategy is to focus on limiting contacts in the "other" category. This in particular suggests that limiting diffuse social contacts in settings such as bars, transportation, restaurants, parties, concerts and lecture halls is far more effective than limiting the same amount of contact events in the "home" and "school/workplace" setting.

Conclusions: To appreciate effect of heterogeneity in various social spheres we need to rethink disease transmission models. Doing so, we found that wide distribution of infectivity favours strategies that reduces the max, while leaving the typical behaviour undisturbed. We found that most workplaces may be opened without much influence on the epidemic, while one could snuff out transmission with an effective avoidance of "other" contacts. Interestingly, including a consideration of "superspreaders" can help explain the dramatic success of even moderate lock-down strategies as that practiced in Denmark and Sweden. And it points to the need to avoid mass gatherings until either flock immunity has been achieved or an effective vaccine is available.

Introduction

The emerging SARS-CoV-2 virus that is causing the COVID-19 pandemic which emerged in late 2019 has lead countries around the world to an unprecedented lock-down strategy. We see that these lock-downs work extremely well in terms of halting the spread, at enormous socioeconomic cost. However, we do not know what aspects of the mitigation efforts are causing the effect. As many countries already or soon will be looking to open up the country again, the relative contribution of different aspects – reducing contacts at home, schools, work or in society – should be better understood.

For this pandemic, a pattern of "superspreader" events has been documented to occur[1]. For example, as South Korea reopened the country carefully recently, a single infected person visiting a night club led to at least 50 new infections. In the US, a 2,5 hour long choir rehearsal tragically led to 75% of the participants getting infected and several deaths. These, as well as outbreaks in prisons and hospitals and following carnevals are vivid examples of SARS-CoV-2 superspreader events illustrate that these are important in the spread of this novel virus. It is also reminiscent of documented superspreader events in the dissemination of SARS-CoV in the 2003 outbreak in Asia and Canada. One recent study estimated that 80% of infections may

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.17.20104745; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

be caused by 10% of the infected population, pointing to transmissibility (R_0) being highly heterogeneous[2]. Thus, the practice of relying on average R_0 in dynamic disease models can obscure considerable individual variation in infectiousness including extreme transmitters[3,4]. It is also understood that such heterogeneity will not affect the epidemic spread by much in standard dynamic SEIR models. One might ask why are we often seeing superspreading events as a precurser to outbreaks? Also, why do we see marked heterogeneity in the geographical distribution and intensity of outbreaks across countries and regions?

Such heterogeneity in transmission risk is a well understood phenomenon in infectious diseases. In 2003 Riley et al. pointed out that 'superspreading events' could explain 80% of transmission events for a wide variety of diseases and thus being a defining characteristic of many epidemic diseases, including vector-borne diseases due to large variability in parasite density in blood in infected individuals[5]. In the 2014 Ebola outbreak in west Africa, superspreader events are thought to have played a key role in sustaining onward transmission of the epidemic[6]. Overall, this means that certain individuals infect far greater numbers of persons than the average.

For SARS-CoV-2 the observed patient-to-patient variation in virus load appears to be many orders of magnitude[7]. Similarly, observed infectivity varies greatly[2]. Therefore we choose to focus on variation in person specific infectivity, while assuming that there is no variation in individuals exposure.

We ask here whether the phenomenon of superspreading might play an important role in our ability to control the ongoing COVID-19/SARS-CoV-2 pandemic with non-pharmaceutical mitigation strategies. To investigate the effect of heterogeneity in various spaces we needed to rethinking our disease transmission models. One can easily do this in a agent-based model in which each person can be assigned individual infection properties. Our findings suggest a way forward in terms of rating the importance of limiting spread in the home, the workplace versus in the environment where many new contacts are encountered and superspreaders events may occur.

Methods

We use the agent-based model which was previously described in[9], see also Fig. 1. Other agent based models with similar structure can be found in[10]. Briefly, this is an age stratified variant of that model, a modification that allow us to quantify the epidemics in terms of the need for intensive care units (ICU). The model is simulated in discrete time-steps of 6 hours, during which each infected agent has a probability of infecting another agent, potentially leading to transmission of the disease. We have checked that smaller time steps do not change our results. As time progress an infected agent may change state, with rates set by the duration of each interval in the top panel of Fig. 1. As we further follow the occupation of ICU, then the model also simulate the agents through the latter steps of the disease using probabilities from table 1.

Throughout the paper we calibrate the infection rates to fit an exponential growth of the unconstrained epidemics to be the 23% per day that was reported by[11]. With parameters from Fig. 1 we find that this correspond to an R_0 of 2.9 in a randomly mixed population. For each simulation we measure R_0 as the average number of infection attempts per person that have been sick. Typical simulations have a population of 200000 and the epidemic is initiated at time zero by 50 infected persons.

In the paper we first consider a "well-mixed" population without any social or age structure. The results in Fig 2, was thus obtained by selection of pairs of agents randomly across the entire population for each infection attempt.

Fig.3, 4 simulate a population where each person is assigned 3 different types of contacts as outlined in Fig. 1. This choice of contacts aims to represent a simplification of society into 3 equally weighted interactions: "Home", "Work" and "Other". In reality, the observed social activity have been estimated to be more tilted toward the "Home" sector and less toward to work/school type of contacts[8,12,13]. By making the relative weights equal it becomes simpler to compare them in our analysis.

The "Home" interactions are build from assigning each person a random home with an average size of 2.1. In these homes we constrain people above 20 to be within 1 age group from each other. We further assign children to parents that are 20-40 years older.

The "Work" interactions are modeled as clusters of average size 8 (Poisson distributed). Further we assign each person connections to two random persons outside this cluster. All ages between 20 and 70 are mixed in the work places. For persons under age 20, we use school classes of average size 24. Social activity per person across home and school is the same, apart from a re-scaling to the fit overall age dependent activity.

The interactions within both "Home" and "Work" are fixed throughout the epidemic, and are highly clustered. Noticeably, the "Home" clusters are by far the smallest.

The "Other" group in principle contains all other interactions, including in particular all public interactions. At each infection attempt, the interactions are chosen at random from the entire population.

The probability for selecting "Home", "Work" or "Other" are calibrated such that they occur equally often. Furthermore, the agents is assigned an age dependent social activity that is fitted to reproduce the contact data of[8], see insert of Fig. 1.

Modeling superspreaders: The eventual superspreaders (i) are selected at the beginning of the simulation and assigned an individual activity (s_i). Once selected, a superspreader will engage in s_i infection attempts in the allocated time interval of 6 hours, where the probability of infecting someone is the same in each encounter. All other persons are simply assigned an $s_i = 1$.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.17.20104745; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

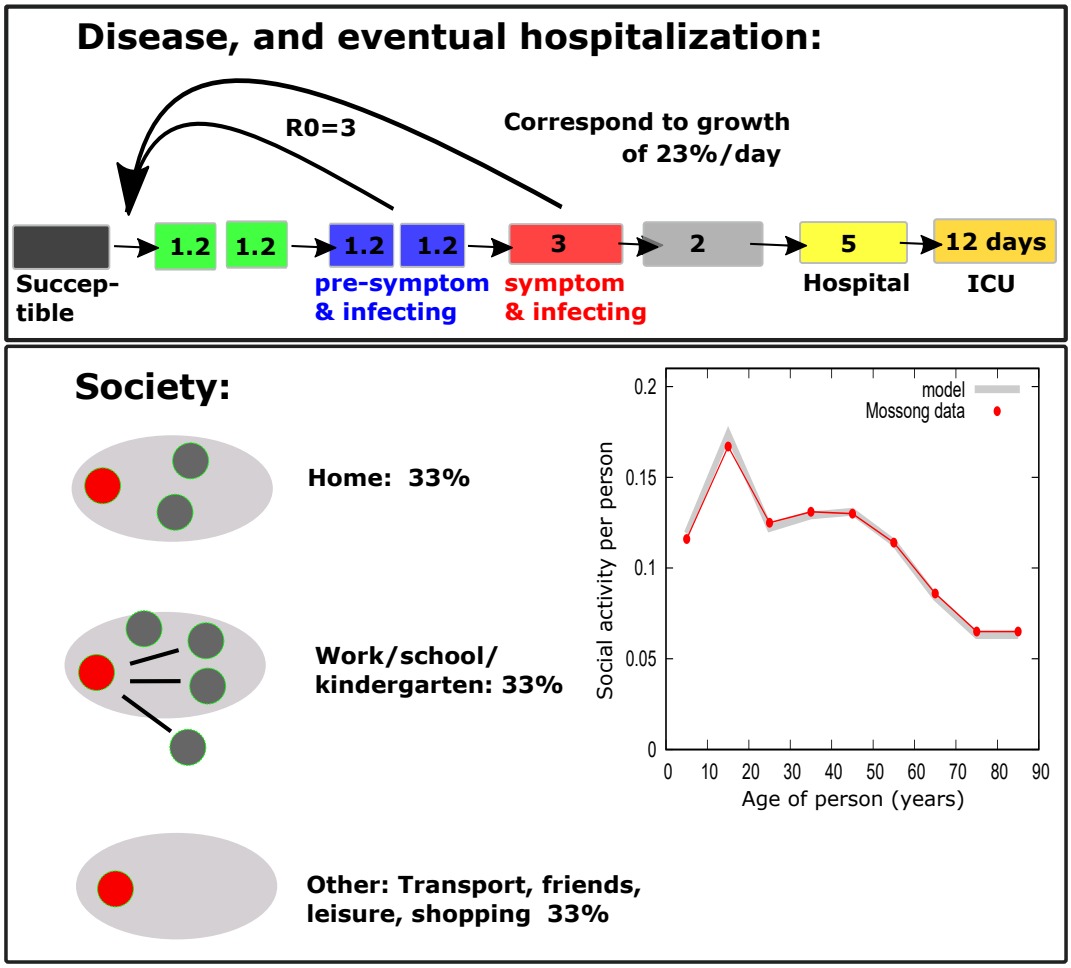

Figure 1. Agent based model with social structure. The model include 45% pre-symptomatic infection. The top panel show various stages of the disease, each simulated with a corresponding rate. People can infect from blue and red boxes. The bottom illustrate the model society, where interactions of each individuals are divided into 3 equal size groups: “Home", “Work" and “Other". Each group is weighted with 1/3. Home is the smallest unit, while work (and school) is larger. The possible home and work interactions are fixed throughout the epidemic. The “Other" group is treated as random contacts. Persons are further assigned a social activity as function of age shown in the insert. The inserted graph show that the social activity as function of age is adjusted to fit the data from ref.[8].

medRxiv preprint doi: https://doi.org/10.1101/2020.05.17.20104745; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Age group	Probability of hospitalisation	Probability of ICU given hospital	Population distribution
0-9	0%	5%	10.9%
10-19	0.013%	5%	11.9%
20-29	0.37%	5%	13.3%
30-39	1.1%	5 %	11.7%
40-49	1.4%	6.3%	13.6%
50-59	2.7%	12.2%	13.6%
60-69	3.9%	27.4%	11.7%
70-79	5.5%	43.2%	8.9%
80-	5.5%	70.9%	4.3%

Table 1. *Distribution of our simulated population, in age groups and with probability to end in Intensive Care Unit given hospitalization. The values for hospitalization and ICU occupancy are from the Norwegian health authorities:* $https://www.fhi.no/sv/smittsomme-sykdommer/corona/koronavirus-modellering/$ or equivalently from[14] calibrated to an estimated fatality rate for SARS-CoV-2 of 0.3%.

This way of introducing superspreaders in the model is mimicking the introduction of a person-specific reproductive number as it was done by Lloyd et al in 2005[15]. For the base model, and for all dynamic simulations in Fig. 2 and 4 we use $s_i = 50$ for 10% of the population. In the sensitivity analysis (see Fig. 4e,f) we first consider a range of prevalence of superspreaders (from 0% to 20% of the population). Then we consider a distribution of superspreader intensity, where s_i are drawn from a distribution $p(s) \propto 1/s^{\gamma}$, $s \in [1, 1000]$. The choice of the upper cut-off also influences the results, as the the average infections per person are divergent with the upper limit when $\gamma \leq 2$. A high γ corresponds to less intense superspreaders.

Results: Mitigation strategies

The agent based perspective on the spread of an epidemic in a population allow us to address different types of questions and mitigation strategies that are possible with the more commonly used SEIR models. Naturally an agent based model resembles its corresponding SEIR variant when there is many agents and they are simulated without any social structure. But when there is repeated social contacts, the results of the models may differ.

First we modeled a COVID-19 like epidemic (as described above) and without "superspreaders and social structure and using $R_0 = 2.7$. Figure 2a) at 100% contact rate corresponds to such an idealization. Fig. 2a corresponds to an SEIR model without social structure; When we reduced the contact rate so that only 50% and 25% of contacts remain, we see that the SARS-CoV-2 epidemic was mitigated, respectively halted, respectively. The largest reduction was enough to bring the effective R_0 (Re) below 1.

Introducing superspreaders into a model without social structure Next we introduced "superspreaders" into the model (see Fig. 2b) to examine the effect of this phenomenon. Fig. 2b) assume that 10% of the agents are 50 times more infectious than the rest. With this assumption, in this model we count that 85% of all transmission events are from superspreaders. Although for this model, the unmitigated epidemic progresses quite similar to the classical model in panel a), we are able to halt the epidemic by only reducing the maximal number of encounters. In fact, only 25% contacts remaining for the superspreader is enough to halt the epidemic. Thus, when superspreaders are prevented from doing their thing, then a dramatic level of mitigation can be achieved while everybody else can live their normal lives. Of course, this scenario is difficult to achieve in real life where we do not know the identify of the superspreaders.

Introducing social structure (but no superspreaders) into the model Next we considered superspreading in the context of a structured society, by introducing into the model three sectors where interactions can take place (Fig. 1). This allow us to introduce an indirect cap on the maximal social activity without knowing the identity of superspreaders.

We next simulated the epidemic trajectory for a socially structured model without superspreaders (Fig. 3). The infection rate is again adjusted to fit the overall growth rate of 23%, and (as consequence) we observe an epidemic that quite closely resembles the basic SEIR model (see Fig 2a). We examine the effect of lock-downs in each of the three contact sectors (home, work/school, other) – one at the time (see Fig. 3b-d). In panel b) when disallowing any social contact in the "home" sector, we find a 1/3 reduction in the epidemic peak size and ICU usage. Next, we closed down all contacts at the "work/schools" sector and find the peak reduced to about half the unmitigated size; ICU less so due to elderly not being directly affected by this. Mitigating “Work" contacts have a slightly larger influence impact than mitigation "home" contacts because the associated network contacts are more extended. Finally, we disallowed the "other" sector, that is, social contact outside the home and work/school sectors. Doing so we find a sizeable reduction in the epidemic peak size to 1/3 of its original size and an even

medRxiv preprint doi: https://doi.org/10.1101/2020.05.17.20104745; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

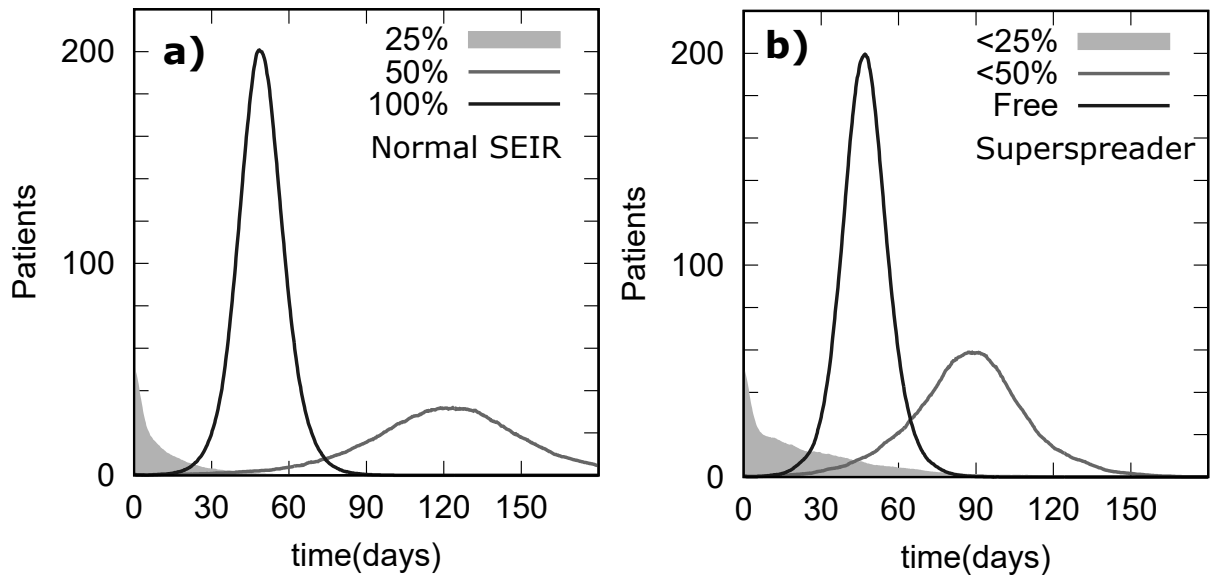

Figure 2. Effects on epidemic trajectory of the proportion of normal contacts in a) a classical SEIR model and b) a superspreader SEIR model. Panel **a**: Classic (SEIR) type model where all agent infect at same rate ($R_0 = 2.7$). The 100% curve marks the unmitigated epidemic, the other two curves marks two level of lock-down in which 50% and 25% of contacts remain, respectively. Panel **b**: Same model, but now assuming that 10% of infected are superspreaders that transmit 50 times more effectively than others. Here the Lock down only imposes a limit on the maximum number of contacts, rather than restraining contacts for all people.

greater reduction in ICU utilization (Fig 3d). This reflects the importance of diffuse contacts, and the fact that elderly are (nearly) only connected to the society through "other" contacts in our model.

The final size of the epidemics for each scenario were inspected (In Figure 3e). We found that the model predicts attack rates of between 80% and 100% of the population, with or without mitigation in the home and work/school sector. However, when closing down the "other" sector of diffuse contacts, it greatly reduces the attack rate, and nearly eliminates illness among elderly, for reasons described above. This results suggests that mitigating "other" events (concerts, etc, but also limiting contacts with family members and sealing off elderly care homes) would greatly reduce infections in the age group at highest risk for severe illness and death.

Agent-based model with social structure and superspreaders We then introduced superspreaders into the socially structured model. Fig. 4 recapitulates the simulations shown in Fig. 3, but this time with the presence of 10% of the population being superspreaders. In this scenario we count that the 10% superspreaders causes 75% of all infections. This is lower than the 85% in the model version without social structure, because social structure put constraint on superspreader activity. This is compatible with finding of others[2] who estimate that 10% of infected causes about 80% of the infections in the Covid-19 epidemic.

In figure 4 panel b) and c) we see that the elimination of work/school or home contacts, respectively, had less influence on the epidemic peak size than in the model without superspreaders (as in Figure 3). Even if one eliminates simultaneously all contacts in both “home" and "work/school" sector, then the peak epidemic size is halved compared to only removing one sector, and the R_0 value is about 1.7. In contrast, Fig. 4d shows that the epidemic is halted by eliminating "other" contacts; to visualize the decline of the epidemic we therefore seeded with 5% of the population infected.

We conclude that superspreaders matter greatly in a mitigation scenario. Epidemics driven by superspreaders are less sensitive to reductions in close contacts at home or work/school, but highly affected by changes in their random contacts.

Sensitivity analyses of superspreader properties Next, we asked whether this striking prediction was sensitive to assumptions about the frequency and intensity of superspreaders. To do so we theoretically explored two types of superspreaders assumptions. To simplify, we considered lockdown effect in each sector in prioritized order: first "other", then "work/school"until the epidemic was contained. We do not consider "home" lockdown as a realistic option. We first allowed the proportion of superspreaders to vary between 0% and 20% while retaining a superspreader intensity of 50 times above normal. In Figure 4e, the reduction in contacts needed to halt the epidemic is indicated in yellow. Thus, the boundary of the yellow region marks epidemic control ($Re = 1$). It can be seen that when the fraction of superspreaders decrease below 8% the epidemic cannot be stopped by closing down the "other" sector alone. Furthermore, the required lock down intensity increases substantially when

medRxiv preprint doi: https://doi.org/10.1101/2020.05.17.20104745; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

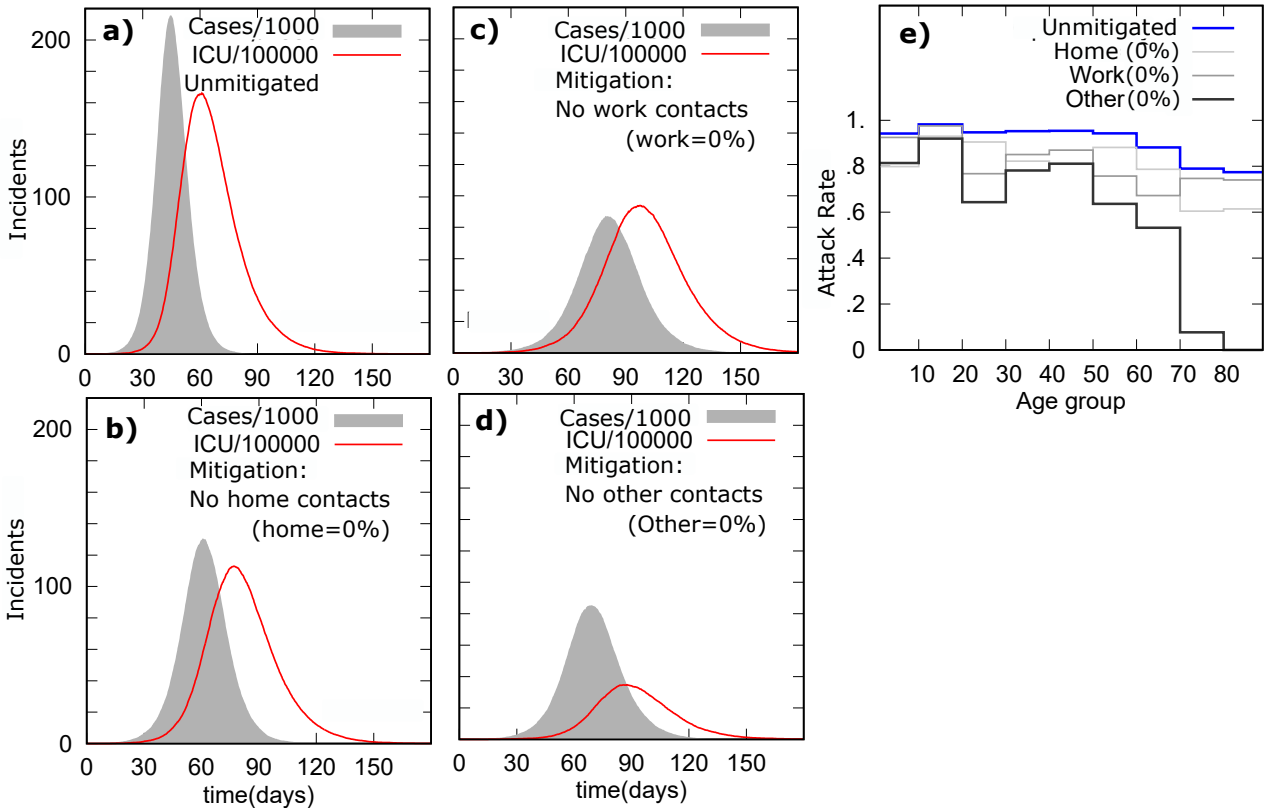

Figure 3. Agent based model with three sectors of contacts (social structure) but no superspreaders. a) Unmitigated epidemic scenario (gray curve) using Norwegian age-specific data for ICU utilization (red curve). **b)** Model simulation without home contacts, and **c)** without work/school contacts, and **d)** without "other" contacts, that is, contacts outside home and work/school. Three quarter reduction in peak daily ICU usage is achieved when limiting the "other" contacts, that is, contacts outside home and work/school. **e)** Show the final population attack rate as a function of age for each of the four simulations in panel a,b,c,d. We find that mitigating "other" contacts result in dramatic reductions in ICU. This is explained by the older generations largely having contact to the remaining society through contacts of the "other" type.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.17.20104745; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

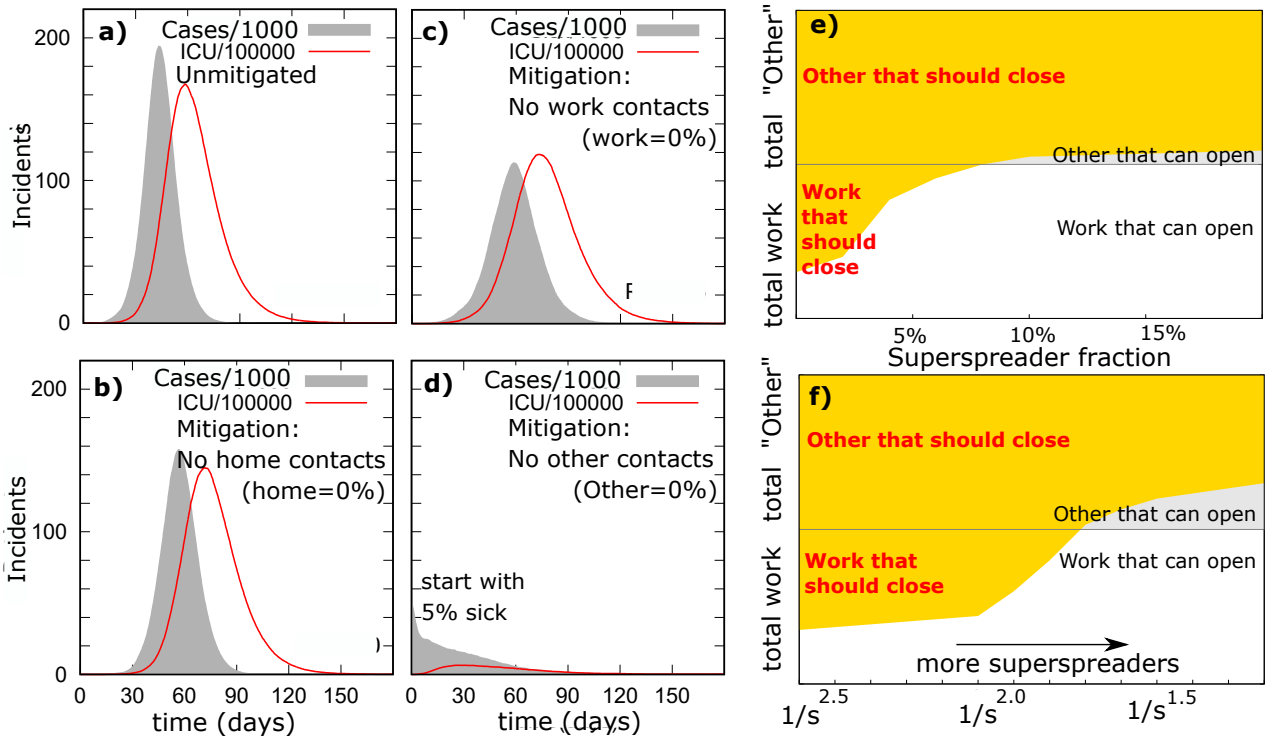

Figure 4. Agent based model with social structure and superspreaders. The panels a-d) recapitulate the simulations done for the social-structured model. **a)** Unmitigated epidemic scenario (gray curve) using Norwegian age-specific data for ICU utilization (red curve). **b)** Model simulation without home contacts, and **c)** without work/school contacts, and **d)** without "other" contacts, **e)** Sensitivity analysis where the yellow area mark the fraction of "Other" and "Work" that needs to be closed to terminate the Covid-19 epidemics. Here the superspreaders are maintained at $s_i = 50$ while the number of them are varied between 0 and 20%. **f)** Sensitivity analysis of model with a distribution of individual infectivity drawn from $p(s) \propto 1/s^{\gamma}$ with γ between 1.2 and 2.5.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.17.20104745; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

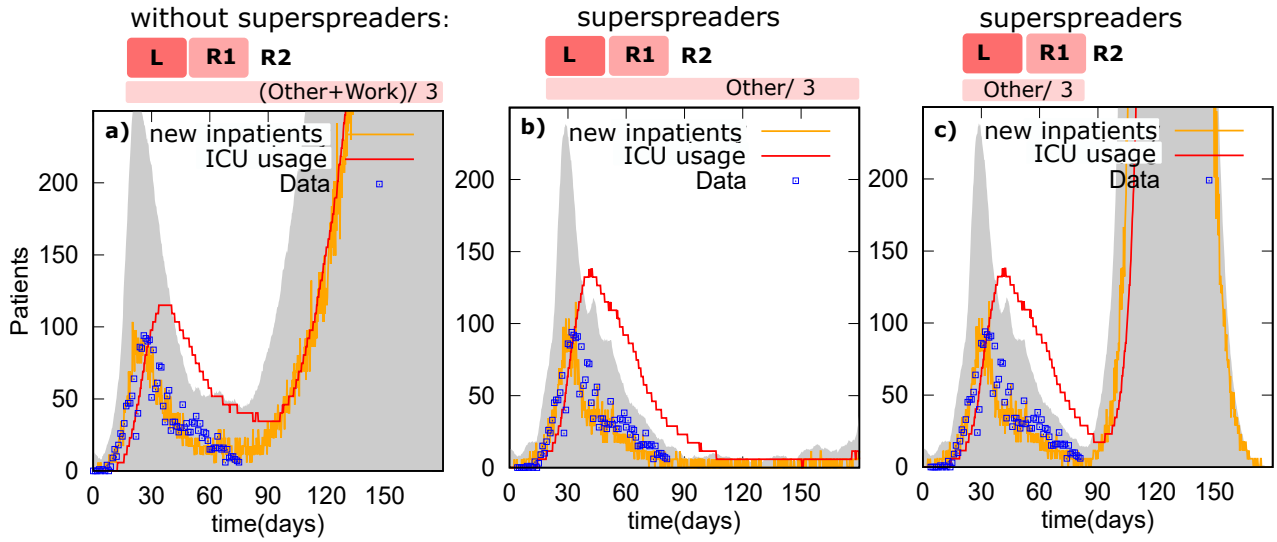

Figure 5. Applying our agent-based social structure model with and without superspreaders, to Danish COVID-19 data. Using 500000 agents and seeding with 100 infected persons, We fitted our model to to danish daily time series of numbers of new inpatients (blue circles). The initial lock down is assumed to reduce contacts in the work/school sector to 25% (L), and then this is partially lifted one month later,to 50% (R1). We let the "other" sector be reduced with the same factor to reflect closing of restaurants, gatherings and concerts etc. R2 mark the period after the second opening. Comparison between panel a,b show that the future trajectory is a lot better is superspreaders are in the model. Panel c demonstrates a dire outlook if social distancing in the "other" sector is not maintained.

the superspreader fraction is small. Below ∼4%; superspreaders they only contribute a minor fraction of all infections in the population and are no longer important.

We next explored the effect of various intensity of superspreaders, rather than an either-or scenario (Panel f). Each person was at the start of the simulation assigned a particular infectivity drawn from a power law distribution, and also imposing an upper limit of contact intensity, at 1000 times the normal level. When this distribution is wider than about $1/s^{1.7}$, corresponding to 22% of superspreaders causing 80% of the infections, the Covid-19 epidemic can be mitigated more easily – by a partial lockdown of the "other" sector. If the distribution is wider, for example $1/s^{1.5}$, corresponding to 16% of infected causing 80% of infections, one can stop the epidemic by reducing “other" sector contacts to 80% of the normal level.

Simulations of a real life lock down sequence. To further explore the importance of superspreading events, we next applied our model to the real life situation that took place in Denmark after a lock-down on March 13. In this country of 5.6 million people the lockdown was achieved early in the epidemic, and was followed one month later by a first reopening stage. And after one additional month by a second reopening. After the lockdown date, the population observed general social distancing in public spaces, as well as careful hygiene measures.

Lets assume that the first lock-down imposed a reduction in the work/school sector to about 25% of the pre-pandemic level, and that a month later this is eased to about 50% of the pre-pandemic level (Figure 5). Then in the 2nd reopening we assume that all restrictions on the ‘Work/school" sector are dropped. We furthermore assumed that the "other" sector would be reduced and reinstated (100% to 25% to 50% to 100%). Our model simulations of this - with and without superspreaders – show that we would fail to reproduce the observed efficiency of the lock down (not shown).

To fit the model to the observed decline in daily numbers of new hospitalizations we further assume that the general social distancing and hygiene efforts were effective outside the home has led to a further reduction of 2/3 in the work and other sector (we multiplied by a factor 1/3) (panel a). Next, in Panel b) we introduce superspreaders and obtain a better fit. In this scenario it is sufficient to assume that the general distance/hygiene factor only reduce the “Other" sector (and not in the work/school sector). Alternatively, one could interpret the reduction in contact rate as the effect of closing down venues where many people meet. The one month forecast past the recent second reopening is far more optimistic with than without superspreaders in the model (compare panel a and panel b). Not surprisingly, a scenario in which we completely return to pre-pandemic interactions and hygiene is far more gloomy – regardless of whether there are superspreaders (See Fig. 5c).

medRxiv preprint doi: https://doi.org/10.1101/2020.05.17.20104745; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Discussion

SARS-CoV-2 is an intriguing pathogen, that challenged our understanding of disease transmission already from its beginning in Wuhan. How could the epidemic be so slow to spread in the epicenter in Wuhan with an estimated growth of 10% a day[16], while it later expanded with 30% per day. Why did it spread so slowly in Japan and why did it never cause a major epidemic in Africa while causing disaster with growth in death rates of up to 40%/day in northern Italy and Spain. Are we missing something? Was it the elimination of superspreading opportunities that ended the SARS outbreak in 2003?

Using an agent-based model we were able to investigate the likely effect of superspreaders on the effect of mitigation schemes. Superspreading is a phenomenon known to occur for coronavira outbreaks of SARS, MERS and now also for COVID-19[1,2]. While we have observed even dramatic superspreading events in which a single person manages to infect up to hundreds in a short time span, it nevertheless appears to be important for the dissemination of COVID-19 epidemic. But the mechanism behind the phenomenon is not well understood; are for example some individuals genetically more prone to spreading widely? Or is anyone a superspreader the day of symptom onset? or could it be that certain activities or gatherings with singing, yelling facilitates superspreading? Because we do not know this we modeled superspreading as a fixed proportion of individuals who transmits 50-fold more than others. Then, in sensitivity analyses, we explored other possibilities such as a smaller fraction and a power-log distribution of superspreaders among all infected individuals.

We have demonstrated that the success of a lockdown and the effects of subsequent re-openings are highly dependent on the frequency and intensity of superspreaders in the population. Our used distribution of superspreaders for SARS-Cov-2 compatible with recent findings by others[2] where they estimate that 10% causes 80% of infections. We used $p(s) \propto 1/s^{1.5}$ which makes 16% causing 80% of the infections in our structured population.

Assuming arbitrarily that interactions are evenly distributed in three sectors – home vs work/school vs "other" sectors –, we investigated the effect of superspreaders in such a socially structured environment. As expected, the presence or absence of superspreaders does not matter much to an unmitigated epidemic. However, in a mitigated epidemic, as COVID-19 just now, superspreaders matters profoundly. With this model and assumptions we simulated COVID-19 epidemics with a focus on the magnitude, trajectory of the outbreak and noting the peak level of ICU utilization (the latter as a measure of impact on the health system). We found the following major results for our agent-based model where superspreaders was acting within a socially structure:

1. In comparison to mitigating in the "home" and "work/school" sector, mitigating the "other" sector has a profound impact - it halts the epidemic. This finding suggests that limiting diffuse/random social contacts such as that occurring in transportation, gatherings, weddings, religious meetings, and frequent trips to shops, bars and restaurants is what drives the COVID-19 epidemic. This dramatic finding only obtains when superspreaders are in the model.
2. When we fit our model to actual observed daily Danish COVID-19 hospitalization data from the epidemic pre-intervention period and two months after the lockdown was in effect (March 13), we found that our one month forecast of the epidemic trajectory following the reopening of the country looked far more promising with than without superspreading, providing one continue to limit social contacts in the "other" sector.

Since superspreaders are important, then the way to optimize mitigation is probably to limit large gatherings where superspreaders can do their deed, but also multiple visits to smaller gatherings. Similar mitigation strategies have been suggested by others in[2,17].

In terms of caveats, we did not consider superspreader "events". Anecdotes about SARS-CoV-2 include the deadly choire rehersal in United States and the seeding of the South Korean epidemic from religious group events in which one or a few individuals created a burst of cases. Should this sort of events be common and important, then our strategy of capping the maximal social activity outside work and home would both reduce the chance and impact of such events. A more elaborate model could benefit from separating large events in the "other" encounter sector.

In our simulations we relegated all diffuse contacts to an “other" sector, and assumed all other contacts (with known persons) occurred through fixed social networks at home and work/school. In reality part of the interactions in the “other" sector is of known persons, for example friend and extended family. On the other hand, part of the interactions in the “work" sector would involved non-known persons, such as casual contacts at cafeterias, conferences and workshops. Importantly, all of this is not too concerning, because what really matters is how many different people a superspreader meets during the period where he/she infects excessively.

Now, about 4 months into the pandemic we see a heterogeneous epidemic patterns across European countries. One possible reasons for this could be the work of superspreaders. An epidemic take-off may well require a superspreader because the R_0 would be below 1 for the rest of the population. But if a superspreader is infected, the disease may spread to other superspreaders. Although hard to prove, it would not surprise us if the observed major differences across cities, regions and countries could be understood as a stochastic phenomenon modulated by a larger chance for the epidemic to percolate among superspreaders in large cities.

It is in this perspective that we propose that COVID-19 models used to model the epidemic trajectory and forecast the effect

medRxiv preprint doi: https://doi.org/10.1101/2020.05.17.20104745; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

of mitigation strategies and re-openings are inaccurate unless they include superspreaders. Without this element, such models will easily overestimate the epidemic size as well as the mitigation intensity needed in order to obtain control. Better models depend on insights about the superspreader phenomenon, especially from direct observation of the frequency of its occurrence in the population, and whether it is an intrinsic or extrinsic property, that is, linked to genetics or to environmental and social opportunities.

Acknowledgements This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program under grant agreement No. 740704, and from the Carlsberg Foundation in Denmark.

References

1. Liu, Y., Eggo, R. M. & Kucharski, A. J. Secondary attack rate and superspreading events for sars-cov-2. *The Lancet* **395**, e47 (2020).
2. Endo, A., Abbott, S., Kucharski, A. J., Funk, S. *et al.* Estimating the overdispersion in covid-19 transmission using outbreak sizes outside china. *Wellcome Open Res.* **5**, 67 (2020).
3. May, R. M. & Anderson, R. M. Transmission dynamics of hiv infection. *Nature* **326**, 137–142 (1987).
4. Woolhouse, M. E. *et al.* Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proc. Natl. Acad. Sci.* **94**, 338–342 (1997).
5. Riley, S. *et al.* Transmission dynamics of the etiological agent of sars in hong kong: impact of public health interventions. *Science* **300**, 1961–1966 (2003).
6. Lau, M. S. *et al.* Spatial and temporal dynamics of superspreading events in the 2014–2015 west africa ebola epidemic. *Proc. Natl. Acad. Sci.* **114**, 2337–2342 (2017).
7. Jones, T. C. *et al.* An analysis of sars-cov-2 viral load by patient age (2020).
8. Mossong, J. *et al.* Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS medicine* **5** (2008).
9. Eilersen, A. & Sneppen, K. Estimating cost-benefit of quarantine length for covid-19 mitigation. *medRxiv* (2020).
10. Aleta, A. *et al.* Modeling the impact of social distancing, testing, contact tracing and household quarantine on second-wave scenarios of the covid-19 epidemic. .
11. Remuzzi, A. & Remuzzi, G. Covid-19 and italy: what next? *The Lancet* (2020).
12. Klepac, P., Kissler, S. & Gog, J. Contagion! the bbc four pandemic–the model behind the documentary. *Epidemics* **24**, 49–59 (2018).
13. Klepac, P. *et al.* Contacts in context: large-scale setting-specific social mixing matrices from the bbc pandemic project. *medRxiv* (2020).
14. Verity, R. *et al.* Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infect. Dis.* (2020).
15. Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E. & Getz, W. M. Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359 (2005).
16. Li, Q. *et al.* Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia. *New Engl. J. Medicine* (2020).
17. Ebrahim, S. H., Ahmed, Q. A., Gozzer, E., Schlagenhauf, P. & Memish, Z. A. Covid-19 and community mitigation strategies in a pandemic (2020).