Assessing the Risk of Cascading COVID-19 Outbreaks from Prison-to-Prison Transfers

Todd L Parsons, Lee Worden

Epidemics 37 (2021) 100532

Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

Assessing the risk of cascading COVID-19 outbreaks from prison-to-prison transfers

Todd L. Parsons [a], Lee Worden [b,*]

[a] CNRS & Laboratoire de Probabilités, Statistique et Modèlisation, Campus Pierre et Marie Curie, Sorbonne Université, Paris, France
[b] Francis I. Proctor Foundation, UCSF, San Francisco, USA

ARTICLE INFO

Keywords:
Household model
Two-level mixing
Disparities
Human rights
Decarceration

ABSTRACT

COVID-19 transmission has been widespread across the California prison system, and at least two of these outbreaks were caused by transfer of infected individuals between prisons. Risks of individual prison outbreaks due to introduction of the virus and of widespread transmission within prisons due to poor conditions have been documented. We examine the additional risk potentially posed by transfer between prisons that can lead to large-scale spread of outbreaks across the prison system if the rate of transfer is sufficiently high.

We estimated the threshold number of individuals transferred per prison per month to generate supercritical transmission between prisons, a condition that could lead to large-scale spread across the prison system. We obtained numerical estimates from a range of representative quantitative assumptions, and derived the percentage of transfers that must be performed with effective quarantine measures to prevent supercritical transmission given known rates of transfers occurring between California prisons.

Our mean estimate of the critical threshold rate of transfers was 27 individuals transferred per prison per month, with standard deviation 26, in the absence of quarantine measures. Available data documents transfers occurring at a rate of 61 transfers per prison per month. At that rate, estimates of the threshold rate of adherence to quarantine precautions had mean 61%, with standard deviation 32%. While the impact of vaccination and possible decarceration measures is unclear, we include estimates of the above quantities given reductions in the probability and extent of outbreaks.

We conclude that the risk of supercritical transmission between California prisons has been substantial, requiring quarantine protocols to be followed rigorously to manage this risk. The rate of outbreaks occurring in California prisons suggests that supercritical transmission may have occurred. *We stress that the thresholds we estimate here do not define a safe level of transfers, even if supercritical transmission between prisons is avoided, since even low rates of transfer can cause very large outbreaks.* We note that risks may persist after vaccination, due for example to variant strains, and in prison systems where widespread vaccination has not occurred. Decarceration remains urgently needed as a public health measure.

1. Introduction

As the COVID-19 pandemic continues in the United States, its dynamics in congregate settings of heightened transmission, including prisons, is crucial to understanding its spread, addressing racial disparities in the burden of the disease, and strategizing effective control.

Prisons are often overcrowded, unsanitary, and provide poor health care, and have been the site of many of the most concentrated and brutal outbreaks of the pandemic so far. Prevention of prison outbreaks is essential because standard control measures such as social distancing and self-isolation are not generally available to prison residents. One in five prisoners in the United States has been infected with SARS-CoV-2, compared to one in 20 in the U.S. overall, more than 1700 have died, and prisoners continue to become infected (The Marshall Project, 2020). The New York Times reported on January 29, 2021 that of the ten largest outbreaks in U.S. correctional facilities to date, six of them have been in California state prisons (The New York Times, 2021). Every single one of California's 35 prisons has reported 200 or more cases (Covid Behind Bars, 2021).

Likely routes of introduction of the disease into prisons are via infected prison staffers, admission of infected prison residents from outside the prison system, and transfers of residents from other prisons. A widely reported outbreak at San Quentin prison in California, which

* Corresponding author.
E-mail addresses: worden.lee@gmail.com, Lee.Worden@UCSF.edu (L. Worden).

https://doi.org/10.1016/j.epidem.2021.100532
Received 12 April 2021; Received in revised form 31 October 2021; Accepted 12 November 2021
Available online 25 November 2021

infected over 2200 of the 3563 inmates and killed 28, was caused by a transfer of prisoners from the Correctional Institute for Men in Chino, California (Office of the Inspector General, 2021), and a subsequent outbreak at California Correctional Center in Susanville, California was likely caused by transfer from San Quentin. Multiple outbreaks in winter 2020 appear to have been caused by importation via staff members.

Prison outbreaks can be very large – the outbreak at San Quentin infected over 60% of the prison population, and the outbreak in California's Avenal State Prison topped 80% – and because of well-known inequities in the criminal justice system, they contribute to racial inequity in the burden of COVID infection (Fortuna et al., 2020; Okonkwo et al., 2020; Lee and Ahmed, 2021; Franco-Paredes et al., 2020). In October 2020, the California Court of Appeals ruled that the California Department of Corrections and Rehabilitation (CDCR) has been guilty of "deliberate indifference" and that California prison populations must be reduced by half to address the ongoing risk of SARS-CoV-2 transmission (Anon, 2020a). This decision reflects the recommendation to decarcerate California prisons to 50% of capacity published by public health experts during the San Quentin outbreak (The Amend Project, 2020). The decision is undergoing appeal, substantial decarceration has not occurred, and multiple outbreaks have occurred in California state prisons in the time since the decision.

While the risks of outbreaks sparked by staff introductions or transfers and spread within prisons due to poor conditions are well known, here we examine the potential danger from another, potentially less apparent risk: the possibility that transmission from prison to prison via transfer of prison residents may be sufficient to lead to uncontrolled spread across the prison system. If such conditions should occur, the disease could be expected to spread to substantially more prisons than otherwise, and infect far more individuals (Fig. 1). The risks to prison residents, staff, and surrounding communities could be considerably increased.

The CDCR currently has quarantine and testing policies in place to prevent transfer of infective individuals. Unfortunately, adherence to CDCR policies has not always been universal, and it cannot be assumed that no risky transfers occur.

We have addressed this question by using established theory of disease transmission, specifically a patch model to be defined below, to estimate the threshold rate of transfer associated with supercritical transmission between prisons, and the rate of adherence to transfer policies needed to prevent supercritical transmission at known rates of transfer.

2. Methods

2.1. Data

Numbers of Covid-19 cases in each CDCR facility were obtained from data from California Department of Corrections and Rehabilitation, collected by the UCLA Covid Behind Bars project (Covid Behind Bars, 2021). Overall population sizes as of March 26, 2021 were extracted directly from data released by CDCR (California Department of Corrections and Rehabilitation, 2021c) (California City Correctional Facility population extracted from California Department of Corrections and Rehabilitation (2021b))

Records released in the course of ongoing legal proceedings document the rate of transfers between prisons in the period from September 21 through October 11, 2020 at about 500 individuals transferred per week (Anon, 2020b, p. 7, line 8). There are 35 institutions in the California prison system, making that equivalent to approximately 61 transfers per prison per month.

2.2. Analysis

A patch model of disease transmission can model a collection of discrete populations in which transmission happens within a population, and at a separate rate between populations. One such approach, the so-called *household model* (Ball et al., 1997), assumes that the population is divided into many small groups (the *households*) in which *local* contacts occur frequently, whereas *global* contacts may occur between any two individuals in the population, albeit at a much lower rate.

In such a model, there are two types of outbreaks: local ones, in which the infection spreads widely within a single group, but remains confined to that group, and global outbreaks, in which the epidemic spreads among many groups. Global outbreaks are governed by a group-to-group reproduction number R_* whose value is the expected number of groups infected by transmission from a single group: the critical value is 1, and a large global outbreak is possible if the value is greater than 1 (Fig. 1).

For our purposes, we assumed that all contacts are local, within groups, except when a transfer occurs of an individual from one group to another. We assumed also that the rate of transfer is low enough that an individual will transfer to at most one group while infective. We found (see Appendix A) that the group-to-group reproduction number has the form

$$R_* = \mu p_G,$$

where μ is the expected size of a major outbreak at a randomly chosen facility and p_G is the probability that an individual will be transferred to a new site *and* cause a major outbreak in the new site (see Appendix A for details).

In order to evaluate how prison transfers affect prison-to-prison transmission, we modelled the group-to-group reproduction number in terms of the average number of individuals transferred between prisons. We expressed the probability p_G in terms of the transfer rate, used empirical prison data to obtain upper and lower bound estimates for μ, and then solved for a threshold rate at which the critical value R_* is equal to one.

We modelled an ensemble of scenarios for these quantities, to cover the range of possibilities. We characterize these as optimistic or pessimistic according to whether they will lead to a lower or higher estimate of R_* within the current modelling framework.

1. Optimistic vs. pessimistic reproduction number. As an estimate of the basic reproduction number R_L within a prison we used the value 8.44 (95% credible interval: 5.00–13.13) estimated from a COVID-19 outbreak in a large urban jail in the U.S. (Puglisi et al., 2020). Because this value is estimated from a setting in which a large outbreak occurred, and conditions in some prisons may be less conducive to transmission than those in which the largest outbreaks have occurred, we took the above number as a pessimistic estimate for R_L. For an optimistic estimate, we calculated the probability p_G that a transfer event leads to transmission between prisons using the more optimistic value of 2.87 (95% CI, 2.39–3.44) that was estimated for a basic reproduction number for COVID-19 in general community transmission (Arif Billah et al., 2020), and cut the probability in half to reflect the possibility that conditions may be better in roughly half of prisons. The factor of one half was chosen arbitrarily, to obtain a conservative bound.

2. Optimistic vs. pessimistic outbreak sizes. We constructed lower and upper bound estimates of the size-weighted mean final outbreak size μ from reported case counts in CDCR prisons. We calculated outbreak sizes to date by taking sequences of reported resident cases at a prison separated by 14 days or more of no cases as separate outbreaks. Because our model results exclude "outbreaks" that end after only a few cases, we exclude these from our estimation. In Appendix E, we estimate the mean and standard deviation of such small outbreaks by computing the total number infected in a branching process conditioned on extinction.

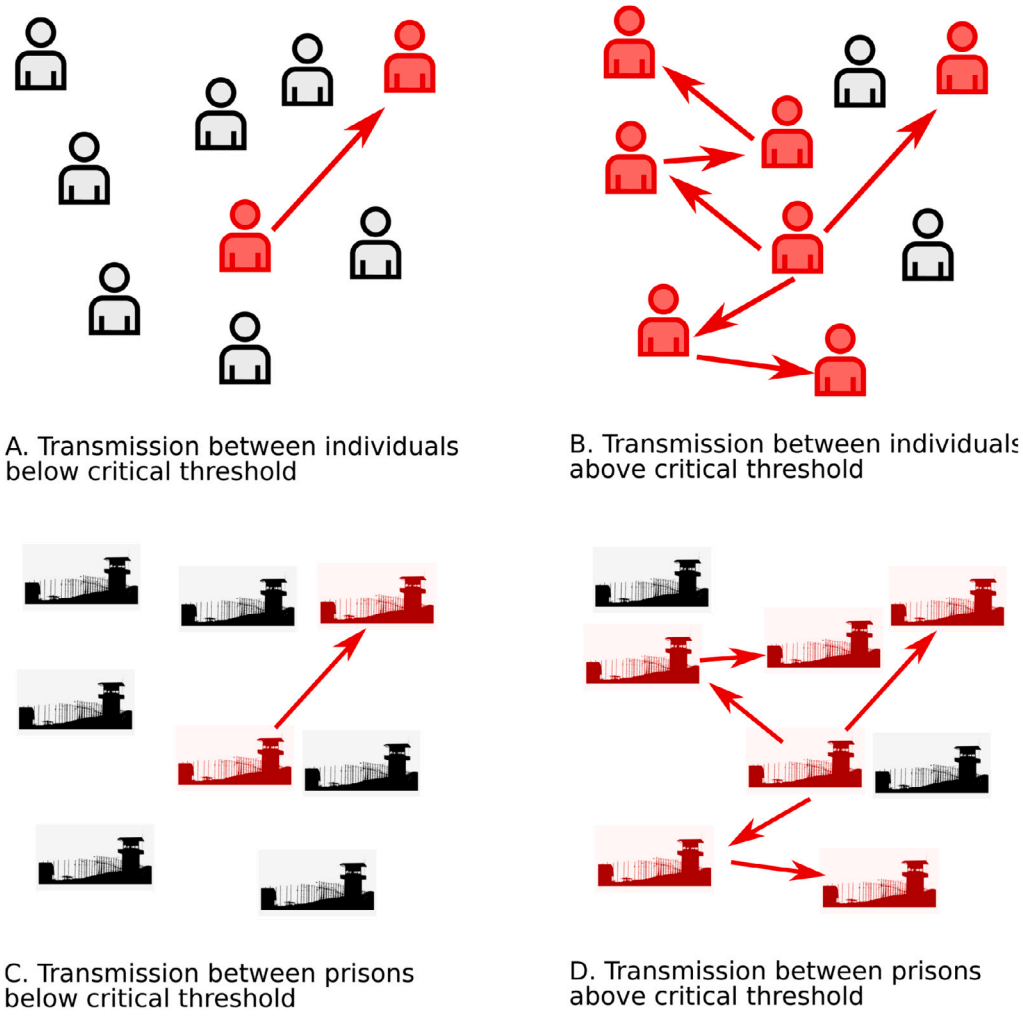

Fig. 1. Scenarios for transmission between individuals and between prisons. A. When the reproduction number R between individuals – the mean number of cases caused by a case – is below the critical threshold of 1, transmission chains are short and outbreaks are small. B. When R is above the critical threshold, a large outbreak is possible. C. When the reproduction number R_* between prisons – the mean number of prison outbreaks caused by a prison outbreak – is below the critical threshold of 1, transmission between prisons may still occur, but spread between prisons will be relatively limited. D. When R_* is above the critical threshold, transmission between prisons can cascade and cause spread throughout the prison system.

Using outbreaks of size 3 and larger (Figs. 2, 3), the size-weighted mean outbreak size μ was 660 cases. Given the realities of asymptomatic infections, incomplete and imperfect testing, and under-reporting, this is likely to underestimate the true sizes of outbreaks. We thus took this value as an estimated lower bound of final outbreak sizes in California prisons. We take the mean overall population of each prison as a conservative upper bound for μ (see Appendix D), which was 3000 as of March 26, 2021.

3. Optimistic vs. pessimistic secondary case distribution. Evidence is accumulating that transmission of SARS-CoV-2 has an overdispersed pattern, in which many people cause few or no infections and a relatively large number of infections are caused by a few people (Althouse et al., 2020; Adam et al., 2020; Susswein and Bansal, 2020). This pattern may make the probability p_G lower than it could be, because relatively more people infect nobody, which reduces the likelihood of a major outbreak (Lloyd-Smith et al., 2005). We estimated p_G given this pattern by assuming a negative binomial distribution of secondary cases (mean = R_L, shape = 0.5 (Susswein and Bansal, 2020)), as is standard. However, this overdispersed pattern may be caused partially by wide variation in the number of people contacted by individuals socially (Susswein and Bansal, 2020), and it is not clear that this variation in contact structure is possible to the same degree in a prison setting, where individuals' movements and locations are heavily constrained and regulated. For this reason, we also considered the possibility that secondary cases may be Poisson distributed within the prison setting, though they are more highly dispersed in community transmission.

4. Optimistic vs. pessimistic timing of transmission events. We also considered that the way in which the timing of transmission events is distributed can affect the probability of transmitting to another prison. If transmission tends to occur in bursts, for example driven by exceptional events when multiple people gather, such a burst might happen either before or after an individual is transferred from prison to prison. If transmission events are independent and happen all at different times, on the other hand, it is more likely that at least one of them will occur after a transfer. We model both of these cases.

We used each combination of the above assumptions to estimate a threshold transfer rate for the California prison system, above which transfers may create a risk of global spread of the coronavirus across the prison system. Calculations detailed in Appendix C use branching process approximations to express the probability p_G in terms of the rate ρ_G of transfer between prisons per person per day for each combination of the above assumptions. These were used, together with our upper and lower bounds on μ, to numerically solve for the threshold rate $\rho_G^\star$ at which $R_* = 1$ under each scenario. We transformed $\rho_G^\star$ to a threshold rate $n^\star$ of transfers per prison per month, using conversion factors of 30 days per month and the average number 2700 of individuals per prisons in the CDCR system as of March 26, 2021.

We note that this threshold number of transfers concerns potentially infective transfers who are exposed to the prison resident population in the facility where they arrive. Prisons, of course, have policies for quarantine of transferred residents and for testing before transfer to prevent transfer of infected individuals, and these policies are likely to reduce the risk due to transfer.

Transfers between California state prisons are regulated by a policy called the movement matrix (California Department of Corrections and Rehabilitation, 2021a). Residents are tested five days before transfer, rapid tested one day before transfer, and quarantined for 14 days after transfer. Quarantine is in celled housing with a solid door if possible, and otherwise in cohorts of no more than four people. Residents in quarantine are screened for symptoms daily, tested if symptomatic, and isolated if they test positive. All of them are tested after five days post-transfer, again after 12 to 14 days post-transfer, and then released if negative and asymptomatic. Both residents and staff are to wear N95 masks during transfer. Residents who have been diagnosed with COVID-19 and subsequently resolved are considered immune for 90 days and exempt from quarantine and testing. After 90 days they are considered susceptible again and subject to the above measures.

We assume that these procedures are likely to reduce the risk of transmission by transfer substantially. However, compliance with safety policies may not be perfect. For example, the California Inspector General has documented extensive noncompliance with mask guidelines in the California prisons, including during the San Quentin outbreak (Office of the Inspector General, 2020). If protective policies reduce the number of transfers who can potentially transmit the virus by some percentage, it is the number of unprotected transfers that must be compared to the threshold value. If transfers are occurring at a known rate n_t, and the threshold transfer rate for uncontrolled transmission between prisons is n^*, then the percentage of transfers that must be conducted in adherence to the protective policy in order to reduce unprotected transfers to the threshold rate is $a = 100(n_t - n^*)/n_t$, or zero if n^* exceeds n_t.

2.3. Vaccination and decarceration

Vaccination in the California prison system is underway, with CDCR reporting that 40% of prison residents have received COVID-19 vaccination. A recent legal filing reported that accounting for previously infected prisoners, 76% of incarcerated people may have immunity (Miller, 2021).

Both increasing immunity and decarceration are likely to affect the spread of infections in at least two important ways, firstly by reducing the reproduction number and relatedly the probability that an introduction leads to an outbreak, and second by reducing the sizes of outbreaks if they occur. Both of these changes will affect our estimates of the rate of transfers needed to produce cascading outbreaks.

We look at the relation between increasing immunity and the critical threshold for cascading outbreaks by estimating the threshold transfer rate and associated quarantine adherence rate, as above, while reducing the local reproduction number parameters discussed above (R_L) by half (which affects our estimate of how often a transfer causes an outbreak, but not of outbreak size), reducing the characteristic outbreak size μ by half, and reducing both by half simultaneously.

3. Results

We first estimated the threshold transfer rate (n^*) under all combinations of the above listed model assumptions, without protective measures (Table 1, Fig. 4). The values estimated for n^* ranged from 3.6 to 92 individuals transferred per prison per month, with mean 27 and standard deviation 26. The generation time distribution used in these estimates was that estimated in a recent meta-analysis (Ferretti et al., 2020): a Weibull distribution with mean 5.5 days and standard deviation 1.8 days (parameters $\alpha = 3.37$, $\beta = 6.12$).

Table 1
Estimates of critical threshold for supercritical transmission between prisons in individuals transferred per prison per month and needed levels of adherence, given partial adherence with California's transfer policy.

Optimistic μ	Optimistic R	Optimistic case distribution	Optimistic timing	n^*	Threshold adherence
Y	Y	Y	Y	92.0	0
N	Y	Y	Y	20.0	67
Y	N	Y	Y	30.0	51
N	N	Y	Y	6.2	90
Y	Y	N	Y	48.0	21
N	Y	N	Y	10.0	83
Y	N	N	Y	22.0	64
N	N	N	Y	4.2	93
Y	Y	Y	N	77.0	0
N	Y	Y	N	18.0	71
Y	N	Y	N	24.0	62
N	N	Y	N	5.6	91
Y	Y	N	N	39.0	36
N	Y	N	N	9.6	84
Y	N	N	N	16.0	75
N	N	N	N	3.6	94

We converted our threshold estimates to the percentage of transfers that must be conducted in compliance with the safety policy in order to achieve the threshold number of unprotected transfers or below, given a total of 61 transfers per prison per month (Table 1, Fig. 4). Our estimates of this threshold rate of adherence to quarantine precautions ranged widely but clustered in the upper third of the percentage scale, with mean 61% and standard deviation 32%.

As a look at the sensitivity of our estimates to reductions in risk due to vaccination and/or decarceration, we estimated the same quantities while reducing the probability of an outbreak, the size of outbreaks, or both, by half (Fig. 5). We estimated that while under our most optimistic assumptions the risk of cascading outbreaks is reduced substantially at 61 transfers per facility per month, to the point where quarantine measures could be ignored entirely without exceeding the estimated threshold (which could of course cause substantial risks other than cascading outbreaks), the change in the median and mean estimates is much more modest. We estimated the mean threshold transfers per prison per month (n^*) at 30, 54, and 61 respectively when reducing the assumed parameter R by half for estimation of the probability of an outbreak, reducing the characteristic outbreak size μ by half, and both (standard deviation 26, 53, 53). The mean estimate of threshold rate of adherence to policy was 55%, 40%, and 34% respectively, with standard deviation 33, 34, 35.

4. Discussion

We have constructed a range of values of a threshold rate of mixing between prisons above which transmission between prisons is likely to be supercritical. Supercriticality between prisons means that a prison outbreak is expected to produce more than one other prison outbreak on average, potentially leading to uncontrolled spread throughout the prison system.

We estimate that the reported rate of transfers that has been occurring in the California prison system has likely exceeded this threshold. We estimate that at these rates of transfers, the quarantine precautions must be highly effective and rates of compliance must be high to avoid risk of supercritical transmission between prisons. The rate of outbreaks occurring in California prisons suggests that supercritical transmission may already have occurred or may be occurring.

We offer these estimates as a way of assessing one of the multiple risks posed by infectious disease transmission in the prison system. *It is important to note that this threshold cannot be understood as providing a safe or acceptable rate of transfers, since transfer rates below the critical threshold can still cause huge outbreaks in multiple prisons*. We are discussing an additional risk beyond the clear dangers of prisons'

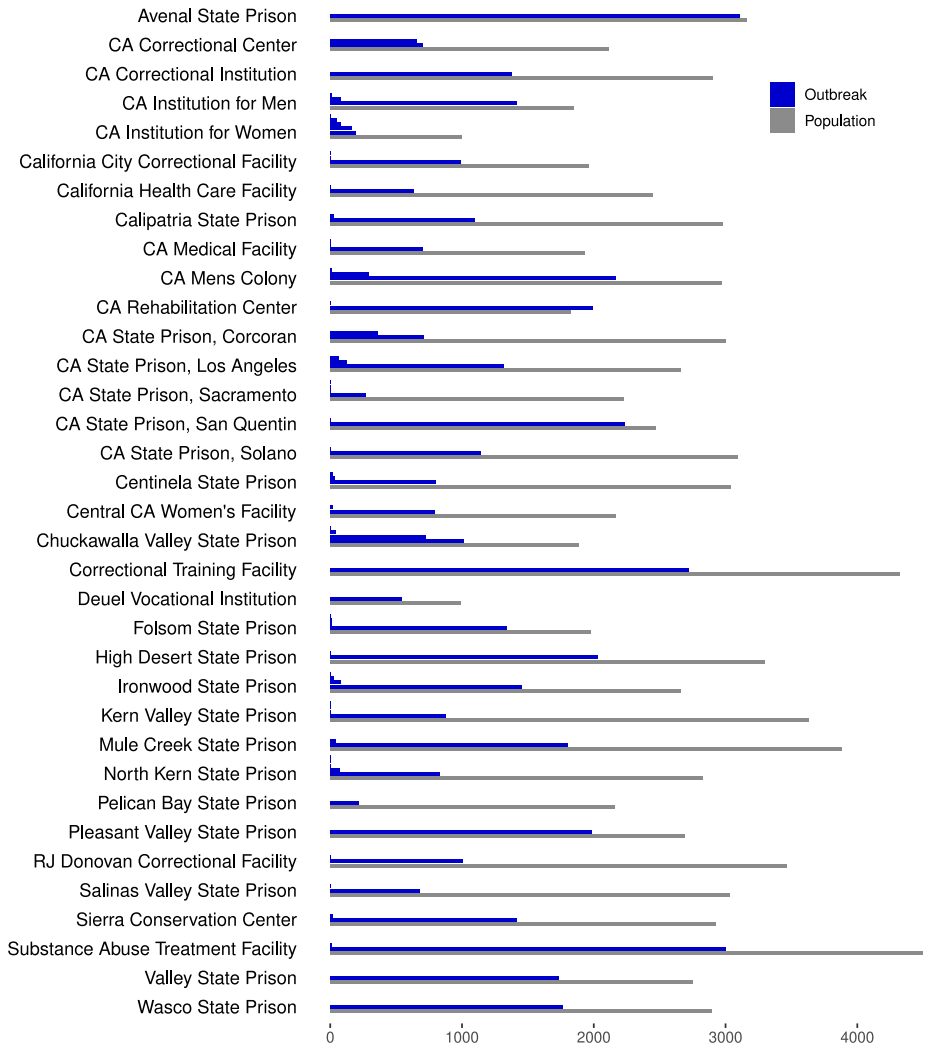

Fig. 2. Sizes of COVID-19 outbreaks in California prisons as of March 26, 2021. Bars show the number of cases per outbreak (blue) for each outbreak of size 3 or more, and total population (grey) at each prison. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

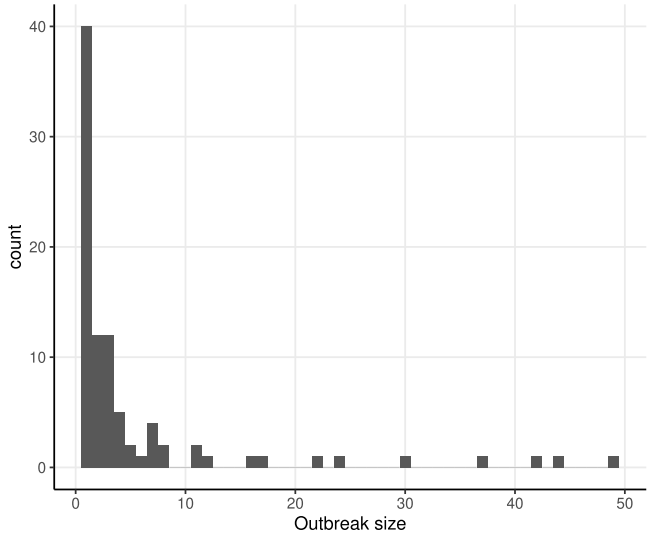

Fig. 3. Distribution of sizes of COVID-19 outbreaks in California prisons, up to size 50.

unsafe conditions and spread due to transfers: the risk that in addition to having multiple large and deadly prison outbreaks of COVID-19, the rate of transfers could be sufficient to cause uncontrolled spread across a large portion of the prison system. This situation would likely lead to a great deal more harm to prison residents and staff than even the known risks of multiple large prison outbreaks, and could place communities throughout the state at risk as well.

The programme of vaccination that is underway in the California prison system is crucial in reducing transmission and saving lives, and will likely help to end the pandemic more broadly as prison transmission poses risks to communities beyond the prison walls. We caution that substantial risks may continue to exist in the CDCR system, as spread of the SARS-CoV-2 virus can still occur, prison conditions continue to be overcrowded and unsanitary, and the effects of variant strains are yet unknown, not to mention the other diseases currently circulating and the potential of future emerging pandemics. Our results and methods may also be relevant to other prison systems where vaccination is not yet widespread. Decarceration remains a crucial public health measure to bring disease spread under control.

While these estimates are necessarily imprecise due to limited availability of data, such that risks could in fact be lower than we have estimated, we note as well that in addition to the mechanism of transfer of prison residents considered here, transmission between prison facilities may also be occurring resulting from travel of infected staff who work at multiple facilities. For this reason, the risk of uncontrolled

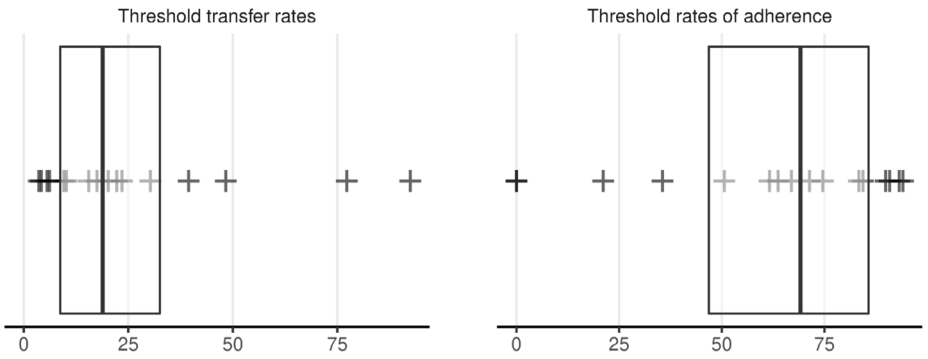

Fig. 4. Estimated threshold transfer rates and levels of adherence to policy. (Left) threshold number of individuals transferred per prison per month, under multiple scenarios (Table 1), (Right) rate of adherence to transfer policy needed to reach threshold number of transfers, under the assumption of 61 total transfers per prison per month. Box plots display median and inter-quartile range.

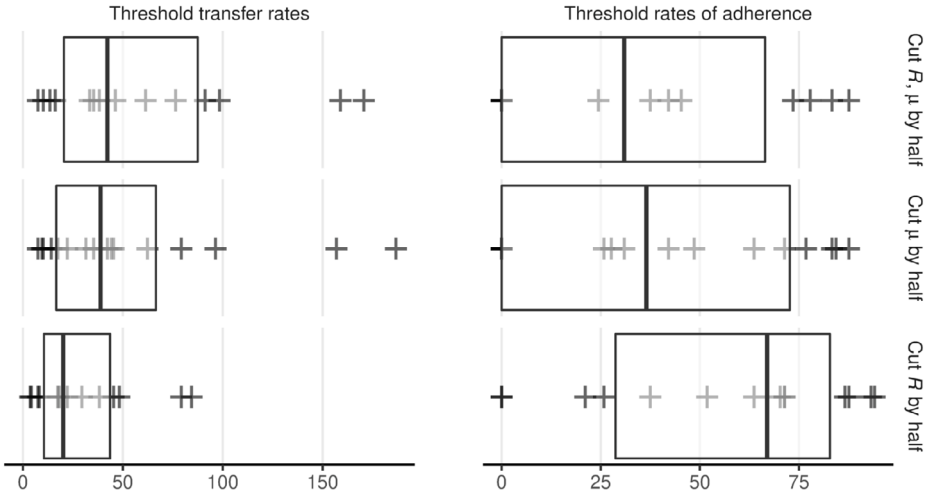

Fig. 5. Estimated threshold transfer rates and levels of adherence to policy given vaccination and/or decarceration, as in previous figure with reductions in reproduction number and/or outbreak size assumed. (Left) threshold number of individuals transferred per prison per month, under multiple scenarios as above, (Right) rate of adherence to transfer policy needed to reach threshold number of transfers, under the assumption of 61 total transfers per prison per month. Box plots display median and inter-quartile range.

transmission between prisons may in fact have been higher than we have estimated here.

These results have a number of limitations. We have assumed that individuals are removed from the epidemic process at the end of their infective period, as we consider the final size of each local epidemic, and thus do not account for the possibility of reinfection. In using a branching process, we have implicitly assumed a very large number of local communities, so that at least initially, each global transmission is to a new site, and ignores the possibility of a second epidemic in the same location. This assumption is reasonable in the context of prisons, where there are indeed many sites. Branching processes are thus most relevant to modelling emerging pathogens, novel strains, and regions with lower rates of vaccination/acquired immunity. We have thus limited ourselves to retrospective conclusions about supercritical transmission in the California prison system; such methods could, however, inform decisions in the face of *e.g.* vaccine resistant variants. The assumption of a homogeneous rate of transfer per individual across all prisons may be limiting as heterogeneity may be important; with sufficient data on site-to-site transfer rates, this could potentially be addressed by dividing facilities into classes with class-specific transfer rates and using a multi-type branching processes, but that is beyond the scope of this study.

This approach is applicable to analysis of risk due to transmission between sites in a variety of hotspot settings of transmission including but not limited to prisons. Transfer, migration, and mixing between sites may be important sources of risk in other locations of high transmission as well, such as jails, ICE facilities, skilled nursing care facilities, meat packing plants, and other agricultural operations.

CRediT authorship contribution statement

Todd L. Parsons: Designed the research, Performed the research, Wrote the paper. **Lee Worden:** Designed the research, Performed the research, Wrote the paper.

Declaration of competing interest

Lee Worden received partial support from the office of the federal receiver, J. Clark Kelso, for research not including this project. Otherwise the authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

LW acknowledges NIH R01 GM130900/GM/NIGMS. We are grateful to Travis Porco, Seth Blumberg, and Jianda Monique, to Martha Lincoln, Sarah Ackley, and multiple colleagues in the Amend and CalPROTECT projects and at CCHCS and CDCR for helpful comments on this work, and to Michael Bien and Ernest Galvan at Rosen, Bien, Galvan & Grunfeld LLP for helpful information regarding transfers.

Appendix A. General household models

A patch model of disease transmission can model a collection of discrete populations in which transmission happens within a population, and at a separate rate between populations. One such approach, the so-called *household model* (Ball et al., 1997), assumes that the population is divided into many small groups (the *households*) in which *local* contacts occur frequently, whereas *global* contacts may occur between any two individuals in the population, albeit at a much lower rate. In such a model, there are two types of outbreaks, local ones, in which the infection spreads widely within a single subpopulation, but remains confined to that subpopulation, and global outbreaks, in which the epidemic spreads among many subpopulations. Here, we consider a variant of the model of Ball et al. (1997); whereas there, global contacts were assumed to occur between individuals that remain in their respective households, here we consider a pandemic caused by individuals moving between households.

A branching process approximation allows us to compute the probability that a global outbreak occurs: if there are a large number of households, at the beginning at least, the probability that the same household receives two or more infected individuals is negligible, so, we can assume to first approximation that each global contact is with a new household. Thus, each infected individual potentially starts a branching process of infected households. The so-called "merciless dichotomy" (Jagers, 1992) tells us that a branching process either goes extinct rapidly, say with probability q, or grows indefinitely, with probability $p = 1 - q$. The latter corresponds to a major pandemic. If the branching process goes extinct, then necessarily, the branching processes started by each infected individual in the first household go extinct as well, which occurs independently for each branching process with probability q as well; this gives us a recursive formula for q.

Let π_i be the probability that a global contact is made with a household with i individuals. Suppose that the fraction of households with i individuals is h_i. If we were to choose a household at random, taking $\pi_i = h_i$ gives the probability of choosing a household of size i. if, on the other hand we choose an *individual* at random, then the probability that individual comes from a household of size i, π_i, is proportional to ih_i (we have equal chance of choosing each individual from the same household), and households are chosen according to a size-biased distribution. The latter is the distribution used in Ball et al. (1997), where transmission is assumed to happen between individuals circulating globally but residing in fixed households, so larger households with more members are more likely to have infectious contacts. The distribution $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots)$ only enters our model in the calculation of the mean household outbreak size; we will use *empirical* upper and lower bounds for this mean size, so we shall not need to specify $\boldsymbol{\pi}$.

Given a household of size i, let P_{ij} be the probability that $j \leq i$ individuals in that household are ultimately infected by an incoming infected individual ("patient zero" to whom we associate the label 0). We then have

$$q = \sum_{i=1}^{\infty} \pi_i \sum_{j=1}^{i} P_{ij} \mathbb{E}[q^{Z_0 + Z_1 + \cdots + Z_j}]$$

where $Z_1, \ldots, Z_j$ are the number of *households* subsequently infected by each of the j infected individuals from the household into which patient zero was introduced ("household one"). We will assume that $Z_1, \ldots, Z_j$ are identical and independent copies of a random variable Z, and all are independent of Z_0, the number of households subsequently infected by patient zero. Because patient zero has already infected one household, we have additional information on that individual, and Z_0 may be distributed differently in light of that information. For example, when an outbreak follows the arrival of patient zero, they may be subsequently quarantined and thus not relocate again during their infectious period, so $Z_0 = 0$.

By independence,

$$q = \sum_{i=1}^{\infty} \pi_i \sum_{j=1}^{i} P_{ij} \mathbb{E}[q^{Z_0}] \mathbb{E}[q^{Z}]^j. \tag{A.1}$$

(*n.b.* it is here that we are assuming the number of households is very large, so that each infected individual is making contact with distinct households). Write $F(q)$ for the expression on the right in (A.1); the function $F(z)$ is the probability generating function (PGF) for the total number of households infected by patient zero. $F(z)$ and $F'(z)$ are positive and increasing for $z > 0$. $F(1) = 1$ and $F'(1)$ is the expected number of households infected by individuals from household one (see *e.g.* Bartlett (1956), Jagers (1975) for results on PGFs for branching processes). The branching process has the possibility of growing indefinitely (*i.e.* a major pandemic occurs) if and only if the equation $F(q) = q$ has a second solution $0 < q < 1$.

Solving $F(q) = q$ analytically is generally impossible, but we can easily derive a criterion for the existence of a second solution. Since

$$F(0) = \sum_{i=1}^{\infty} \pi_i P_{i0}$$

is the probability that the patient zero infects no other individuals in their household, we must have $F(0) > 0$. Rolle's Theorem tells us that $F(q) = q$ has a solution $q < 1$ if and only if $F'(q') = 1$ for some $q < q' < 1$. Since $F'(z)$ is increasing, this shows that there is a second solution if and only if $F'(1) > 1$.

Now,

$$F'(1) = \sum_{i=1}^{\infty} \pi_i \sum_{j=1}^{i} P_{ij} \left(\mathbb{E}[Z_1] + j\mathbb{E}[Z] \right) = \sum_{i=1}^{\infty} \pi_i \left(\mathbb{E}[Z_1] + \mu_i \mathbb{E}[Z] \right).$$

where $\mu_i = \sum_{j=1}^{i} j P_{ij}$ is the expected number of the i individuals in a household of size i who are infected by the incoming patient zero (see Appendix D for an approximation for $\frac{\mu_i}{i}$ when i is large). Now let

$$\mu = \sum_{i=1}^{\infty} \pi_i \mu_i,$$

so μ is the expected number of individuals infected in household one, when the household is drawn according to $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots)$, and

$$F'(1) = \mathbb{E}[Z_1] + \mu \mathbb{E}[Z]$$

gives the expected total number of households infected by the first infected individual ($\mathbb{E}[Z_1]$) and their contacts (each one makes $\mathbb{E}[Z]$ contacts), *i.e.* R_*.

R_* thus reflects the fact that transmission between households is proportional to the number of individuals within the house, who each individually make global contacts with other households. In the same way that the critical threshold $R_0 = 1$ for the basic reproduction number separates subcritical from supercritical transmission in non-patch models, the boundary $R_* = 1$ is the global critical threshold above which transmission is globally supercritical, meaning that an outbreak in one group is expected to infect more than one other group on average and can cause a large outbreak across the system of groups. To emphasize the distinction between R_* and the usual basic reproduction number, in what follows we will use R_L to denote the basic reproduction number for a local epidemic at a single site.

The branching process approximation we have chosen has considerable flexibility beyond the application presented here. At the cost of making the parameter μ something of a black box, we need only make minimal assumptions about individual epidemic and transfer dynamics: for example, we do not need to make specific assumptions about the duration of the infectious or latent periods, and the results are equally compatible with SIR or SEIR dynamics. In particular, we do not make explicit use of the values π_i, which only enter into the weighted mean outbreak size μ.

Each of these choices, however, will result in different values of μ; while the distribution of prison sizes (h_i) is an empirical and observable

quantity, the mean number of infections at a site will have to be either computed or observed empirically. When individual facilities are sufficiently large that an ordinary differential equation (ODE) model is reasonable approximation to the local dynamics, we can use standard compartmental models of estimate the final size of the epidemic. When local facilities are small, more computationally intensive methods are required (see Picard and Lefevre (1990)). Here, we will use facility sizes as an upper bound for the size of the local outbreak (the total fraction infected for *e.g.* in the SIR compartmental model, the total fraction infected is approximately $1 - e^{-R_L}$, which is very close to 1 for values of R_L estimated for SARS-CoV-2), and the average size of reported epidemics as a lower bound: underreporting and asymptomatic infections ensure that reported sizes underestimate the true size of outbreaks.

In what follows, we will assume that individuals are moved sufficiently infrequently between sites that the probability that they move twice during their infectious period is negligible *i.e.* $\mathbb{E}[Z_1] = 0$ whereas $\mathbb{E}[Z]$ is the probability that an individual moves to a new household during their infectious period *and* causes an outbreak in their new household, p_G. As a result, we are underestimating the probability of a major pandemic.

In large populations, outbreaks tend to two extremes: they either "fizzle", resulting in an average of fewer than $\frac{1}{R_L}$ new infections, or they result in a large outbreak in which almost all of the susceptible population is eventually infected. The former are likely to go undetected, so we neglect outbreaks that fizzle, and only consider large outbreaks in computing p_G. Again, this results in an underestimation of the probability of a major pandemic. After an interlude explaining the approach we take to modelling individual transmission, we turn our attention to computing this probability.

Appendix B. Point processes for infections

B.1. Intensity functions and the generation interval density

The Crump-Mode-Jagers general branching process (Crump and Mode, 1968, 1969; Jagers, 1975) provides an extremely supple framework in which to investigate the initial phase of an epidemic, giving an individual-based stochastic model that corresponds to the renewal equation at the early phase of an epidemic when, in a large population, we may assume that each new contact made by an infectious individual is with a susceptible individual who has not been previously contacted. Fix $t = 0$ as the time that a given focal individual is infected, and assume secondary infections at random times $t_1, t_2, \ldots$. The ensemble of times is an example of a *point process*, which is represented by its *counting measure*, N: given a subset $A \subseteq [0, \infty)$,

$$N(A) = \#\{t_i : t_i \in A\}.$$

We may similarly assign such a point process to each secondary infection, with the points t_i now corresponding to the time elapsed between their infection and the tertiary infections they cause.

One can integrate an arbitrary function f with respect to $N(dt) = \#\{t_i : t_i \in dt\}$:

$$\int_0^t f(t)\, N(dt) = \sum_i f(t_i),$$

whenever the sum on the right hand side exists.

If no two points coincide, *i.e.* if the infection is a *simple point process*, in which case

$$N(dt) = \begin{cases} 1 & \text{if } t_i \in dt \text{ for some } i, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

(*n.b.*, while no two points coincide, we do allow points to be arbitrarily close). A simple point process is characterized by its *intensity function*

$$\lambda(t|\mathcal{H}_t)\, dt = \mathbb{P}\{t_i \in dt \text{ for some } i|\mathcal{H}_t\},$$

where $\mathcal{H}_t$ denotes the information available at time t, which includes the infection times $t_i \le t$, but may also include information about other hidden variables (we illustrate this with an example below). For a set $A \subset [0, \infty)$, $N(A)$ is a Poisson point process with rate $\int_A \lambda(t|\mathcal{H}_t)\, dt$. When $\lambda(t|\mathcal{H}_t)$ is a random function *independent of the prior infection times*, then the point process is a *Cox process* or *doubly stochastic Poisson process*; when it is a deterministic function of time, it is an *inhomogeneous Poisson process*; finally, if the intensity function is a constant λ, then the point process is the classical Poisson process.

Example 1. The classical SIR compartmental differential equation model can be interpreted as the large population limit of a stochastic model in which, upon being infected, makes infectious contacts at rate β during an infectious period that is exponentially distributed with (recovery) rate γ. If L denotes the infectious period of the first infectious individual, and we measure time with $t = 0$ corresponding to their time of infection, then their infection point process has intensity $\lambda(t|\mathcal{H}_t) = \beta \mathbb{1}_{\{L \ge t\}}$, where $\mathbb{1}_{\{L \ge t\}}$ is the *indicator function* of their infectious period:

$$\mathbb{1}_{\{L \ge t\}} = \begin{cases} 1 & \text{if } L \ge t, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

This is an example of a Cox process: the intensity function is random, as it depends on the infectious period L, but the probability of an infectious contact at time t is independent of any previous contacts. Here, $\mathcal{H}_t$ contains the information of whether or not $L \ge t$ (*n.b.* it does not include the value L itself unless $L \le t$, as $\mathcal{H}_t$ contains no information about the future beyond time t).

In general, each individual has a distinct intensity function, which may depend on one or more hidden variables (*e.g.* in Example 1 above, each individual has a distinct – random – infectious period). We may also consider the expected intensity function of the point process, $h(t)$, given by

$$h(t) = \mathbb{E}[\lambda(t|\mathcal{H}_t)].$$

$h(t)$ the corresponds to the ensemble average intensity function over a large population. If $\int_0^\infty h(s)\, ds < \infty$, then we can normalize $h(t)$ to obtain a probability density function,

$$g(t) = \frac{h(t)}{\int_0^\infty h(s)\, ds},$$

which is the generation interval density. In particular, the total expected number of infections caused by patient zero, the *local reproduction number*, is

$$R_L = \mathbb{E}[N([0, \infty))] = \mathbb{E}\left[\int_0^\infty N(ds)\right] = \int_0^\infty h(s)\, ds,$$

Thus, $h(t) = R_L g(t)$ *e.g.* in Example 1, we see that

$$h(t) = \mathbb{E}[\beta \mathbb{1}_{\{L \ge t\}}] = \beta \mathbb{P}\{L \ge t\} = \beta e^{-\gamma t},$$

and $\int_0^\infty h(t)\, dt = \frac{\beta}{\gamma} = R_L$.

Given the abundance of data regarding the generation interval, and the relative paucity of information on individual hidden variables, we will use $h(t) = R_L g(t)$ for the empirical generation interval distribution for each individual's intensity function — approximating each individual by the population average. This allows us to make full use of available information about the timing of infection events, rather than imposing *e.g.* a compartmental model. It is not, however, without consequence, *e.g.* the total number of infections caused by "patient zero", $N([0, \infty))$, is a Poisson process with rate

$$\int_0^\infty h(t)\, dt = R_L \int_0^\infty g(s)\, ds = R_L.$$

On the other hand, if we compute the probability generating function (the PGF, which uniquely characterizes a random variable) for the SIR compartmental model of Example 1 above, where the total number

of infections caused by a single individual is Poisson distributed with (random) rate

$$\int_0^\infty \beta \mathbb{1}_{\{L \geq t\}}\, dt = \beta L,$$

we have

$$\begin{aligned}\mathbb{E}[z^{N[0,\infty)}] &= \mathbb{E}\left[\mathbb{E}[z^{N[0,\infty)}|L]\right] \\ &= \mathbb{E}\left[e^{\beta L(z-1)}\right] \\ &= \frac{\gamma}{\gamma - \beta(z-1)}\end{aligned}$$

i.e. the total number of infections is geometrically distributed with success probability $p = \frac{\gamma}{\beta+\gamma} = \frac{1}{1+R_L}$, which, despite having the same mean, is not Poisson distributed with rate R_L.

This might otherwise might be considered a deficiency, but for us it becomes a virtue: the geometric distribution has variance to mean ratio (VMR) of $\frac{1}{p} = 1+R_L > 1$ and the Poisson process has VMR one, whereas the VMR for empirically observed individual transmission lies between the two. As we discuss below, however, our inhomogeneous Poisson process model can be used to construct a family of processes with the same reproduction number R_L and generation interval $g(t)$ whose total number of infections per infected are negative binomially distributed, and thus include both the Poisson and geometric distributions as well as processes with VMRs taking all values in $[1, \infty)$.

B.2. Incorporating overdispersion

To allow for varying dispersion – as well as individual variation in contact rates – we consider a doubly stochastic Poisson (or Cox) process, drawing an individual reproductive ratio, r, for each individual from a fixed distribution with mean R_L.

Given $A \subset [0, \infty)$, set $g_A = \int_A g(t)\, dt$. If we assume each individual's r is gamma distributed with shape parameter k and scale parameter $\frac{R_L}{k}$ for some $k > 0$, then the probability of m infections in an interval $[a, b]$ caused by a randomly chosen individual is

$$\begin{aligned}\int_0^\infty \frac{(rg_A)^m}{m!} e^{-rg_A} r^{k-1} e^{-\frac{k}{R_L} r}\, dt = \frac{\Gamma(m+k)}{m!\,\Gamma(k)} \left(\frac{k}{R_L g_A + k}\right)^k \\ \times \left(\frac{R_L g_A}{R_L g_A + k}\right)^m,\end{aligned}$$

which we recognize as a negative binomial distribution with success probability $\frac{R_L g_A}{R_L g_A + k}$ and k failures. The mean number of infections occurring at times in A is thus $R_L g_A$ and the VMR of the number of those infections is $1 + \frac{R_L g_A}{k}$. In particular, taking A to be the whole real line, we see that the mean total number of infections caused by a single individual is negative binomially distributed with mean R_L and index of dispersion $1 + \frac{R_L}{k}$; taking $k = 1$, we obtain a geometric distribution with success probability $\frac{1}{1+R_L}$, whereas taking $k \to \infty$ yields a Poisson distribution with rate R_L.

B.3. Compound Poisson processes

More generally, we can consider the case when events occur with intensity function $h(t)$, but now the number of infections occurring at the i^{th} contact is given by a random variables ν_i, where $\nu_1, \nu_2, \ldots$ are independent and identically distributed (i.i.d.). For example, given low-resolution temporal data, we may be unable to distinguish nearby infection times. The number of infections occurring at times in the set A is then

$$\tilde{N}(A) = \sum_{i=1}^{N(A)} \nu_i,$$

where, as before, $N(dt)$ is a simple point process and

$$\mathbb{E}[\tilde{N}(dt)] = \mathbb{E}[\nu] h(t)\, dt,$$

and

$$R_L = \mathbb{E}[\tilde{N}([0,\infty))] = \mathbb{E}[\nu] \int_0^\infty h(t)\, dt.$$

Unlike previously, this relation does not fix the value of $\int_0^\infty h(t)\, dt$. Rather, for any $\lambda \in (0, \infty)$, we can have $\int_0^\infty h(t)\, dt = \lambda$, provided $\mathbb{E}[\nu] = R_L/\lambda$. We then have $h(t) = \lambda g(t)$, so λ determines the number of points: an increased λ will correspond to fewer infections on average per event and *vice versa*. In particular, for a given R_L, we can potentially have arbitrarily large clusters of infections, provided they occur sufficiently rarely.

B.3.1. Conditioning on a single event

As an extreme case, we consider a scenario in which all transmission from a given individual occurs at a single point in time. The probability of exactly one event in the inhomogeneous Poisson point process is

$$\int_0^\infty h(t)\, dt\, e^{-\int_0^\infty h(t)\, dt} = \lambda e^{-\lambda},$$

whereas the joint probability of having a single point at t is

$$h(t) e^{-\int_0^\infty h(t)\, dt},$$

so the probability of a single point at t conditional on only one point is the ratio of these two probabilities,

$$\frac{h(t) e^{-\int_0^\infty h(t)\, dt}}{\int_0^\infty h(t)\, dt e^{-\int_0^\infty h(t)\, dt}} = \frac{h(t)}{\int_0^\infty h(t)\, dt} = g(t).$$

Since we require that the expected number of infections remains equal to R_L, we must have that their number, ν_1, satisfies $\mathbb{E}[\nu_1] = R_L$.

Appendix C. Probability of a large local outbreak, p_G

We now use the modelling framework described above to determine the compute the probability, p_G from Appendix A, that an infected individual is transferred and causes a major outbreak in the new site. We will work in the context of our prison to prison model, taking the prisons as the "households" of the model. Recall that a minor outbreak occurs when the branching process approximating the initial transmission goes extinct, and a major outbreak occurs otherwise. In practice, it is more transparent to compute the probability $q_G = 1 - p_G$. that a major outbreak does not occur, which we do below.

By assumption, an individual can infect a new site if and only if they are transferred while infectious. We will assume that transfers occur as a Poisson point process with rate ρ_G (transfers per individual per day), which for simplicity is assumed to sufficiently small that the probability that a given individual is transferred multiple times while infectious is negligibly small. Thus, each individual waits an exponentially distributed time with mean $1/\rho_G$ before being transferred.

To obtain a range of plausible values for q_G/p_G, we will consider each of the models discussed in Appendix B:

(i) a compound Poisson model in which infections happen individually and independently with a time-dependent intensity function $h(t) = R_L g(t)$,
(ii) a doubly-stochastic Poisson model, where individual reproduction number varies,
(iii) a compound Poisson process model where each event leads to multiple infections, and
(iv) a "burst" model in which all infections occur simultaneously.

C.1. Poisson distributed infections

To compute the probability of a local large outbreak under our generation interval approach, we must take care to distinguish between the patient zero who initiated the epidemic, for whom some fraction of the infectious period has already elapsed, and the newly infected individuals in the new site.

Under the assumption of a large local population, we can continue to use branching process recursive formulas to determine the probability that no locally infected individual gives rise to a significant outbreak, say q_L:

$$q_L = \mathbb{E}\left[q_L^{N([0,T))}\right]. \tag{C.1}$$

Here $N([0,T))$ is the number of individuals infected by a given infectious individual before that infectious individual is transferred at a random time T after infection. As before, T is exponentially distributed with rate ρ_G, whereas conditional on T, $N([0,T))$ is Poisson distributed with rate $R_L \int_0^T g(t)\,dt$. Now,

$$q_L = \mathbb{E}\left[\mathbb{E}\left[q_L^{N([0,T))}|T\right]\right],$$

where the outer expectation is for the random variable T. Recognizing the inner expectation as the probability generating function for a Poisson random variable gives us

$$\begin{aligned} q_L &= \mathbb{E}\left[e^{R_L(q_L-1)\int_0^T g(t)\,dt}\right] \\ &= \int_0^\infty \rho_G e^{-\rho_G t} e^{R_L(q_L-1)\int_0^t g(u)\,du}\,dt. \end{aligned}$$

Now consider patient zero. Let T' (still exponentially distributed with rate ρ_G) be the time after the start of their infectious period at which patient zero was introduced into the local community. Then, they will infect $N([T',\infty))$ individuals in the new site, whence the probability that there is not a major outbreak is

$$q_G = \mathbb{E}\left[q_L^{N([T',\infty))}\right].$$

Again, first taking the conditional expectation of $N([T',\infty))$ given T', which is a Poisson random variable with rate $R_L \int_{T'}^\infty g(t)\,dt = R_L G(T')$, and then over T', gives

$$q_G = \int_0^\infty \rho_G e^{-\rho_G t} e^{-R_L(1-q_L)G(t)}\,dt \tag{C.2}$$

e.g. for the Weibull distribution with shape a and scale b, $G(t) = e^{-\left(\frac{t}{b}\right)^a}$.

C.2. Individually varying reproductive number

More generally, as before, we can consider the possibility that each individual has an i.i.d. reproductive number, say R_i drawn according to a Gamma $\left(k, \frac{R_L}{k}\right)$ distribution, in which case, we may proceed as before to obtain (C.1), which we must now average over the gamma distribution to obtain

$$\begin{aligned} q_L &= \mathbb{E}\left[q^{N([0,T))}\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\mathbb{E}\left[q^{N([0,T))}|R_i,T\right]|T\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[e^{R_i\int_0^T g(t)\,dt(q_L-1)}|T\right]\right] \\ &= \mathbb{E}\left[\left(1-\frac{R_L(q_L-1)}{k}\int_0^T g(t)\,dt\right)^{-k}\right] \end{aligned}$$

and, recalling $G(t) = \int_t^\infty g(u)\,du$,

$$\begin{aligned} &= \mathbb{E}\left[\left(1-\frac{R_L(q_L-1)(1-G(T))}{k}\right)^{-k}\right] \\ &= \int_0^\infty \rho_G e^{-\rho_G t}\left(1+\frac{R_L(1-q_L)(1-G(t))}{k}\right)^{-k} dt, \end{aligned}$$

which we may solve numerically for q_L. Proceeding similarly, averaging (C.2) over the gamma distribution and interchanging the order of integration yields

$$\begin{aligned} q_G &= \int_0^\infty \rho_G e^{-\rho_G t}\left(1-\frac{R_L}{k}G(t)(q_L-1)\right)^{-k} dt \\ &= \int_0^\infty \rho_G e^{-\rho_G t}\left(1+\frac{R_L}{k}G(t)p_L\right)^{-k} dt, \end{aligned}$$

where we recall $p_L = 1 - q_L$.

C.3. Compound Poisson processes

Now, suppose that at the i^{th} time of the inhomogeneous Poisson processes $N([0,T))$ and $N([T',\infty))$ (as previously, T and T' indicate the transfer time out of, or into, the focal site respectively) the individual independently infects ν_i individuals, where ν_i is a random variable with probability generating function $P_\nu(z) = \mathbb{E}[z^\nu]$.

Now, since the ν_i are independently and identically distributed and independent of their arrival times in $N([0,T))$,

$$\begin{aligned} q_L &= \mathbb{E}\left[q_L^{\sum_{i=1}^{N([0,T))}\nu_i}\right] = \mathbb{E}\left[\prod_{i=1}^{N([0,T))} q_L^{\nu_i}\right] = \mathbb{E}\left[\prod_{i=1}^{N([0,T))} \mathbb{E}\left[q_L^{\nu_i}|N([0,T))\right]\right] \\ &= \mathbb{E}\left[\prod_{i=1}^{N([0,T))} \mathbb{E}\left[q_L^{\nu_i}\right]\right] = \mathbb{E}\left[\prod_{i=1}^{N([0,T))} P_\nu(q_L)\right] = \mathbb{E}\left[P_\nu(q_L)^{N([0,T))}\right]. \end{aligned}$$

Now, proceeding as before,

$$\begin{aligned} \mathbb{E}\left[P_\nu(q_L)^{N([0,T))}\right] &= \mathbb{E}\left[\mathbb{E}\left[P_\nu(q_L)^{N([0,T))}|T\right]\right] \\ &= \mathbb{E}\left[e^{R_L(P_\nu(q_L)-1)\int_0^T g(t)\,dt}\right] \\ &= \int_0^\infty \rho_G e^{-\rho_G t} e^{R_L(P_\nu(q_L)-1)\int_0^t g(u)\,du}\,dt, \end{aligned}$$

giving us the relation $q_L = \int_0^\infty \rho_G e^{-\rho_G t} e^{R_L(P_\nu(q_L)-1)\int_0^t g(u)\,du}\,dt$.

Proceeding similarly gives us

$$q_G = \int_0^\infty \rho_G e^{-\rho_G t} e^{R_L G(t)(P_\nu(q_L)-1)}\,dt.$$

C.4. Bursts of infections

Suppose each individual waits a randomly distributed time with probability density function $g(t)$ before causing a random number ν of infections, independent of the time of transmission. Fix a given individual, and suppose that they are transferred at time T', exponentially distributed with rate ρ_G and that their transmission event happens at time T. Then, their local chain of infection goes extinct if either they are transferred prior to transmission, or if not, if all those they infect have finite chains of infection:

$$\begin{aligned} q_L &= \mathbb{E}\left[\mathbb{1}_{\{T'<T\}} + \mathbb{1}_{\{T'<T\}} q_L^\nu\right] \\ &= \mathbb{P}\{T'<T\} + \mathbb{P}\{T'>T\}\mathbb{E}\left[q_L^\nu\right]. \end{aligned}$$

Again, the expectation on the right is the probability generating function for ν evaluated at q_L, $P_\nu(q_L)$.

The calculation thus depends on the choice of law for the random variables ν_i, or equivalently the choice of distribution and its probability generating function $P_\nu(z)$. For example, if each ν_i is negatively binomially distributed with mean R_L and k successes (and thus success probability $p = \frac{R_L}{k+R_L}$) we have $P_\nu(z) = \left(\frac{k}{k+R_L(1-q_L)}\right)^k$, whereas for a Poisson process with the same mean R_L we have $P_\nu(z) = e^{R_L(z-1)}$.

On the other hand,

$$\begin{aligned} \mathbb{P}\{T'>T\} &= \mathbb{E}\left[\mathbb{P}\left\{T'>T|T\right\}\right] \\ &= \mathbb{E}\left[e^{-\rho_G T}\right] \\ &= \int_0^\infty e^{-\rho_G t} g(t)\,dt \\ &= \hat{g}(\rho_G), \end{aligned}$$

where $\hat{g}(p)$ indicates the Laplace transform of $g(t)$. Thus,

$$q_L = (1-\hat{g}(\rho_G)) + \hat{g}(\rho_G) P_\nu(q_L).$$

For the initial individual, we need to take into account the possibility that the individual had their "burst" prior to being transferred to the focal site,

$$\begin{aligned} q_G &= \mathbb{P}\{T'>T\} + \mathbb{P}\{T'<T\}\mathbb{E}\left[q_L^\nu\right] \\ &= \hat{g}(\rho_G) + (1-\hat{g}(\rho_G)) P_\nu(q_L) \end{aligned}$$

$$= \hat{g}(\rho_G) + (1-\hat{g}(\rho_G))\frac{q_L - (1-\hat{g}(\rho_G))}{\hat{g}(\rho_G)}$$
$$= \hat{g}(\rho_G) + \frac{(1-\hat{g}(\rho_G))(\hat{g}(\rho_G) - p_L)}{\hat{g}(\rho_G)}$$

or, rearranging,

$$p_G = p_L \frac{1-\hat{g}(\rho_G)}{\hat{g}(\rho_G)}.$$

C.5. Threshold mixing rates

Synthesizing the above results to derive a threshold rate of transfers, we recall that the criterion for a major outbreak on the congregate level is $R_* > 1$, where $R_* = \mu\mathbb{E}[Z]$, μ is the (size-biased) mean size of a major outbreak, and Z is the number of facilities in which a given individual causes a major outbreak. Assuming that each individual is transferred at most once, then $\mathbb{E}[Z] = p_G = 1-q_G$, where q_G is calculated as above. To calculate the critical transfer rate ρ_G^*, one needs to solve $R_* = 1$ for ρ_G.

Appendix D. The final size of a large well-mixed epidemic

In this appendix, we sketch an argument that, in the limit of large population sizes and under well-mixed assumptions, the final size of a closed epidemic (*i.e.* the fraction of susceptible hosts who have been infected over the duration of the epidemic, assuming no new individuals are introduced) is equal to that for the Kermack–McKendrick SIR epidemic for a wide variety of models, provided all individuals are indistinguishable. As will become clearer as we formulate the argument, the result applies equally well to *e.g.* a SIR or SEIR epidemic, but does not apply when *e.g.* individuals vary in their susceptibility or in the size of their contact networks.

We consider the epidemic from a retrospective vantage point. Suppose that the local population consists of n individuals, labelled (arbitrarily) $1,\dots,n$. Let $\mathbb{1}_i$ be an indicator variable, equal to 1 if the individual with label i is eventually infected, and 0 if they are never infected. Let N_i be the number of *potential* transmissions by individual i: if individual i is never infected ($\mathbb{1}_i = 0$), then none of these potential transmissions result in a new infection; if on the other hand, i is eventually infected ($\mathbb{1}_i = 1$), then the potential infections result in new infections if the contacted individual is susceptible. Thus, $\mathbb{E}[N_i] = R_L$ The variables N_i and $\mathbb{1}_i$ are assumed to be independent (*e.g.* we exclude the possibility that some individuals are highly susceptible and highly transmissive, as might occur with a weakened immune system.

Let $R_n(\infty) = \sum_{i=1}^n \mathbb{1}_i$ denote the total number of individuals who were infected over the course of the epidemic, and let $S_n(\infty) = n - R_n(\infty)$ be the number who were never infected.

Now, consider the probability, $p_1 = \mathbb{P}\{\mathbb{1}_1 = 0\}$, that the individual with label 1 was never infected; because the labelling is arbitrary, this is simply the probability that we uniformly drew the one of the $S_n(\infty)$ individuals that were never infected, $\frac{S_n(\infty)}{n} = 1 - \frac{R_n(\infty)}{n}$. Since all individuals are identical up to an arbitrary labelling, $p_i = p_1$ for all i.

On the other hand, the probability that the first individual was never infected can be written as the probability that they were not "chosen" as a contact by an infectious individual (under the well-mixed assumption, any individual i has a contact with individual 1 with probability $\frac{1}{n-1}$):

$$\prod_{i=2}^n \left(1-\frac{1}{n-1}\right)^{N_i\mathbb{1}_i} = e^{\sum_{i=2}^n N_i\mathbb{1}_i \ln\left(1-\frac{1}{n-1}\right)}$$
$$= e^{-\sum_{i=2}^n N_i\mathbb{1}_i\left(\frac{1}{n-1}+O\left(\frac{1}{n^2}\right)\right)}.$$

Now, consider the sum $\sum_{i=2}^n N_i\mathbb{1}_i$. Let $\tilde{R}_n(\infty)$ be the number of individuals with labels $\{2,\dots,n\}$ who were infected (thus $\tilde{R}_n(\infty) = R_n(\infty)$ or $\tilde{R}_n(\infty)-1$ according to whether individual 1 was infected or not). If the limit

$$r(\infty) = \lim_{n\to\infty} \frac{\tilde{R}_n(\infty)}{n}$$

exists and is non-zero, then the central limit theorem tells us that

$$\lim_{n\to\infty} \sum_{i=2}^n \frac{N_i\mathbb{1}_i}{n-1} = \lim_{n\to\infty} \frac{\tilde{R}_n(\infty)}{n-1} \sum_{\substack{i=2\\ \mathbb{1}_i=1}}^n \frac{N_i}{\tilde{R}_n(\infty)}$$
$$= r(\infty)\mathbb{E}[N_i]$$
$$= r(\infty)R_L.$$

On the other hand, if $r(\infty) = 0$, which would be the case if the branching process approximating the epidemic went extinct, then for any $\varepsilon > 0$ Markov's inequality tells us that

$$\mathbb{P}\left\{\sum_{i=2}^n \frac{N_i\mathbb{1}_i}{n-1} > \varepsilon\right\} \le \frac{1}{\varepsilon}\mathbb{E}\left[\sum_{i=2}^n \frac{N_i\mathbb{1}_i}{n-1}\right]$$
$$= \frac{1}{\varepsilon}\frac{R_L\mathbb{E}[\tilde{R}_n(\infty)]}{n-1},$$

which tends to 0 as $n\to\infty$. Since $\varepsilon > 0$ is arbitrary, the sum is zero with probability 1.

Equating the two expressions for the probability that the first individual is never infected, and taking the limit as $n\to\infty$, we get the tautology $1 - r(\infty) = 1$ if $r(\infty) = 0$, and the Kermack–McKendrick final size formula otherwise:

$$1 - r(\infty) = e^{-R_L r(\infty)}$$

Rewriting this in terms of the final fraction susceptible, $s(\infty) = 1-r(\infty)$, we get $s(\infty) = e^{R_L(1-s(\infty))R_L}$, or, equivalently, $-R_L s(\infty)e^{-R_L s(\infty)} = -R_L e^{-R_L}$. Thus,

$$s(\infty) = -\frac{1}{R_L}W_0\left(-R_L e^{-R_L}\right),$$

where $W_0(x)$ is the principal branch of Lambert's W-function (Corless et al., 1996), the transcendental (multi-)function satisfying $x = W(x)e^{W(x)}$, where the principal branch is real valued and increasing on $[-e^{-1},\infty)$.

Moreover, $W_0(x) = x + O(x^2)$, so

$$s(\infty) = e^{-R_L} + O(R_L e^{-2R_L}).$$

For $R_L = 4$, we have that $e^{-R_L} \approx 0.0015$, whereas $s(\infty) \approx 0.0198$, so the relative error in approximating $s(\infty)$ by e^{-R_L} is already less than 8%, whereas approximately 98% of the population will have been infected, thus justifying the use of the mean household size as a reasonable upper bound on the final size of the epidemic.

Appendix E. The final size of a small outbreak

In what follows, we derive an estimate for the final size of a local outbreak conditioned on extinction: for reasons of tractability, we will compute the final size of an outbreak assuming no transfers, and assuming that patient zero arrives at the time of their infection.

E.1. Conditioning on extinction

Let $N = N([0,\infty))$ denote, as previously, the total number of infections caused by a focal individual in any of the models discussed in Appendix B, and let

$$\mathcal{F}(z) = \mathbb{E}[z^N] = \sum_{n=0}^\infty \mathbb{P}\{N=n\}z^n$$

be the corresponding PGF. Then, as previously the unique solution $0 < \tilde{q} < 1$ to $\tilde{q} = \mathcal{F}(\tilde{q})$ is the probability that the branching process of infections (*i.e.* as before, we ignore exhaustion of susceptible individuals) eventually goes extinct. Now, consider the process conditioned on extinction: the branching process goes extinct if and only if the branching processes initiated by all individuals infected by patient zero also go extinct; the independence of these secondary branching

processes means the probability that each goes extinct is independent wit probability $\tilde{q}$, and thus,

$$\mathbb{P}\{N = n|\text{extinction}\} = \frac{\mathbb{P}\{N = n, \text{extinction}\}}{\mathbb{P}\{\text{extinction}\}} = \frac{\mathbb{P}\{N = n\}(\tilde{q})^n}{\tilde{q}}.$$

In particular, the PGF for the total number of infections conditioned on extinction is $\mathcal{F}_{ext}(z) = \frac{\mathcal{F}(\tilde{q}z)}{\tilde{q}}$.

E.2. The PGF for the final size

In what follows, we will refer to *generations* of infections: the first generation consists of all individuals infected by patient zero, the second generation consists of all individuals infected by individuals in the first generation, *etc.* Let X_k be the number of infected individuals in the k^{th} generation, and let

$$Y_k = \sum_{i=0}^{k} X_i$$

be the total number infected in the first k generations. Let $\mathcal{G}_k(w) = \mathbb{E}[w^{Y_k}|\text{extinction}]$ be the PGF for Y_k conditional on extinction. Now, suppose that patient zero infects N individuals, who each give rise to branching processes with $X_k^{(i)}$ offspring in the k^{th} generation and $Y_k^{(i)}$ total offspring by the k^{th} generation. Then,

$$\mathcal{G}_k(w) = \mathbb{E}\left[w^{1+\sum_{i=1}^{N} Y_{k-1}^{(i)}} \,\middle|\, \text{extinction}\right]$$

(n.b. the term 1 counts patient zero in the total outbreak size)

$$\begin{aligned}
&= \mathbb{E}\left[w \prod_{i=1}^{N} w^{Y_{k-1}^{(i)}} \,\middle|\, \text{extinction}\right] \\
&= \mathbb{E}\left[w\mathbb{E}\left[w^{Y_{k-1}} \,\middle|\, \text{extinction}\right]^N \,\middle|\, \text{extinction}\right] \\
&= \mathbb{E}\left[w\mathcal{G}_{k-1}(w)^N \,\middle|\, \text{extinction}\right] \\
&= w\mathcal{F}_{ext}(\mathcal{G}_{k-1}(w))
\end{aligned}$$

Taking $k \to \infty$, we get a functional equation for the final size of the epidemic conditioned on extinction,

$$\mathcal{G}_\infty(w) = w\mathcal{F}_{ext}(\mathcal{G}_\infty(w)). \tag{E.1}$$

E.3. The Poisson final size distribution

In general, no exact solution for (E.1) is available, although we can use Lagrange's inversion formula to obtain the coefficients of the series for $\mathcal{G}_\infty(w)$ from those of $\mathcal{F}_{ext}(z)$. In the special case when N is a Poisson(R_L) random variable, we have $\mathcal{F}(z) = e^{R_L(z-1)}$, so that $\tilde{q} = e^{R_L(\tilde{q}-1)}$. Thus, $-R_L\tilde{q}e^{-R_L\tilde{q}} = -R_Le^{-R_L}$, which we can solve as above using Lambert's W function to obtain

$$\tilde{q} = -\frac{1}{R_L} W_0\left(-R_Le^{-R_L}\right),$$

whereas $\mathcal{F}_{ext}(z) = \frac{1}{\tilde{q}}e^{\tilde{q}R_L(z-1)}$. Then, $\mathcal{G}_\infty(w) = \frac{w}{\tilde{q}}e^{\tilde{q}R_L(w-1)}$, so $-R_L\mathcal{G}_\infty(w)e^{-R_L\mathcal{G}_\infty(w)} = -wR_Le^{-R_L}$, and again

$$\mathcal{G}_\infty(w) = -\frac{1}{\tilde{q}R_L} W_0\left(-wR_Le^{-R_L}\right) = \frac{W_0\left(-wR_Le^{-R_L}\right)}{W_0\left(-R_Le^{-R_L}\right)}.$$

Using the power series $W_0(x) = \sum_{n=0}^{\infty} \frac{(-n)^{n-1}}{n!}x^n$ (obtained using Lagrange's inversion formula), we get

$$\mathcal{G}_\infty(w) = -\frac{1}{W_0\left(-R_Le^{-R_L}\right)} \sum_{n=0}^{\infty} \frac{n^{n-1}\left(R_Le^{-R_L}\right)^n}{n!} w^n,$$

whence $\mathbb{P}\{W = n|\text{extinction}\} = -\frac{n^{n-1}\left(R_Le^{-R_L}\right)^n}{n!W_0\left(-R_Le^{-R_L}\right)}$.

While this could be used to establish a probabilistic cut-off to characterize small *vs,* large epidemics in the Poisson case (*i.e.* we could define an outbreak as large if its size, say N_{out}, is such that $\mathbb{P}\{W = N_{out}|\text{extinction}\} < \varepsilon$ for some small $\varepsilon > 0$), we will instead use the mean size of a small epidemic to make the distinction (*i.e.* we will define an epidemic as large if its size is greater than some multiple of the mean small size).

E.4. The expected final size of a small outbreak

We thus turn our attention to computing the expected final size of an outbreak, which we recall is $\mathcal{G}'_\infty(1)$ (*n.b.* $\mathcal{G}_\infty(1) = 1$). Implicitly differentiating (E.1), we have

$$\mathcal{G}'_\infty(w) = \mathcal{F}_{ext}(\mathcal{G}_\infty(w)) + w\mathcal{F}'_{ext}(\mathcal{G}_\infty(w))\mathcal{G}'_\infty(w),$$

whence

$$\mathcal{G}'_\infty(w) = \frac{\mathcal{F}_{ext}(\mathcal{G}_\infty(w))}{1 - w\mathcal{F}'_{ext}(\mathcal{G}_\infty(w))}$$

and $\mathcal{G}'_\infty(w) = \frac{1}{1-\mathcal{F}'_{ext}(1)}$. Finally, $\mathcal{F}'_{ext}(z) = \mathcal{F}'(\tilde{q}z)$, so $\mathcal{F}'_{ext}(1) = \mathcal{F}'(\tilde{q})$, so

$$\mathcal{G}'_\infty(1) = \frac{1}{1 - \mathcal{F}'(\tilde{q})}$$

(n.b. in Appendix A, we show that $\mathcal{F}'(\tilde{q}) < 1$). Recalling that $\mathcal{G}''_\infty(w)$ gives the second factorial moment of the outbreak size, one may proceed similarly to show that the variance in the outbreak size is

$$\frac{\tilde{q}\mathcal{F}''(\tilde{q}) + \mathcal{F}'(\tilde{q})(1 - \mathcal{F}'(\tilde{q}))}{(1 - \mathcal{F}'(\tilde{q}))^3}.$$

In the Poissonian case, we have $\mathcal{F}'(z) = R_Le^{R_L(z-1)}$. Recalling that $W_0(x)e^{W_0(x)} = x$, this gives us

$$\mathcal{F}'(\tilde{q}) = R_Le^{R_L(\tilde{q}-1)} = -R_Le^{-R_L}e^{-W_0\left(-R_Le^{-R_L}\right)} = -W_0\left(-R_Le^{-R_L}\right) = R_L\tilde{q},$$

whereas $\mathcal{F}''(\tilde{q}) = R_L\mathcal{F}'(\tilde{q}) = R_L\tilde{q}^2$. In particular, the mean (standard deviation) of the final size is approximately 1.25 (0.62) and 1.0018 (0.043) for our small and large values for R_L, 2.87 and 8.44. Applying the 99% rule, we take outbreaks of 3 or fewer individuals to be small.

In the negative binomial case, we have $\mathcal{F}(z) = \left(\frac{1-p}{1-pz}\right)^k$, where $p = \frac{R_L}{R_L+k}$ (see Appendix B). No exact solution is available for $\tilde{q}$, but we can compute

$$\mathcal{F}'(z) = k\left(\frac{1-p}{1-pz}\right)^{k-1} \frac{p(1-p)}{(1-pz)^2} = \left(\frac{1-p}{1-pz}\right)^k \frac{kp}{1-pz} = \mathcal{F}(z)\frac{kp}{1-pz} \tag{E.2}$$

and

$$\mathcal{F}'(\tilde{q}) = \mathcal{F}(\tilde{q})\frac{kp}{1-p\tilde{q}} = \frac{kp\tilde{q}}{1-p\tilde{q}} = \frac{R_Lk\tilde{q}}{R_L(1-\tilde{q})+k} \le R_L\tilde{q}.$$

n.b. this inequality is not sufficient to show that the expected final size with the negative binomial distribution is smaller than that for the Poisson distribution, as the extinction probability $\tilde{q}$ is larger for the former than for the latter (Lloyd-Smith et al., 2005). We can, however, find a condition under which this is true: recall from Appendix A that $\tilde{q} < \tilde{q}'$, where $\mathcal{F}'(\tilde{q}') = 1$. We will use this to obtain an upper bound on $\tilde{q}$. Using (E.2), we have

$$\frac{1-p\tilde{q}'}{kp} = \mathcal{F}(\tilde{q}') = \left(\frac{1-p}{1-p\tilde{q}'}\right)^k.$$

Substituting $p = \frac{R_L}{R_L+k}$ yields

$$\frac{R_L(1-\tilde{q}')+k}{kR_L} = \left(\frac{k}{R_L(1-\tilde{q}')+k}\right)^k,$$

whence

$$\frac{1}{R_L} = \left(\frac{k}{R_L(1-\tilde{q}')+k}\right)^{k+1} = (\mathcal{F}(\tilde{q}'))^{\frac{k+1}{k}}.$$

Since $\tilde{q} \leq \tilde{q}'$ and $\mathcal{F}(z)$ is increasing, we have $\mathcal{F}(\tilde{q}') \geq \mathcal{F}(\tilde{q}) = \tilde{q}$, giving us $\tilde{q} \leq R_L^{-1-\frac{1}{k}}$. Now, the final size of the Poisson process is larger than that for the negative binomial if and only if

$$\frac{R_L k\tilde{q}}{R_L(1-\tilde{q})+k} \leq -W_0\left(-R_L e^{-R_L}\right).$$

$\frac{R_L kx}{R_L(1-x)+k}$ is an increasing function of x, so the left hand side is bounded above by $\frac{R_L k\tilde{q}'}{R_L(1-\tilde{q}')+k} = \frac{kR_L^{-\frac{1}{k}}}{R_L+k-R_L^{-\frac{1}{k}}}$, which is in turn less than $\frac{kR_L^{-\frac{1}{k}}}{R_L+k-1}$, giving us the sufficient condition

$$\frac{kR_L^{-\frac{1}{k}}}{R_L+k-1} \leq -W_0\left(-R_L e^{-R_L}\right).$$

Solving for k (details omitted),

$$k \leq \frac{\ln R_L(R_L-1)}{(R_L-1)W_0\left(\frac{-\ln R_L R_L^{\frac{1}{R_L-1}}}{W_0(-R_l e^{-R_L})(R_L-1)}\right) - \ln R_L}.$$

The right hand side is decreasing on $R_L \in [1,\infty)$ and greater than 0.54 for $R_L \leq 9$, so we conclude that the final size under a Poisson distribution is an upper bound for the final size using a negative binomial distribution for the parameters used here ($R_L = 2.87$ and $R_L = 8.44$, $k = 0.5$).

Finally, we note that in the burst model, $N = \nu_1$, where ν_1 is an arbitrary distribution with expected value R_L, and we cannot compute anything further without specifying the law of ν_1; here, we took ν_1 to be negative binomial distribution above, so that the final size with Poisson distribution remains an upper bound.

References

Adam, Dillon C., Wu, Peng, Wong, Jessica Y., Lau, Eric H.Y., Tsang, Tim K., Cauchemez, Simon, Leung, Gabriel M., Cowling, Benjamin J., 2020. Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. Nature Med. 26 (11), 1714–1719.

Althouse, Benjamin M., Wenger, Edward A., Miller, Joel C., Scarpino, Samuel V., Allard, Antoine, Hébert-Dufresne, Laurent, Hu, Hao, 2020. Stochasticity and heterogeneity in the transmission dynamics of SARS-CoV-2. arXiv:2005.13689 [physics, q-bio], arXiv:2005.13689.

2020a. In re Von Staich, on Habeas Corpus (A160122). https://www.courts.ca.gov/opinions/documents/A160122.PDF. (Accessed 29 January 2021). Secondary URL: https://law.justia.com/cases/california/court-of-appeal/2020/a160122.html.

2020b. Joint Case Management Conference Statement (Case 4:01-cv-01351-JST Document 3477). https://rbgg.com/wp-content/uploads/Dkt-3477-PLATA-Joint-CMC-Statement-11-04-20-489-ovr.pdf. (Accessed 29 January 2021).

Arif Billah, Md., Mamun Miah, Md., Nuruzzaman Khan, Md., 2020. Reproductive number of coronavirus: A systematic review and meta-analysis based on global level evidence. PLoS One 15 (11), 1–17.

Ball, Frank, Mollison, Denis, Scalia-Tomba, Gianpaolo, 1997. Epidemics with two levels of mixing. Ann. Appl. Probab. 7 (1), 46–89.

Bartlett, M.S., 1956. Deterministic and stochastic models for recurrent epidemics. In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, vol. 4, pp. 81–108.

California Department of Corrections and Rehabilitation, 2021a. COVID-19 Screening and testing for patient movement. https://www.cdcr.ca.gov/covid19/wp-content/uploads/sites/197/2021/01/COVID-19-Screening-and-Testing-Matrix-Final-21-01-08.pdf. (Accessed 4 February 2021).

California Department of Corrections and Rehabilitation, 2021b. Population COVID-19 tracking. https://www.cdcr.ca.gov/covid19/population-status-tracking/. (Accessed 26 March 2021).

California Department of Corrections and Rehabilitation, 2021c. Weekly report of population as of midnight March 24, 2021. https://www.cdcr.ca.gov/research/wp-content/uploads/sites/174/2021/03/Tpop1d210324.pdf. (Accessed 26 March 2021).

Corless, R.M., Gonnet, G.H., Hare, D.E.G., Jeffrey, D.J., Knuth, D.E., 1996. On the LambertW function. Adv. Comput. Math. 5 (1), 329–359.

Covid Behind Bars, 2021. UCLA Law COVID behind bars data. https://github.com/uclalawcovid19behindbars/data. (Accessed 26 March 2021).

Crump, K.S., Mode, C.J., 1968. A general age-dependent branching process. I. J. Math. Anal. Appl. 24 (3), 494–508.

Crump, K.S., Mode, C.J., 1969. A general age-dependent branching process. II. J. Math. Anal. Appl. 25 (3), 8–17.

Ferretti, Luca, Ledda, Alice, Wymant, Chris, Zhao, Lele, Ledda, Virginia, Abeler-Dörner, Lucie, Kendall, Michelle, Nurtay, Anel, Cheng, Hao-Yuan, Ng, Ta-Chou, Lin, Hsien-Ho, Hinch, Rob, Masel, Joanna, Marm Kilpatrick, A., Fraser, Christophe, 2020. The Timing of COVID-19 Transmission. SSRN Scholarly Paper ID 3716879, Social Science Research Network, Rochester, NY.

Fortuna, Lisa.R., Tolou-Shams, Marina., Robles-Ramamurthy, Barbara., Porche, Michelle.V., 2020. Inequity and the disproportionate impact of COVID-19 on communities of color in the united states: The need for a trauma-informed social justice response. Psychol. Trauma.

Franco-Paredes, Carlos, Ghandnoosh, Nazgol, Latif, Hassan, Krsak, Martin, Henao-Martinez, Andres F., Robins, Megan, Barahona, Lilian Vargas, Poeschla, Eric M., 2020. Decarceration and community re-entry in the COVID-19 era. Lancet Infectious Diseases.

Jagers, P., 1975. Branching Processes with Biological Applications. Wiley, London.

Jagers, P., 1992. Stabilities and instabilities in population dynamics. J. Appl. Probab. 770–780.

Lee, Ian.J., Ahmed, Nasar.U., 2021. The devastating cost of racial and ethnic health inequity in the covid-19 pandemic. J. National Med. Assoc. 113 (1), 114–117.

Lloyd-Smith, J.O., Schreiber, S.J., Kopp, P.E., Getz, W.M., 2005. Superspreading and the effect of individual variation on disease emergence. Nature 438 (7066), 355–359.

Miller, Leila, 2021. 40% Of inmates in California's corrections system have been vaccinated for COVID-19. In: Los Angeles Times. Section, California.

Office of the Inspector General, 2020. COVID–19 Review Series, Part Two: The California Department of Corrections and Rehabilitation Distributed and Mandated the Use of Personal Protective Equipment and Cloth Face Coverings; However, Its Lax Enforcement Led to Inadequate Adherence to Basic Safety Protocols. https://www.oig.ca.gov/wp-content/uploads/2020/10/OIG-COVID-19-Review-Series-Part-2-%E2%80%93-Face-Coverings-and-PPE.pdf. (Accessed 29 October 2020).

Office of the Inspector General, 2021. COVID–19 Review series, part three: california correctional health care services and the California department of corrections and rehabilitation caused a public health disaster at San Quentin State prison when they transferred medically vulnerable incarcerated persons from the California Institution for men without taking proper safeguards. https://www.oig.ca.gov/wp-content/uploads/2021/02/OIG-COVID-19-Review-Series-Part-3-%E2%80%93-Transfer-of-Patients-from-CIM.pdf. (Accessed 4 February 2021).

Okonkwo, Nneoma E., Aguwa, Ugochi T., Jang, Minyoung, Barré, Iman A., Page, Kathleen R., Sullivan, Patrick S., Beyrer, Chris, Baral, Stefan, 2020. COVID-19 and the US response: accelerating health inequities. BMJ Evidence-Based Med..

Picard, Philippe, Lefevre, Claude, 1990. A unified analysis of the final size and severity distribution in collective reed-frost epidemic processes. Adv. Appl. Probab. 22 (2), 269–294.

Puglisi, Lisa B., Malloy, Giovanni S.P., Harvey, Tyler D., Brandeau, Margaret L., Wang, Emily A., 2020. Estimation of COVID-19 basic reproduction ratio in a large urban jail in the United States. Ann. Epidemiol..

Susswein, Zachary, Bansal, Shweta, 2020. Characterizing superspreading of SARS-CoV-2: from mechanism to measurement. MedRxiv https://www.medrxiv.org/content/early/2020/12/11/2020.12.08.20246082.

The Amend Project, 2020. Urgent memo COVID-19 outbreak: San quentin prison. https://amend.us/wp-content/uploads/2020/06/COVID19-Outbreak-SQ-Prison-6.15.2020.pdf. (Accessed 29 January 2021).

The Marshall Project, 2020. 1 in 5 prisoners in the U.S. has had COVID-19. https://www.themarshallproject.org/2020/12/18/1-in-5-prisoners-in-the-u-s-has-had-covid-19, Published in collaboration with The Associated Press. (Accessed: January 29, 2021).

The New York Times, 2021. Coronavirus in the United States: Latest map and case count. https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html. (Accessed 29 January 2021).