Research Report

Social heterogeneity drives complex patterns of the COVID-19 pandemic: insights from a novel Stochastic Heterogeneous Epidemic Model

Alexander V. Maltsev[1] and Michael D. Stern[2]*

[1]Translational Gerontology Branch, National Institute on Aging/NIH, Baltimore, MD, USA

[2]Laboratory of Cardiovascular Science, National Institute on Aging/NIH, Baltimore, MD, USA

**Corresponding Author*:

Michael D. Stern. Email: SternMi@mail.nih.gov

*ORCID:*Alexander V. Maltsev: 0000-0002-2219-1114; Michael D. Stern: 0000-0003-4476-7425

Keywords: COVID-19, SARS-CoV2, infection, stochastic model, hotspots, second wave, social interactions

Author Contributions

A.V.M. and M.D.S. designed research, performed research, analyzed data, and wrote the paper.

This PDF file includes: Main Text: Figures 1 to 5. Supplemental Information (Figures S1 to S2 and movie captions)

Abstract: As of August 2020, it has become evident that regional infection curves of COVID-19 exhibit complex patterns which often differ from curves predicted by forecasting models. We hypothesized that this may be due to social heterogeneity not accounted for by regional models. Here we present a new Stochastic Heterogeneous Epidemic Model (SHEM) to investigate the role of heterogenous societal structure. SHEM is intended to be a general tool with which to explore scenarios and determine the expected consequences of various interventions. We represent a society by an arbitrary network of sub-populations that could represent social as well as geographical strata. We created several scenarios with large clusters of people with R_0 of COVID-19 interacting with multiple smaller local clusters that have larger internal R_0. We find that isolation or embedding of these vulnerable sub-clusters generate complex infection patterns which includes multiple peaks and growth periods, an extended plateau, a prolonged tail, or a delayed second wave of infection, which may or may not form due to stochasticity. We also show that local clusters can either be driving or driven forces in infection progression. Embedded vulnerable groups become hotspots that drive infection despite efforts of the main population to socially distance, while isolated areas suffer delayed but intense infection. Social heterogeneity is a key factor in the formation of complex infection curves. Vulnerable subgroups that cannot implement mitigation strategies can spread infection to socially distanced populations, defeating mitigations. This implies that mitigation of vulnerable groups is essential to control the epidemic.

Significance Statement:

We developed a new multiscale Stochastic Heterogeneous Epidemic Model (SHEM) and demonstrated major roles for social heterogeneity and stochasticity in pandemic development. We simulated viral infection in a theoretical society where small communities that cannot socially distance link to large clusters representing urban populations. Depending on the model parameters, our simulations of COVID-19 infection generated a large variety of dynamic patterns, including multimodal growth periods observed now in the US and worldwide. Infection of small, vulnerable clusters of people defeated mitigation efforts by the main population. The importance of protecting vulnerable subgroups suggests policy implications. Our abstract model could be applied at multiple scales of human societal organization.

Introduction

Coronaviruses represent one of the major pathogens that primarily target the human respiratory system. Previous outbreaks of coronaviruses (CoVs) that affected humans include the severe acute respiratory syndrome (SARS)-CoV and the Middle East respiratory syndrome (MERS)-CoV (1). COVID-19 is a disease caused by the novel coronavirus SARS-CoV2 virus that is both fatal and has a high transmission rate (R_0), almost twice that of the 2017-2018 common influenza (2, 3). The World Health Organization stated that this combination of high health risk and susceptibility is of great global public health concern, and efforts must be directed to prevent further infection while vaccines are still being developed (4). As of August 2020, there are more than ten million confirmed COVID-19 cases with more than a half of million deaths. Older adults seem to be at higher risk for developing more serious complications from COVID-19 illness (5, 6). In today's absence of a vaccine and impactful treatments, the most effective way to combat the virus is to find and implement mitigation strategies for susceptible populations. An invaluable resource in this difficult task is numerical modeling studies that can reveal key factors in pandemic development.

What models could be useful? Direct study of the available data of COVID-19 is complicated because many cases and deaths are underrepresented. However, a simple model that correctly captures large-scale behaviors, but gets some details wrong, is useful, whereas a complicated model that gets some details correct but mischaracterizes the large-scale behaviors is misleading (7). Previously, during the H1N1 pandemic, generic (i.e. non-specific) stochastic influenza models were an important tool to understand and quantify the full effects of the virus in simulations of important scenarios (8). Open source stochastic models such as FluTE (2010) or

GLEaM (2011) were developed to simulate the spatial interaction and clusterization of millions of people (9, 10) to discover epidemic patterns. Now with respect to COVID-19, the FluTE model has recently been used to offer interventions to mitigate early spread of SARS-CoV2 in Singapore (11). Furthermore, GLEaM was also adopted by Chinazzi et al. to model the international propagation of COVID-19 to gain insight into the effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak.

However, complex infection curves are visible in real data in various regions around the world as of August 2020, distinguishing themselves from simple curves generated by models such as SEIR (12). Despite extensive efforts to understand and predict the COVID-19 spread, the key factors that determine the multimodal rise patterns, the asymmetry of the recovery phase, and the emergence of a distinct second wave remain unclear. A novel multi-region pandemic model (LIST) of Friston et al. (13) which incorporates regional factors, suggests that a second wave of infection typically occurs when a population is re-exposed to a virus by an influx of infected people from another region. In a separate study by Althouse et al (14), the transmission of disease is shown to be stochastic, and is commonly dominated by a small number of individuals, driven by super-spreading events.

In principle, agent-based models like FluTE that track the fate of every individual ought to be the most precise. However, these models require detailed statistical information about the social interactions and grouping of individuals. This information is well characterized for seasonal influenza, but has been completely disrupted by mitigations for COVID-19 and continues to be in flux, changing in unpredictable ways that cannot be anticipated. Therefore, instead of a data-based forecasting model like those described above, we chose to develop a

scenario model that can be used as a tool to study the consequences of a set hypothesis-driven conditions in a user-specified network of populations, which could be not only geographical but also social subgroups. Indeed, one underexplored but important factor of pandemic spread is regional heterogeneity which is characterized by various subpopulation clusterization, societal interaction, and disease mitigation strategies. Our hypothesis is that complex infection curves that consist of multiple infection peaks and growth periods are the consequence of asynchronous propagation of infection among groups with widely varying degrees of intra-group interaction and isolation from main hubs (a metapopulation of infections). We also attempt to take account of over-dispersed stochasticity (14), which is usually not incorporated into compartmental models but can be critical in small or virgin populations.

To approach this problem, we developed a novel Stochastic Heterogeneous Epidemic Model (dubbed SHEM) which incorporates heterogeneous aspects of society. The model was inspired by our stochastic models of local calcium release dynamics inside heart cells, driven by explosive calcium-induced-calcium-release (15, 16). As a demonstration of this tool, we examine several key scenarios of heterogeneity where separate communities of various clusterization and transmission capabilities are linked to a large population cluster. For the purposes of investigating COVID-19, the basic reproduction number of infection (R_0) of the bulk of our population was assigned to $R_0 = 2.5$ which is within the range of SARS-CoV-2 basic reproduction number based on the early phase of COVID-19 outbreak in Italy (17). The mean duration of infection (infectious period) was taken to be 7 days and the incubation period 5.5 days. Interplay of various degree of heterogeneity and isolation periods in our model generated various dynamic patterns of infection, including a multi-modal growth periods, an extended plateau, prolonged tail, or a delayed second wave of infection.

Results

Simulations of infection in isolated clusters driven by an urban cluster

In the first set of simulations we examined the virus spread in simple hypothetical scenarios with equal numbers of individuals in urban and isolated populations (Fig. 1A, insert). The large urban cluster was composed of 1 million individuals set to R_0=2.5 (open level, but changing throughout the simulation). The isolated population consisted of 250 clusters, each with 4000 +/- 500 people and with the same internal R_0=2.5 that remained constant throughout all simulation stages. The urban cluster was weakly connected with 0.001% transient contact into the isolated clusters (*alphainpop*) while isolated clusters had 0.1% contact into the urban cluster (*alphaoutpop*), see Methods for the definition of transient contact. This can be visualized as a collection of small suburban neighborhoods or nursing homes that are attempting to isolate themselves from the city. We investigated the 4 scenarios, specified below. In each scenario except #1, the urban cluster closed to R_0=1.25 at t=40 days (closed level, e.g. this was New York City under lockdown, based on 21% antibody positive tests at the peak (18)).

1) No mitigation, i.e. freely expending pandemic: The large cluster of individuals stays always open.
2) Premature, partial reopening to R_0 = 1.9 at 100 days.
3) Moderate lockdown period with full reopening at 225 days to R_0 = 2.5.
4) Long lockdown period with full reopening at 365 days to R_0 = 2.5.

A general tendency throughout all 4 scenarios was that as the lockdown period increased, the magnitude of the infection decreased but its duration increased. At the same time the interplay of the urban cluster and the isolated clusters generated a variety of specific patterns in virus spread dynamics.

In the first “no mitigation” scenario (Fig. 1A) the isolated areas generated a strong second peak at the time when infection in the urban cluster had gone through its peak and was decaying. The second peak substantially extended the overall span of the pattern, nearly twice as much.

The infection rise in the “premature reopening” scenario (Fig. 1B) was multi-modal. The initial rise of infection substantially slowed in the main cluster during the closed period, but a contribution from the isolated areas becomes notable closer to the partial reopening at day 100 (Fig. 1D, inset). In the reopen period, the infection surged in both the isolated and urban clusters. The peak of the isolated clusters happened later than the urban cluster, creating an apparent plateau in active infection cases from day 175 to 225.

The infection dynamics in the “moderate lockdown” scenario (Fig. 1C) was more complex. During the closed stage (of urban center), while infection in the urban cluster declined, the delayed infection in isolated clusters continued to rise forming an additional peak in total infections (Fig. 1E, inset). Yet another peak in total infections emerged in the reopen stage that is generated mainly by the urban cluster, but echoed by the isolated subpopulations.

The “late reopening” (Fig. 1D), decreased infection during the first wave in both urban and isolated clusters, but resulted in a distinct delayed second wave of infection. This second wave pattern provides a delay during which the respective second wave of deaths could be intercepted and prevented with the timely development of a vaccine or effective treatment (for example a year from the time of infection onset in Fig. 1E).

We also performed a control simulation to validate that heterogeneity of isolated clusters is indeed important for the infection pattern. In the most complex scenario of “moderate lockdown” shown in Figure 1C we substitute 250 clusters by one big cluster with the same population of one million people keeping all other parameters the same. The simulations showed

a different pattern in which the second big cluster always generated a peak of substantially larger amplitude (Fig. S1).

Simulations of integrated clusters driving infection in an urban cluster

By altering parameters in the same topology as Figure 1A, we found that the outlying clusters, if they are unable to socially distance, can become potential "hotspots" that can drive the infection in the urban population even against efforts of the latter to lock down. In this scenario the large urban cluster was composed of 1 million individuals with R_0 = 1.25 throughout all simulation stages while the highly susceptible population consists of 250 clusters each with 1200 +/- 500 people and internal R_0 = 3.0 that are partially embedded in the urban cluster. This R_0 value is based on data from four districts in Germany when essential manufacturing sectors were open – 95%-prediction interval: 2.16 – 3.73 (19). The potential hotspot clusters were connected with 20% out-coupling into the urban cluster (*alphaoutpop = 0.20*, see Methods*)*. This mechanism of transient contact implements short-term movement of the same people in and out regularly, which does not dilute the effect of the conditions in hotspots the way that random bidirectional migration would. In other words, the same people "virtually" move back and forth, but spend most of their time in the high- R_0 locations where the infection regenerates. In this scenario, the small number of infections in the urban area are picked up by hotspots, amplified, and then drive a wave of infection among the urban population despite their efforts to keep their internal R_0 at 1.25 by social distancing.

We performed 10 runs of these simulations which demonstrated that the integrated clusters drove infection in the urban cluster as shown in a typical example in Fig 2A, B, leading the late appearance of the epidemic in places that had seen few cases in a microcosm of the

pattern. The stochastic nature of infection in individual hotspots is shown in Figure 2C; many hotspots "explode" as early as between 50 and 100 days (Movie S1). Hotspots substantially increased the peak of infection and shifted it towards much earlier occurrence from about 400 days to 200 days (Fig 2D).

In the second "chain" topology multiple small urban areas (population 100K each) are sequentially connected and 30 potential hotspots with R_0=2.0 drive infection within in each urban cluster and facilitate propagation from cluster to cluster (Fig 3; Supplementary figure S2 and Movie S2 show the stochastic dynamics of individual hotspots). In this model, the first cluster began with R_0 = 2.5, then locked down to 1.25 at day 40, while the unsuspecting clusters down the road kept R_0 = 1.05 throughout, signifying low population density and efforts at social distancing, which were defeated by the hotspots picking up the small number of arriving infections and amplifying them. In the case where the hotspots were closed, infection generally didn't reach the last urban cluster at all in the absence of amplification along the way.

Stochasticity can be a critical factor in infection spread

The previous simulations showed how stochasticity has substantial influence on infection dynamics when small subpopulations are involved. We performed additional simulations to illustrate that stochasticity can influence not only the timing and amplitude of the infection curve within a particular pattern, but critically influence the pattern per se, when the system balances at the edge of criticality. Figure 4 shows a simulation of a single urban area of 1 million population under prolonged lockdown followed by late reopening after the epidemic appeared to be over. Initially R_0 = 2.5 locked down to R_0 = 1.25 in the closed stage on day 70, and reopened to R_0 = 2.5 on day 340. A variably delayed second wave of infection occurred only in half of the

simulations. Closer examination of the nadir (figure 4, bottom panel) shows that the epidemic can escape extinction even when of the order of 10 infected individuals remain.

Discussion

As of August 2020, the infection curves of the COVID-19 pandemic in various locations have been very different from standard smooth bell curves. We hypothesized that multiple, asynchronous waves and plateaus are in part due to stochasticity and excess heterogeneity, as well as changing efforts at mitigation. Geographic heterogeneity is included in forecasting models (13, 20, 21) which use extensive, public databases of population characteristics and travel patterns, but these do not fully account for the stratification of the social behavior that controls the spread of the virus. One sophisticated forecasting model, on the basis of an early antibody survey, predicted in April that California would achieve herd immunity within a few weeks (22), an expectation that was alarmingly refuted by infection numbers on the first week of July.

For these reasons, instead of building a data-based forecasting and estimation model, we chose to develop a scenario-modeling tool that can be used to explore mechanisms of infection dynamics in idealized models of social stratification constructed by the investigator. The model is built as a network of "populations" which could represent be social and behavioral strata of geographic populations. Our model could therefore be considered a metapopulation of SARS-CoV2, when a single species is spread among different environments that determine its local survival or extinction.

An epidemic can be likened to a forest fire, which spreads by diffusion along a front, but can also jump by embers that may or may not start a new blaze. Such spread to virgin areas, with a virus as with a fire, is intrinsically stochastic and such stochasticity, which is not explicitly

included in mean-field models, may contribute to the remarkable patchiness of the COVID-19 epidemic becoming increasingly evident as of August 2020. This has caused the epidemic to appear entirely different to observers in different locations, leading to politicization of the response, which is, itself, a form of social heterogeneity. For rare spread to small, isolated subgroups (embers) this stochasticity is crucial. Patchiness is aggravated by the over-dispersion (super-spreading) of secondary cases of COVID-19, where the majority of infected individuals do not spread the virus, but some can cause up to a hundred secondary infections (14). We therefore built a model that is explicitly stochastic, with a mechanism to account for over-dispersion, which is not straightforward in a time-dependent evolution model.

As a demonstration of this general model, we simulated several scenarios, having in common the presence of multiple small groups that are unable/unwilling to socially distance and so have high internal R_0 (vulnerable groups). These were linked to one or more large urban populations, and depending on their degree of interaction, they can either be driven by infection from the main population, or can act as major drivers of the epidemic.

Isolated subpopulations (*e.g.* nursing homes, prisons, remote suburbs, cloistered religious groups) can have a substantially delayed contribution to total infection cases, ultimately forming an infection curve with various multi-modal growth periods or plateaus (Fig 1). These communities are infection-driven, and due to their isolated nature, have low herd immunity that puts them at risk for explosive scenarios if basic mitigation strategies are not implemented.

Alternatively, vulnerable subpopulations can be partially integrated into the main group (employees of factories, warehouses, meat packing plants, church groups, campuses, shelters, and other essential workers) with substantial and possibly asymmetrical contact, and yet maintain their identity and potential to explosively regenerate the infection. We found that these

potential "hotspots," can actually drive the spread of COVID-19 in the urban population by picking up infection and amplifying it (Fig 2 and 3, movies S1, S2). Hotspots can unpredictably ignite infection even in a locked down population and then propagate and ignite other isolated populations. In general, the results show that infection can spread through socially distanced populations when aided by vulnerable subgroups that cannot implement mitigation strategies. Heterogeneous vulnerability of subgroups has been widely demonstrated in COVID-19 infection data. Mitigating the susceptibility of such groups will be needed to control the epidemic. This is consistent with the mitigation studies of Chao et al. using an agent-based model (23).

Limitations and future studies

We deliberately chose to construct a general, abstract model of heterogeneous, interacting populations as a tool to study mechanism of spread of the epidemic. However, even the idealized and simplified scenarios shown above as an example suggested important considerations for real-world mitigation. More realistic scenarios could straightforwardly be constructed from observational data, provided the required information about subgroup R_0 and topology could be extracted, and the model could be applied to specific scenarios at various levels in the form of a fractal-like structure, i.e. large countries surrounded by small isolated countries, or large highly-populated states (or provinces) surrounded by smaller isolated states, or big counties or cities surrounded by suburban areas. In other words, while the COVID-19 pandemic is a complex multiscale problem, our model can be applied at any given scale and incorporate specific heterogeneous structure relevant at that scale.

In the present version of the model, we do not distinguish between symptomatic and asymptomatic infections, since the relative infectivity of the two is uncertain and controversial,

and there is no direct way to determine the number of asymptomatic infections. We also did not take account of recent suggestions that infectivity is concentrated in a short time window just before and after symptom onset. We used the standard SEIRD assumption that infections are generated throughout the period of infection, whose clinical mean duration (7 days) was used. If infectivity is front-loaded that would have an effect similar shortening this period, which would accelerate the time course of the epidemic. We also did not take account of the fact that many recovered patients do not quickly re-enter their normal social circles either due to prolonged symptoms or prejudice, delaying their contribution to herd immunity. Any of these issues could be dealt with by elaborating the model, but in keeping with our philosophy of studying simplified situations to draw qualitative conclusions, we deferred these improvements. We also did not consider the controversial evidence regarding the physical mechanisms of transmission of COVID-19; these enter the model only as black-box parameters, which are not well determined by present data. For this reason, we did not discuss the most pressing example of vulnerable subgroups: schools, for which these parameters are especially sketchy.

We implemented a novel mechanism to account for over-dispersed secondary infections (super-spreading) within a time-evolution scheme, by keeping a partial history of individual infections. This is not as accurate as tracking each individual in an agent-based model, but much less demanding computationally. It does have the disadvantage that it tags each individual with his number of secondary infections (most common: 0) which requires special consideration when individuals create infections in a new location. A better method is needed for distinguishing among types of super-spreading events, both in the model and in real data.

Methods

Model purpose

In view of the constantly changing behavioral environment for COVID-19 in the United States, data-based predictive modeling of the future of the epidemic is difficult. Our model is intended to be a general tool with which to explore scenarios and determine the expected consequences of various interventions. The model is specifically intended to examine the effect of heterogeneity, including not only geographic but also social heterogeneity, i.e. the existence of groups within one geographic location that have different social interaction patterns and may be partially isolated from neighboring groups, *e.g.* nursing homes, prisons, campuses. Alternatively, subgroups can be partially embedded in the main population, e.g. meat processing plants or warehouses who are unable to socially distance at work, but spend part of their day in the main community where they can acquire and amplify infection, defeating community mitigation efforts. The model is fully stochastic and, unlike most compartmental models, incorporates the effect of over dispersion of secondary infections (super spreading).

Structure of the Model

The general model consists of a number of subpopulations ("villages") whose number is limited only by memory. The simulation is based on a generalization of the SEIRD representation. The state of each village is represented by the numbers of individuals in each of 5 states: *Susceptible*, *Exposed* (destined to become infected), *Infected*, *Recovered* (immune) and *Dead* (however, see below under Super-spreading for additional state-dependence). Each village is, by definition, homogeneous and mixed. Villages could represent actual geographic units, but could also be groups or sub-regions that have different social interactions or behavior.

Each village J is characterized by its population, the expected mortality of virus infections, and its local value RINN(J) of the basic reproduction number R_0. R_0 is defined as the mean expected number of secondary infections spawned by one infected individual over the duration of their illness, *if the population were totally susceptible*. It is a property of both the virus and the behavior of individuals in the population, but is distinct from *R(t)*, the realized, time dependent, reproduction number that depends also on the fraction of susceptible individuals remaining during the epidemic.

Villages are connected by a user-specified network of unidirectional links along which infection or individuals can travel at user-specified rates, including links from each village to itself to represent internal infection/recovery processes. Infection can spread by two processes: transient contact between groups (*alpha* process) *e.g.* nursing home staff coming from the city, or actual migration of individuals from one village to another (*beta* process). Each non-self-link is characterized by 4 user-supplied parameters: *alphain* and *alphaout* describe the degree of transient contact (see below) along or against the direction of the link respectively; *betain* and *betaout* are rates of migration of individuals (time^{-1}).

Transient Contact (alpha) Process

Infection transmitted by transient contact is modeled as though members of one village spend some (small) fraction *alpha(in/out)* of their time (*i.e.* of their inter-personal contacts) "visiting" the opposite village at the other end of the link, adjusted for any mitigations (an example would be staff working at a nursing home, or meat-packing plant employees, treated as a separate, high-risk population but living in the surrounding county). The spread of infection in

each direction of the link has two components: (1) exposure of susceptibles by visiting infectious individuals and (2) exposure of visiting susceptibles in the visited village, who then carry the infection back to their village. This formulation allows for the possibility that transmission is asymmetric. The generation of exposure by these "visitors" at home and abroad is scaled so that each infected individual, generates (in an otherwise susceptible population) his destined number of secondary cases (see below under super-spreading).

This arrangement allows for the possibility that "visitors" from different villages could cross-infect while visiting a common hub (picture UPS and FEDEX drivers) even if there is no direct link between them. To represent this process, "virtual links" are generated between pairs of physical links that meet in a hub (in graph-theory terms these are links of the adjoint graph of the network). Infection by this indirect process is second order in the alpha's so it makes very little contribution in the case of highly isolated sub-populations (*e.g.* nursing homes, prisons) but could be important for embedded sub-populations with high contact with the hub.

Simulation Method

The entire collection of populations is simulated as a single, continuous-time Markov chain (birth-death process). There are 14 types of possible events associated with each link:

1) Infection from source to target by transient contact
2) Infection from target to source by transient contact
3) Infected individual moves from source to target
4) Exposed individual moves from source to target
5) Susceptible individual moves from source to target

6) Infected individual moves from target to source
7) Exposed individual moves from target to source
8) Susceptible individual moves from target to source
9) Susceptible gets exposed inside village (self-link only)
10) Exposed converts to infected inside village (self-link only)
11) Infected recovers inside village (self-link only)
12) Infected dies inside village (self-link only)
13) Recovered moves from source to target
14) Recovered moves from target to source

At each step, rates of these 14 possible events (computed as in classic SEIRD differential equations, except as below under Super-spreading) are summed over all links in the network to give a total transition rate R_{tot}. A uniformly distributed random number rn is generated and $-log(rn)/R_{tot}$ is taken to be the waiting time until the next event, considered as a Poisson point process. Time t is advanced by that amount and then a particular transition is selected by a second random number rn_2 by partitioning R_{tot} into segments proportional to their relative rates and finding which segment contains $rn_2 * R_{tot}$. The action associated with the prescribed event (*e.g.* increment/decrement numbers in states SEIRD of the villages connected by the link) is taken, and then the above is repeated for the next step. This is continued until t reaches t_{max} specified in the input parameter file or *tswitch*, the time set for a discrete change in parameters. Output is produced whenever t crosses *tout*, which is then incremented to t plus a user-specified interval (usually one day). This time-binning is only for output and plays no role in the continuous-time evolution of the epidemic.

Super-spreading

It is known that the distribution of secondary COVID-19 infections generated by a single, infected individual is over-dispersed (*i.e.* has a long tail compared to the Poisson distribution of infections expected if transmission were random). Although the average R_0 is estimated to be 2.5-4 in the absence of social distancing mitigations, contact tracing has shown that single individuals have infected up to a hundred others. This is known as super-spreading events, and can occur by several possible mechanisms, involving either a predilection of an individual (*e.g.* a celebrity who travels widely and contacts many other people) or a situation in which individuals were placed in unusually close contact (*e.g.* a church choir in an indoor location). On the other hand, the majority of infected individuals do not appear to spread the infection to anyone. It has been shown (14) that this over-dispersed distribution can be approximated by a negative binomial distribution, with mean R_0 (by definition) and dispersion parameter $r<<1$, for example 3 and 0.16. By iterating this distribution for several generations of viral spread, it is found that the eventual distribution of epidemic size is predicted to be quite different than found for a hypothetical stochastic transmission by Poison-distributed secondary infections with the same R_0. A recent model of contact tracing assumed, based on data from the Netherlands, that the distribution of number of personal contacts outside the family is distributed as a negative binomial and used this to generate random changes to infection levels at 1-day intervals (24).

Unfortunately, viral generations do not remain synchronous in time, so it is not straightforward to incorporate super-spreading in a time-dependent epidemic evolution model except by following the interactions and infections of each individual in the population, as done for example in the FLuTE simulation for influenza (8). This is very compute-intensive, but a

more significant objection from our point of view is that it depends on knowing (statistically) the social interaction groups and travel behavior of the population at a fine-grained scale, and these have been severely disrupted by mitigation efforts during the current pandemic. Rather than speculate on these variables, we have developed a modified Markov scheme that tries to reproduce the observed distribution of secondary infections by replacing R_0 in the event-rate calculations by an infectivity that is itself stochastic.

Supposing that an individual produces, *a posteriori*, K secondary infections over the lifetime t_{dur} of his infection, the required rate is K/t_{dur} . However, t_{dur} is itself stochastic: in our model, as in simple SIR models, recovery is treated as a Poisson point process with a rate $1/trec$ per infected individual where *trec* is the observed mean duration of infections (it is not presently known whether that is true of asymptomatic infections). When there are multiple infected individuals present, if recovery events in the population are random with a rate $k_i/trec$, where k_i is the number of infected individuals in the village, a super-spreader is likely to be "recovered" before (or after) doing his "job". To avoid this, we have adopted the following scheme:

- In each village j, at each event, an infectivity *inf(*j*)* is maintained that takes the place of k_i*R_0 in the SEIRD rate equations.
- Whenever a new infection is created (by conversion of an exposed individual), a random number K is drawn from a negative binomial distribution of mean R_0 and dispersion r_{eff} , the latter to be determined. *Inf* is incremented by K and the individual infectivity K is placed on the top of a linked list.
- Whenever a random recovery event is generated at the above-mentioned rate, the oldest individual infectivity is removed from the bottom of the list and subtracted from *inf*.

The overall rate of infection events, based on *inf*, will be a superimposition of the rates of the individual infection events generated by each infected individual. The mean number of secondary infections actually realized by one infected individual over the life of his infection will be $K*t_{dur}/trec$. Since infections recover in the order in which they were created, if there are n infections active, t_{dur} will be the nth waiting time of the Poisson point process of rate $n/trec$ (where $n=k_i(j)$), whose probability distribution is proportional to $poisson(n-1,n/trec)$. The individual infections generated by individual K are a Poisson point process, so the probability that the individual actually generates j secondary infections is $poisson(j, K*t_{dur}/trec)$. Integrating this over the distribution of t_{dur}, multiplying by the probability mass distribution of the negative binomial distribution and summing over K we find:

$$p(j,n) = \frac{n^n \Gamma(n+j) r^r \sum_{k=0}^{\infty} \frac{r_0^k k^j (n+k)^{-n-j} (r+r_0)^{-r-k} \Gamma(r+k)}{\Gamma(k+1)}}{j\Gamma(j)\Gamma(n)\Gamma(r)}$$

as the distribution of the actual, realized number of secondary infections. This is a long-tailed probability distribution that can be fit, by an appropriate choice of r_{eff}, to approximate the empirical negative binomial distribution with $r=0.16$ over the relevant range. With more than a few active infections present, the distribution converges to:

$$p(j,inf) = \frac{n^n \Gamma(n+j) r^r \sum_{k=0}^{\infty} \frac{r_0^k k^j (n+k)^{-n-j} (r+r_0)^{-r-k} \Gamma(r+k)}{\Gamma(k+1)}}{j\Gamma(j)\Gamma(n)\Gamma(r)}$$

We choose r_{eff} to give the best least-squares fit on a linear scale of the case $n=1$, which is the most important stochastic case since it governs the chance that a single infected individual

can start an outbreak, giving the chance that an infected individual causes no secondary infections, $p(0,1)=0.62$ similar to the empirical distribution. These distributions are all normalized and have mean R_0 and differ dramatically from the Poisson distribution (Fig. 5, dashed line) assumed in the classic SEIR model. Larger values of n are decreasingly important because the aggregate distribution of the actual infection rate controlled by the sum *inf* behaves similarly to *negbinomial* $(R_0, n*r)$ which converges to Poisson so stochastic effects become less important once there are many active cases.

Super-spreaders vs super-spreading events

Super-spreading can be a property of the individual or of the circumstances. What happens when an individual infected patient migrates to a new village? Does he keep his identity or does he assume the infectiousness typical of the local R_0 of his new environment? In the model we can make the choice is determined by a logical variable SPREADR (default TRUE. controlled in the demos by the input parameter *spreads*). If SPREADR is true, a migrant keeps his prior K value which simply migrates from the top (newest) link to be added to the top of the infection list in the new village, thereby preserving his infectious lifetime in his new home. If *SPREADR* is false then the K value of migrants is re-randomized using the local R_0 and r_{eff} and the infectivity of transient visitors in the alpha process is re-scaled to the local value of R_0. In the current version of the program, SPREADR is a single variable governing all events, but it could easily be made specific to individual links to distinguish groups that are vulnerable due to high density in their home village (e.g. factory or warehouse) versus groups that are intrinsically super-spreaders due to their individual behavior.

Software Considerations

The model software is written in Fortran 77/90. The main simulation engine, described above, is in the form of a single Fortran module SIMULATOR. It is intended to be driven by a front-end program that sets up the distribution of the populations, topology of the network, and parameter changes at discrete time points, and connects to SIMULATOR by calling subroutine EPISIM (28 variables). Ideally, the front end should use some kind of scripting language based on network concepts. For purpose of these demonstrations, we hand-coded a front end – epichainF, describing a chain of urban clusters (or a single cluster) connected by bidirectional travel, each linked to a large set of small subpopulations whose characteristics differ from the urban cluster.

The single Markov-chain structure of the model is intrinsically serial, and is implemented in a single processor thread. For configuration with a large number of links sharing a common hub, the large number virtual links can make this slow. The main inner loop will probably be parallelized in future versions.

Acknowledgments

The work was supported by the Intramural Research Program of the National Institute on Aging, National Institute of Health. This study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD. (http://biowulf.nih.gov).

References

1. Rothan HA & Byrareddy SN (2020) The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *J Autoimmun* 109:102433.
2. Xiaofang F*, et al.* (2019) A Severe Seasonal Influenza Epidemic During 2017–2018 in China After the 2009 Pandemic Influenza: A Modeling Study. *Infectious Microbes & Diseases* 1(1):20-26.
3. Liu Y, Gayle AA, Wilder-Smith A, & Rocklov J (2020) The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J Travel Med* 27(2).
4. Organization WH (2020) Coronavirus disease 2019 (COVID-19) Situation Report – 72. *https://apps.who.int/iris/bitstream/handle/10665/331685/nCoVsitrep01Apr2020-eng.pdf*.
5. Liu K, Chen Y, Lin R, & Han K (2020) Clinical features of COVID-19 in elderly patients: A comparison with young and middle-aged patients. *J Infect* 80(6):e14-e18.
6. Li P*, et al.* (2020) Clinical features and short-term outcomes of elderly patients with COVID-19. *Int J Infect Dis* 97:245-250.
7. Siegenfeld AF, Taleb NN, & Bar-Yam Y (2020) Opinion: What models can and cannot tell us about COVID-19. *Proceedings of the National Academy of Sciences*:202011542.
8. Chao DL, Halloran ME, Obenchain VJ, & Longini IM, Jr. (2010) FluTE, a publicly available stochastic influenza epidemic simulation model. *PLoS Comput Biol* 6(1):e1000656.
9. Van den Broeck W*, et al.* (2011) The GLEaMviz computational tool, a publicly available software to explore realistic epidemic spreading scenarios at the global scale. *BMC Infect Dis* 11:37.
10. Balcan D*, et al.* (2010) Modeling the spatial spread of infectious diseases: the GLobal Epidemic and Mobility computational model. *J Comput Sci* 1(3):132-145.
11. Koo JR*, et al.* (2020) Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. *Lancet Infect Dis* 20(6):678-688.
12. Weiss H (2013) The SIR model and the Foundations of Public Health. *MATerials MATemàtics* 3:1-17.
13. Friston KJ*, et al.* (2020) Second waves, social distancing,and the spread of COVID-19 across America. *https://arxiv.org/abs/2004.13017*.
14. Althouse BM*, et al.* (2020) Stochasticity and heterogeneity in the transmission dynamics of SARS-CoV-2. *https://arxiv.org/abs/2005.13689*.
15. Stern MD*, et al.* (2014) Hierarchical clustering of ryanodine receptors enables emergence of a calcium clock in sinoatrial node cells. *J Gen Physiol* 143(5):577-604.
16. Maltsev AV, Maltsev VA, & Stern MD (2017) Clusters of calcium release channels harness the Ising phase transition to confine their elementary intracellular signals. *Proc Natl Acad Sci U S A* 114(29):7525-7530.
17. Rahman B, Sadraddin E, & Porreca A (2020) The basic reproduction number of SARS-CoV-2 in Wuhan is about to die out, how about the rest of the World? *Rev Med Virol*:e2111.
18. Schultz C (2020) More than one in five people who were tested for virus antibodies in N.Y.C. had them. *The New York Times: https://www.nytimes.com/2020/04/23/nyregion/coronavirus-new-york-update.html*.
19. The-Robert-Koch-Institute (2020) COVID-19 Situation Report 21/06/2020. *https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/2020-06-21-en.pdf*.
20. Chinazzi M*, et al.* (2020) The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* 368(6489):395-400.

21. Pei S, Kandula S, & Shaman J (2020) Differential Effects of Intervention Timing on COVID-19 Spread in the United States. *https://www.ncbi.nlm.nih.gov/pubmed/32511526*.
22. Friston KJ*, et al.* (2020) DYNAMIC CAUSAL MODELLING OF COVID-19. *https://arxiv.org/abs/2004.04463*.
23. Chao DL, Oron AP, Srikrishna D, & Famulare M (2020) MODELING LAYERED NON-PHARMACEUTICAL INTERVENTIONS AGAINST SARS-COV-2 IN THE UNITED STATES WITH CORVID. *https://www.medrxiv.org/content/10.1101/2020.04.08.20058487v1.full.pdf*.
24. Kretzschmar ME*, et al.* (2020) Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. *Lancet Public Health* 5(8):e452-e459.

Figure legends

Fig. 1: Complex dynamic patterns of SARS-CoV2 infection in simulations in a heterogeneous society when infection in isolated clusters are ignited by an urban cluster implementing various lockdown strategies. **A,** isolated clusters generate a second delayed peak when no intervention is implemented. Inset schematically illustrates the society structure in this scenario. Contributions are shown by different colors. **B,** an apparent plateau after early reopening and complex rise pattern during close period (inset). Green shade shows the lockdown periods. **C,** A multimodal rise (inset) with additional peak generated by rural cluster after full reopening at day 225. **D,** A delayed second wave emerged after full reopening at day 365. **E,** The dynamics of total number of deaths in each scenario (A-D).

Fig. 2: Highly susceptible integrated clusters (hotspots) drive SARS-CoV2 infection in an urban cluster. A and B, Initial rise of infection in hotspot clusters is followed by the infection in urban cluster with a delay of about 30 days. Y-axis represents active infections in % population reflecting for hotspots (red line) the ratio of all active cases in all hotspots to entire population of all 250 hotspots. Inset shows schematically the society structure in this scenario. **C,** Infection in individual hotspots (multiple colors) substantially fluctuates in terms of time of ignition and magnitude from the mean (red bold curve). See also Movie S1. **C**, Explosive infection in hotspots within locked urban cluster substantially increased the peak of infection in the entire society and shifted it towards much earlier occurrence from about 400 days to 200 days. Shown are 10 simulation runs for each scenario.

Fig. 3: Complex infection propagation patterns in multiple urban areas containing hotspots. A, Schematic illustration of the heterogeneous society used in simulations. **B,** Total infection count oscillates as infection propagates. While individual oscillations exhibit substantial variations in timing and amplitude, the patterns remain the same (i.e. 4 oscillations, reflecting infection surge in each urban cluster). **C,** The infection in hotspots is delayed before the lockdown at day 40, but then is always in the lead (red curves), driving infection in each urban cluster (blue curves) and facilitating infection propagation among clusters (Movie S2).

Fig. 4: Stochasticity can be a critical factor in the formation of complex infection patterns including secondary waves of infections. A, Simulations of infection dynamics in an urban cluster with 1 million population that closes at day 70 and reopens at day 350. The second wave of infection occurs with variable delays in only 50% of simulation runs (5 out of 10). **B,** Finer details of the infection dynamics near criticality.

Fig. 5: The distribution of secondary infections generated by infectious individuals. Black: Observed negative binomial distribution (14); Green, blue, magenta, red: The actual realized number of secondary cases generated over the lifetime of one infection in the presence of n other infections individuals according to our scheme. All distributions have mean R_0 =3.0. Dashed line: Poisson distribution with mean R_0 as implicit in mean-field SEIR models.

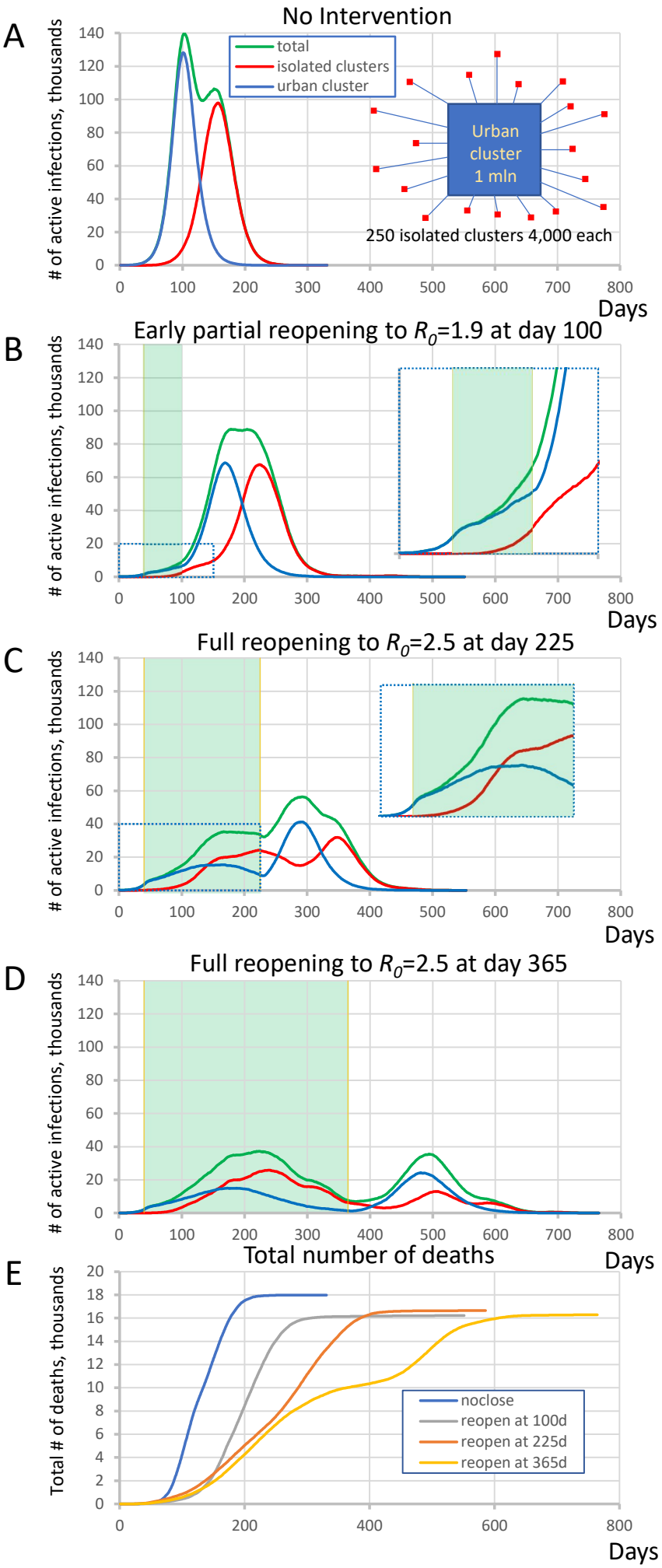
A
No Intervention
total
isolated clusters
urban cluster
Urban cluster 1 mln
250 isolated clusters 4,000 each
# of active infections, thousands
Days
B
Early partial reopening to R_0=1.9 at day 100
C
Full reopening to R_0=2.5 at day 225
D
Full reopening to R_0=2.5 at day 365
E
Total number of deaths
Total # of deaths, thousands
noclose
reopen at 100d
reopen at 225d
reopen at 365d

Figure 1

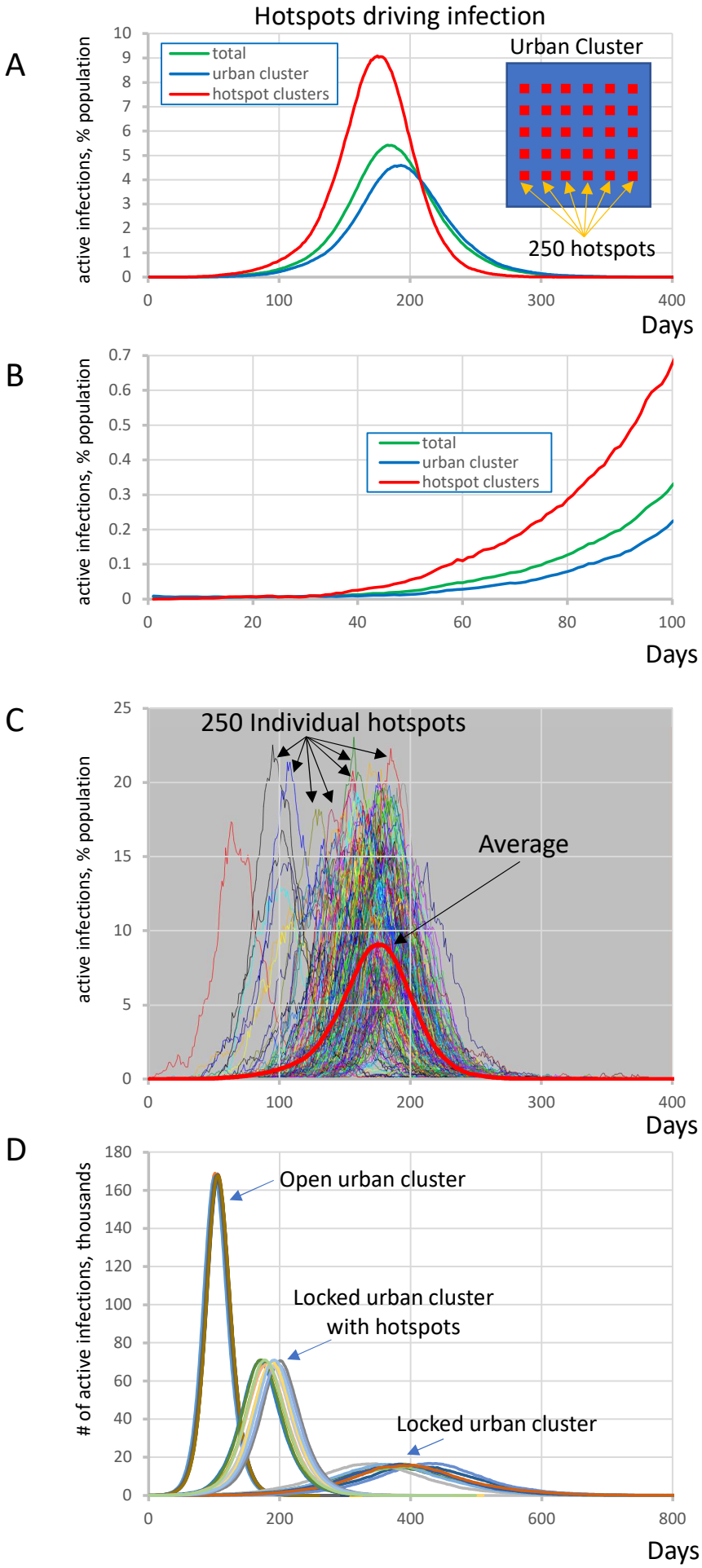
A
Hotspots driving infection
active infections, % population
total
urban cluster
hotspot clusters
Urban Cluster
250 hotspots
Days
B
active infections, % population
total
urban cluster
hotspot clusters
Days
C
active infections, % population
250 Individual hotspots
Average
Days
D
# of active infections, thousands
Open urban cluster
Locked urban cluster
with hotspots
Locked urban cluster
Days

Figure 2

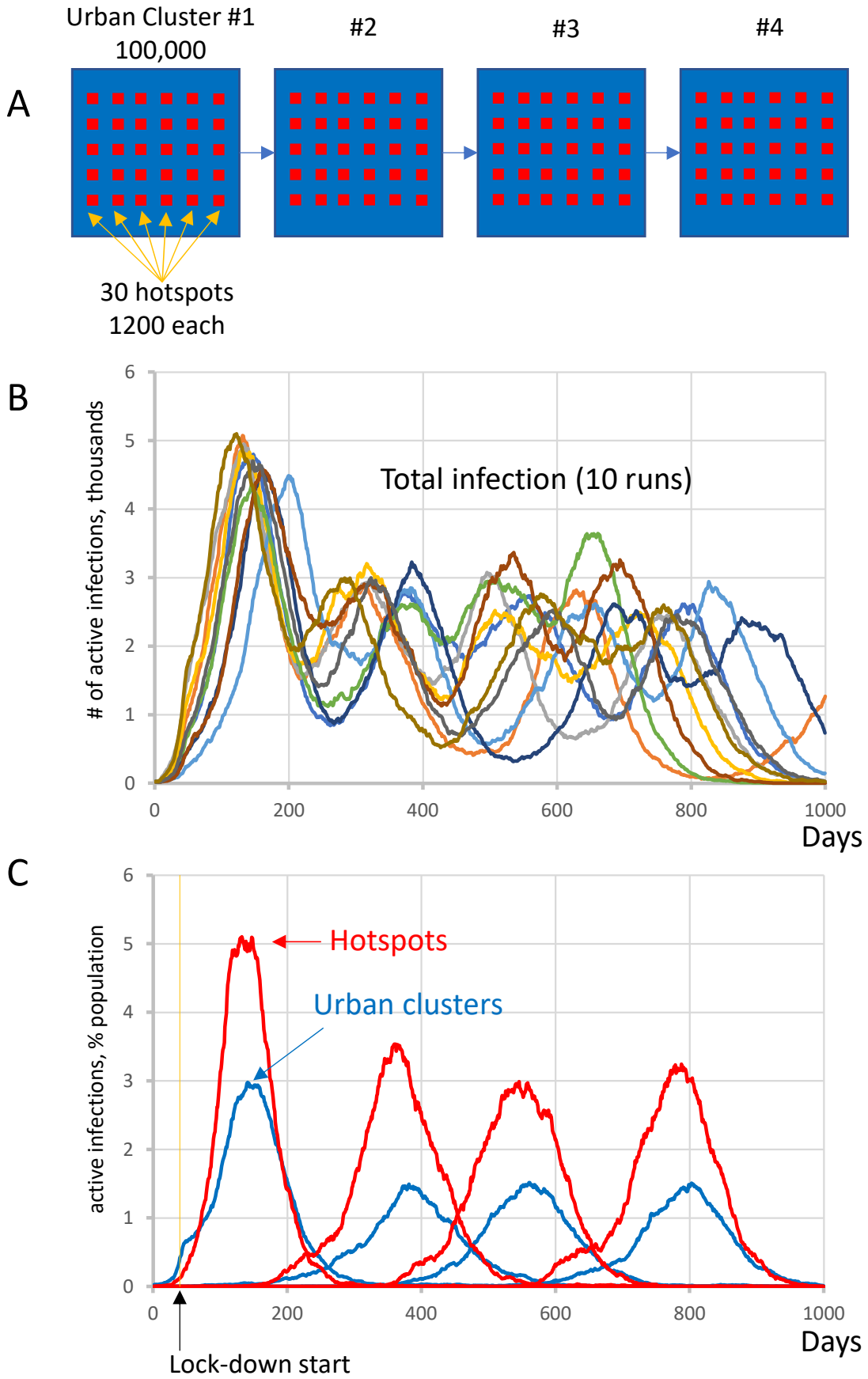
A
Urban Cluster #1
100,000
#2
#3
#4
30 hotspots
1200 each
B
Total infection (10 runs)
# of active infections, thousands
Days
C
Hotspots
Urban clusters
active infections, % population
Days
Lock-down start

Figure 3

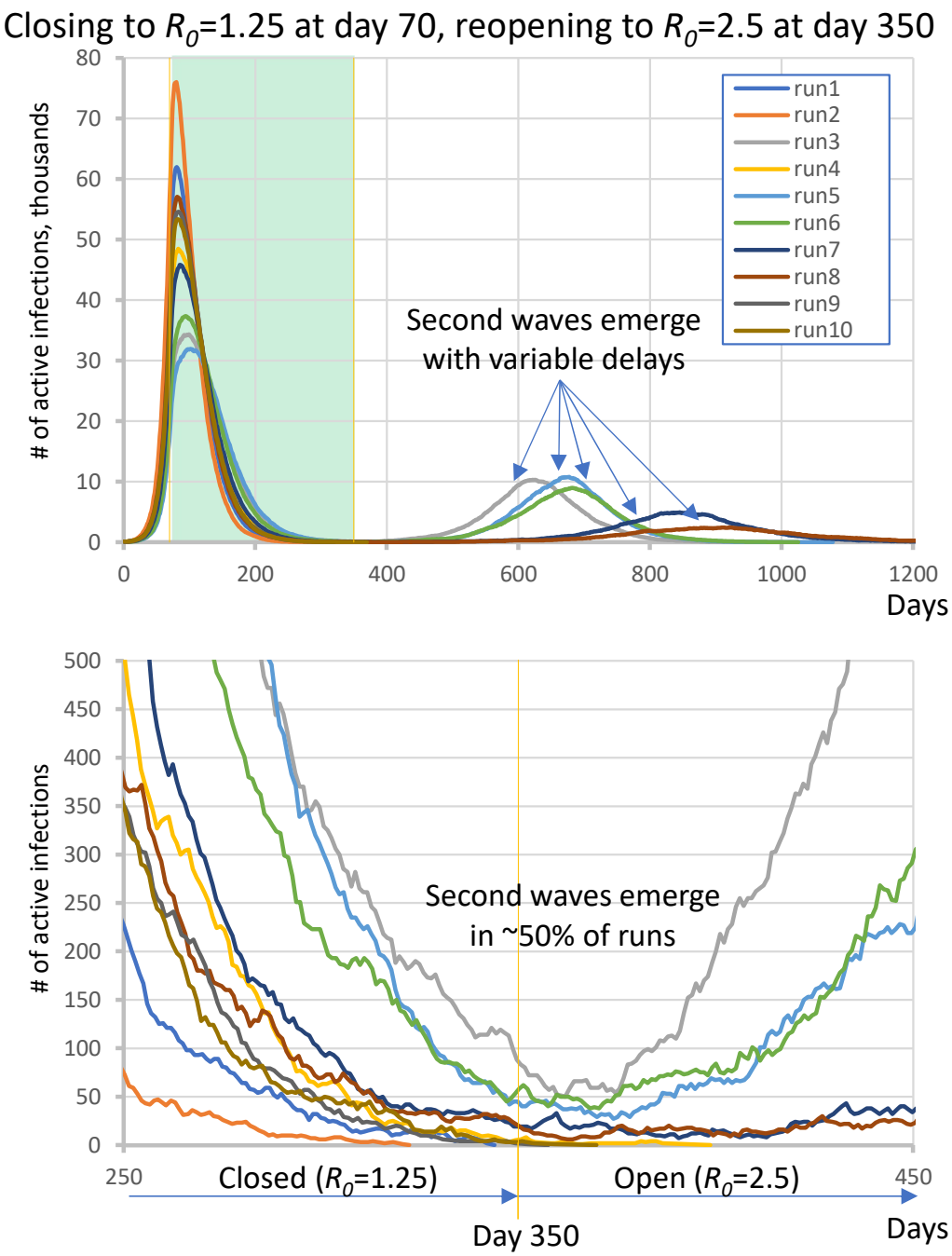
Closing to R_0=1.25 at day 70, reopening to R_0=2.5 at day 350
# of active infections, thousands
Days
run1
run2
run3
run4
run5
run6
run7
run8
run9
run10
Second waves emerge
with variable delays
# of active infections
Second waves emerge
in ~50% of runs
Closed (R_0=1.25)
Open (R_0=2.5)
Day 350
Days

Figure 4

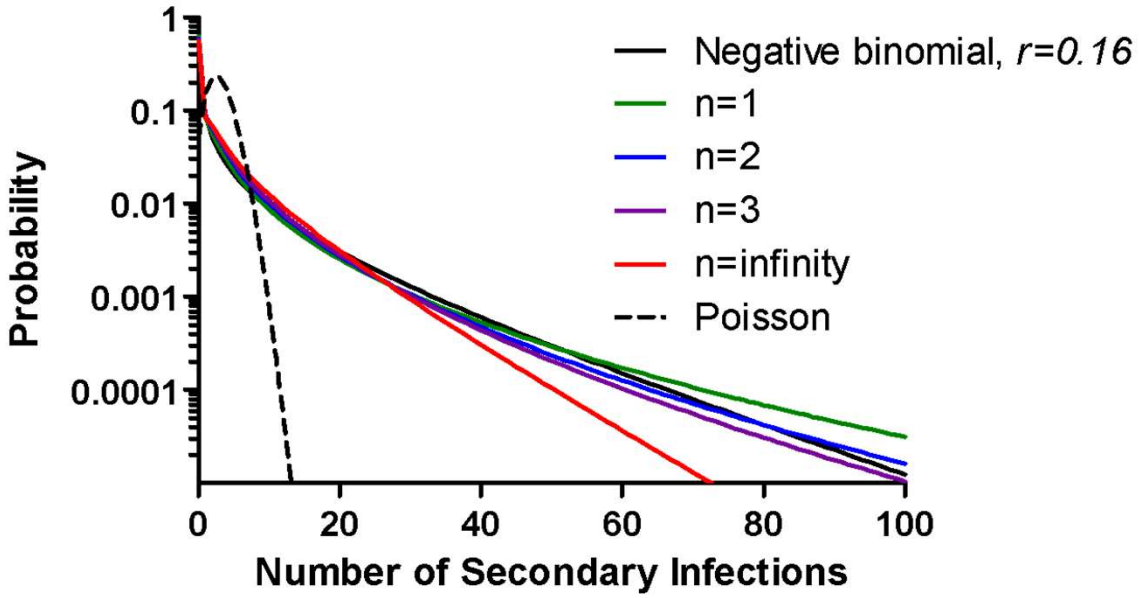
Negative binomial, r=0.16
n=1
n=2
n=3
n=infinity
Poisson
Probability
1
0.1
0.01
0.001
0.0001
0
20
40
60
80
100
Number of Secondary Infections

Figure 5

Supplementary Information for

Social heterogeneity drives complex patterns of the COVID-19 pandemic: insights from a novel Stochastic Heterogeneous Epidemic Model

Alexander V. Maltsev and Michael D. Stern*

**Corresponding Author*:

Michael D. Stern,

Email: SternMi@mail.nih.gov

This PDF file includes:

Figures S1 and S2
Legends for Movies S1 and S2

Other supplementary materials for this manuscript include the following:

Movies S1 and S2

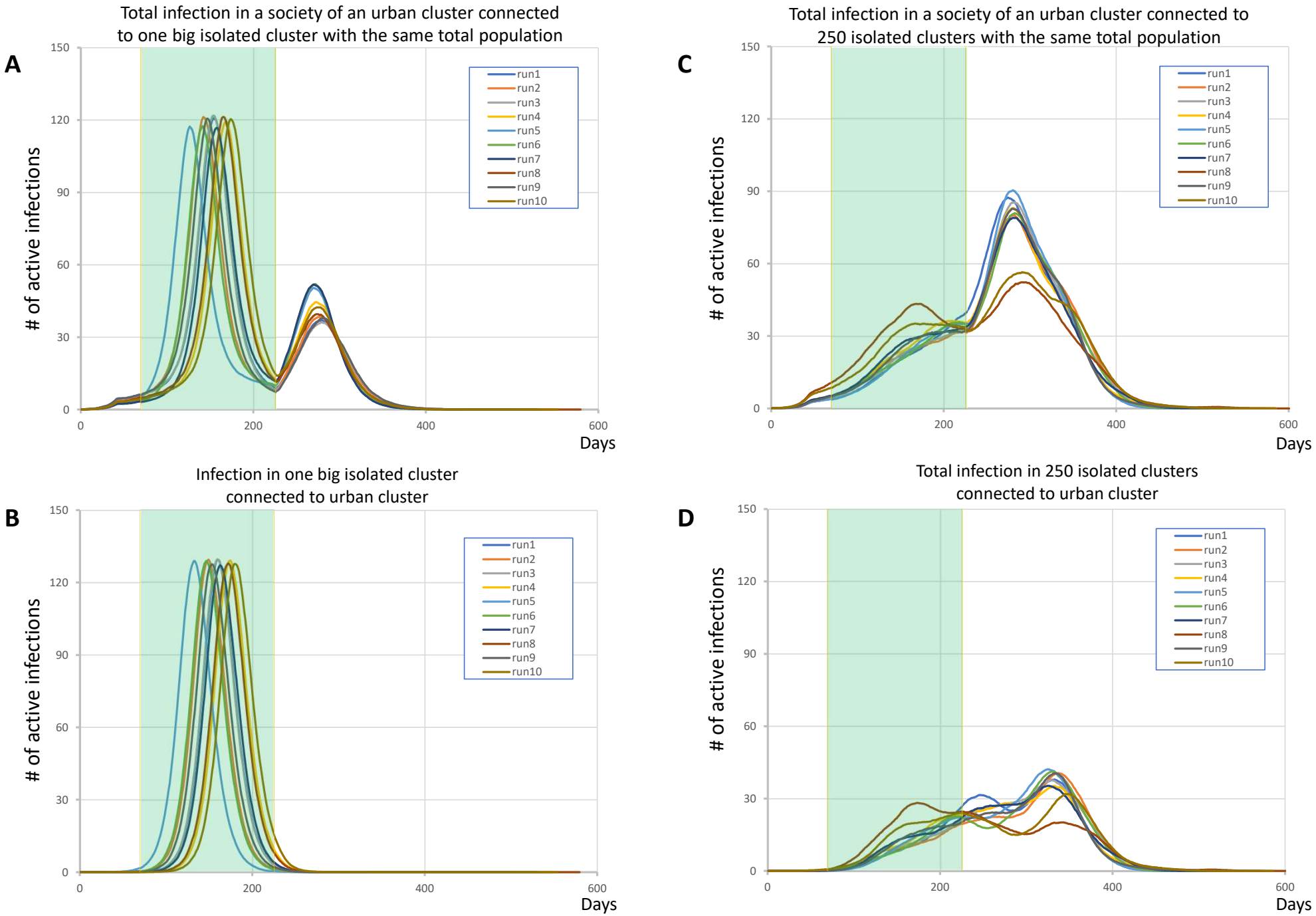

Fig. S1. Heterogeneity of isolated clusters is important for the infection pattern. In the most complex scenario of "moderate lockdown" (Fig. 1C in main text) we substituted 250 clusters by one big cluster with the same population of one million people keeping all other parameters the same. The big isolated cluster generated substantial and sharp infection peak (**panels A and B**) that is absent or very small in case of 250 isolated clusters (**panels C and D**). Each panel shows 10 simulation runs (overlapped multi-color curves). The lockdown period from day 40 to day 225 is shown by green shade.

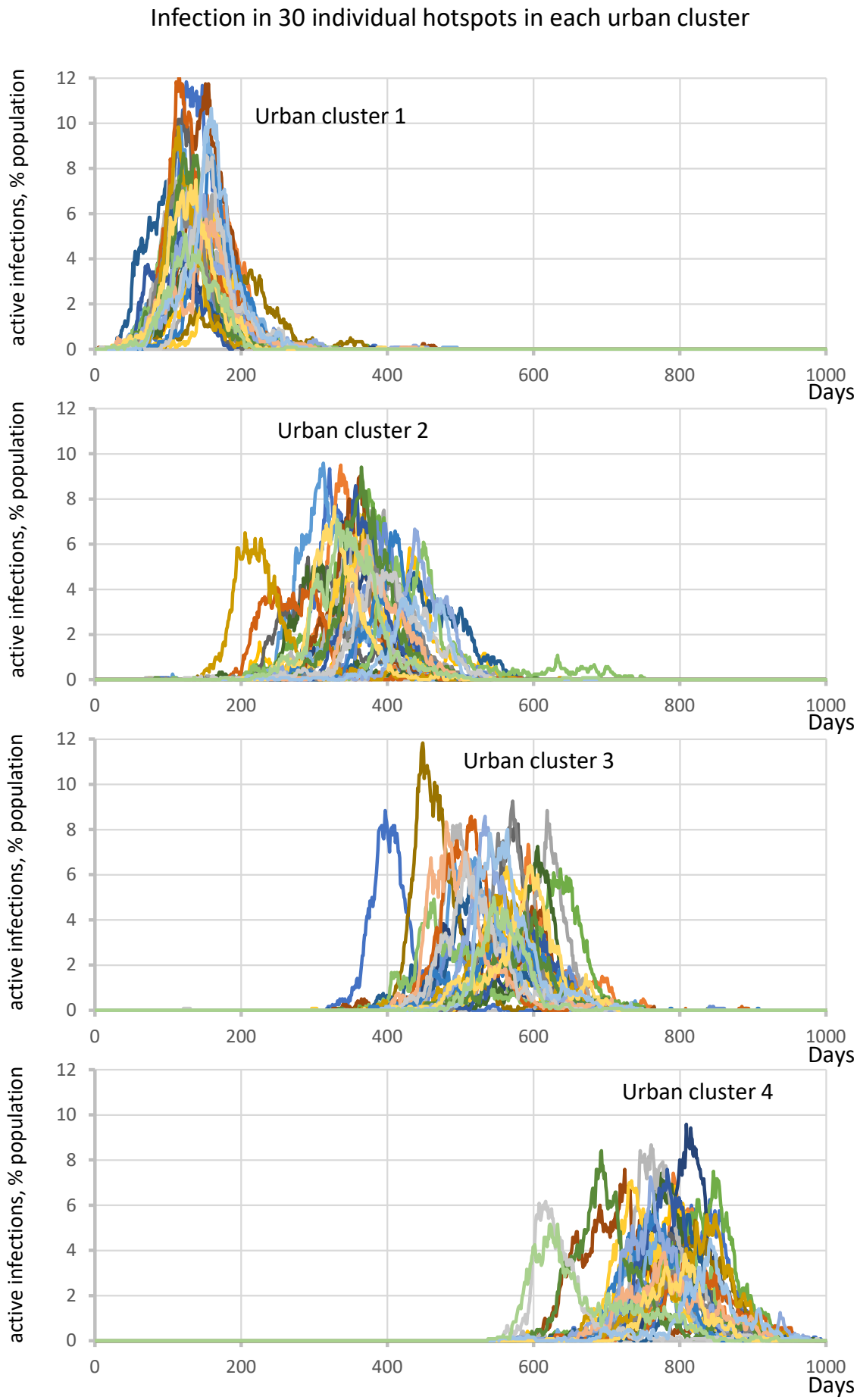

Fig. S2. Stochastic propagation of infection from one urban area to another via hotspots in a society of 4 connected urban areas, each in lockdown but having hotspots. Each plot from top to bottom shows infection explosions in each individual hotspot for each urban area (specified by labels). See main text for details and also Movie S2.

Movie S1 (separate file). Highly susceptible integrated clusters (hotspots) drive SARS-Cov2 infection in an urban cluster in stochastic simulations of SHEM model. Infection time-dependent changes in hotspots (small squares) are coded by red shades saturating (pure red) at 5% of infection in each individual cluster. Infection in the urban area (big square) are is coded by blue shades saturating (pure blue) at 5% of infection in the area. The time is shown in the left upper corner in number of days. Large urban cluster had 1 million individuals with $R_0 = 1.25$ while 250 hotspot clusters with 1200 +/- 500 people had the same internal $R_0 = 3.0$.

Movie S2 (separate file). Hotspots drive SARS-Cov2 infection in each urban cluster and infection propagation among urban clusters in stochastic simulations of SHEM model in a society of 4 connected urban areas. Time-dependent changes of infection in hotspots (small squares) are coded by red shades saturating (pure red) at 3% of infection in each individual cluster. Infection in the urban area (big square) is coded by blue shades saturating (pure blue) at 3% of infection in the area. The time is shown in the left upper corner in number of days. Each urban area of 100,000 people at day 40 became closed from $R_0 = 2.5$ to $R_0 = 1.25$ at day 40. Each urban area has 30 hotspots with 1200 +/- 500 people that avoid closing and keep the same internal $R_0 = 2$.