A simple model of how varying exposure at gatherings impacts the emergence of variants and their detection

Bradford P. Taylor and William P. Hanage

Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA

October 25, 2023

Abstract

Understanding how epidemics spread within societies is key for establishing adequate infection control responses. Dynamical models provide a means to translate surveillance data into predictions of future disease spread, yet many epidemic models do not capture empirically observed features of socialization. Here, we build epidemic models that incorporate transmission heterogeneity, homophily, and awareness of the risk of infection by explicitly modeling attendance at gatherings. Our risk-SIR model captures how diseases tend to spread initially within more social subpopulations and sequentially spread to less social subpopulations. This dynamic complicates surveillance as many metrics become dynamic when comorbidities vary with socialization. We also show fast-growing lineages that emerge within highly social subpopulations can be misidentified as potential variants of concern, and that these lineages will transiently grow in frequency as they remain distributed among more social individuals. Finally, we show that the structure of society, namely the amount of homophily where similarly social individuals preferentially interact, sets the magnitude and duration of our observed biases, making spurious variants of concern all but inevitable. Together, we provide a simple framework for incorporating socialization and behavior in epidemic models and argue that this is necessary to contextualize estimates during surveillance.

Introduction

Directly transmitted infectious agents rely on the behavior of their hosts to start new infections and maintain themselves – specifically the contacts that

their hosts make along which transmission may then occur. The contact networks that link individuals in a population are hence the routes along which communicable diseases spread, and considerable effort has been made to understand the nature of these networks and the implications for control of infectious agents [3]. The numbers of contacts individuals make can vary greatly, just as the nature of contacts can differ; some may be merely fleeting opportunities for a pathogen to initiate infection while others are sustained exposures to a large inoculum. Such heterogeneity in secondary contact rates appears to be common across epidemics, given over-dispersion of empirically observed secondary infection distributions, meaning that few individuals disproportionately contribute more towards ongoing transmission [25, 44, 23, 35]. While some of this may be due to infections with unusual properties, those infected must still make contacts in order to transmit to them [32, 10, 24, 22, 14, 7].

Another feature that contact networks exhibit is assortativity: the tendency for contacts to be made between nodes with similar properties [29]. In social networks this is termed homophily, and in the case of communicable disease the consequence is that most transmission events take place among groups of similar individuals, which can be explicitly incorporated with a contact matrix (e.g., for age-structured models) [27, 33]. Varying contact networks can be incorporated into dynamical models of transmission and have produced insights into subjects such as the optimal allocation of vaccines (those which prevent transmission are more effective when used in those with the most contacts) or the importance of core groups in maintaining sexually transmitted infections and as a focus of interventions [36, 9].

A special case of assortativity and homophily is when individuals share increased risks of exposure, and increased numbers of contacts – for example through gathering in groups of larger size. Larger gatherings both increase the chance that an infectious person attends, and the numbers to whom they may transmit [1, 13, 6]. The resulting variations in risk of exposure may result from voluntary social interactions (and contact rates tend to be higher in younger people), but can also include sexual networks in which the majority of at-risk contacts occur between a small proportion of the population, or crowded settings such as workplaces in which employees are unable to mitigate their own risks of exposure (e.g., meatpacking plants in the early stages of the covid pandemic are a well-known case of workplaces in which large outbreaks occurred) [45, 40, 37]. In the context of an emerging pathogen or variant such transmission heterogeneity and homophily mean that any estimates from local surveillance efforts may not reflect the epidemic potential in other communities. The social context of disease spread can produce biases in surveillance given differences in resources between settings.

Mathematical models may be used to project the expected consequences of outbreaks. For example, in the classic SIR model for a directly transmitted pathogen the basic reproduction number R_0, determines the final size of the epidemic, but the model makes well known unrealistic assumptions about mixing patterns and must be extended to account for heterogeneity and homophily [20, 16]. Network models haven proven informative means to interrogate the effect

of social structure on epidemic spread, yet by their nature they are demanding to parameterize [19, 18]. Nonetheless, capturing these heterogeneous contacts is necessary to contextualize the rapid growth of an epidemic as a result of the social context where the outbreak arises, rather than specific properties of the pathogen. This sort of "founder effect" is familiar in population genetics, but less so in epidemiology [26, 38].

Here we present a simple extension of the classic SIR framework that specifically includes gathering sizes and homophily among individuals on the basis of the risk of infection. The resulting case curves are related to the gathering sizes and enable us to quantify how incorporating socialization alters our expectations for early spread of emerging viruses and variants and how this is reflected in seroprevalence and genomic surveillance.

Transmission at gatherings: Group-SIR model

We propose modifying SIR models to account for varying contact rates and risk by explicitly modeling gathering dynamics using sampling processes. For any SIR-like model there are two key events, transmission and recovery, yielding infection dynamics:

$$\frac{dI}{dt} = \overbrace{B(S,I,R)}^{\text{transmission}} - \overbrace{\gamma I}^{\text{recovery}}$$

We focus on how gathering dynamics alter transmission, $B(S,I,R)$ and ignore the effects of different recovery rates by normalizing time by the mean infectious period, $\frac{1}{\gamma}$. Transmission can occur between each pair of the n attendees who were randomly sampled from the overall population. Figure 1a illustrates this sampling process for the classic SIR model where n=2 and Figure 1b illustrates this for the generalized "group-SIR" model. The expected number of infections at a gathering determines the overall transmission rate:

$$B_{\text{group}}(n,S,I,R) = \mathbb{E}[\Delta I] = \sum_{\{s,i,r\}} s\mathbb{P}_{\text{group}}(s,i,r|S,I,R)\mathbb{P}_{\text{infect}}(i)$$

with probabilities

$$\mathbb{P}_{\text{infect}}(i) = 1-(1-p_\beta)^i,$$
$$\mathbb{P}_{\text{group}}(s,i,r|S,I,R) = \frac{n!}{s!i!r!}\left(\frac{S}{N}\right)^s\left(\frac{I}{N}\right)^i\left(\frac{R}{N}\right)^r$$

where lower-case s, i, and r denote the integer number of sampled susceptibles, infected, and recovered attending the gathering and uppercase S, I, R denote the population-level densities such that $S+I+R=N$. To average, we sum over all possible combinations of n SIR attendees sampled from the population densities according to a multinomial distribution, $\mathbb{P}_{\text{group}}$. A susceptible is infected

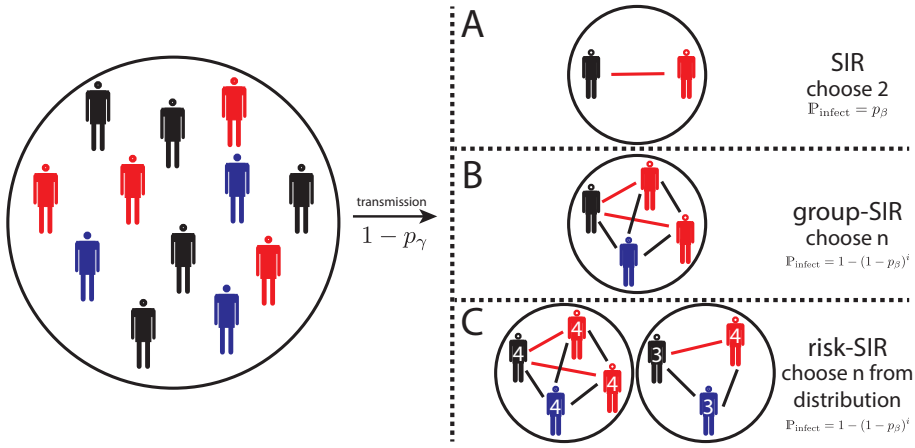

Figure 1: Schematic of sampling processes leading to different epidemic models described. A) The classic SIR model arises from sampling 2 individuals from a population of susceptible (black), infected (red), and recovered (blue). B) Sampling n individuals yields a generalization of the SIR model, the group-SIR. Infections occur independently between each S-I pair present in a congregation. C) Denoting individuals by the maximum congregation size they are willing to attending yields the risk-SIR model. Societies can be defined by a distribution of congregation sizes which determine the epidemiological dynamics [4].

at the gathering with probability, $\mathbb{P}_{\text{infect}}$ whenever any attending infected transmits according to a pairwise probability of infection, p_β. The dynamics simplify to a classic SIR model when gatherings are restricted to pairs, n=2, whereas the transmission term becomes increasingly nonlinear for larger gatherings.

Increasing gathering sizes leads to sharper epidemics for otherwise fixed parameter values (Figure 2a). This is expected since increasing group sizes implicitly increases the total number of contacts relative to the recovery rate, i.e., increasing the basic reproduction number, R_0. Rescaling the transmission rate by the total number of pairwise interactions at a gathering, $\binom{n}{2}$, collapses the group-SIR dynamics onto each other (Figure 2a inset). There are fewer cases at the epidemic peak for larger gatherings as a result of multiple infectious individuals infecting the same susceptible, effectively reducing the number of pairwise interactions. In short, congregating in larger groups increases the rate of epidemic spread, but has minor impact on the dynamics when correcting for the number of interactions except by reducing the cumulative number of infections, in line with prior observations [41].

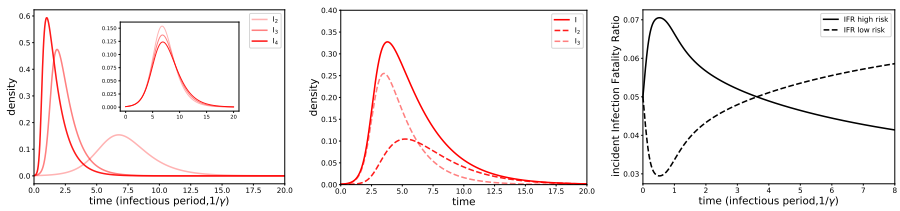

Figure 2: Generalized SIR dynamics. A) Group-SIR dynamics, where transmission occurs within groups with sizes here denoted by the subscript. B) Risk-SIR dynamics, where transmission occurs within groups according to a distribution here set with sizes n=2,3, and 4 at equal frequency. Subpopulations are denoted by their risk level shown, which specifies the maximum gathering size they are willing to attend. C) The sequential peaks of infection cause the overall infection fatality ratio (IFR) to be dynamic when it varies between subpopulations. Here we compare IFR dynamics when the IFR is positively correlated with risk level (IFR high risk) and when the IFR is negatively correlated with risk level (IFR low risk).

Gathering size dependent attendance: Risk-SIR model

During infectious disease outbreaks individuals may vary in how much they alter their behavior in response to perceived risk [30, 8, 2, 15]. To capture this we can further generalize the group-SIR dynamics where individuals choose among differently sized, concurrent gatherings and avoid those larger than their "risk level":

$$\frac{dI_k}{dt} = B_{\text{risk}}(k) - \gamma I_k$$

where risk levels are denoted by subscripts. Transmission follows from repeatedly sampling from the population for a given distribution of differently sized gatherings, $\mathbb{P}_n$:

$$B_{\text{risk}} = \sum_{n=2}^{k} \mathbb{P}_n B_{\text{group}}(n; S_k, \sum_{m \geq n} I_m, -S_k + \sum_{m \geq n} (R_m + S_m))$$

Each transmission term captures new infections within a particular focal subpopulation of individuals with the same risk level, k, such that only the respective susceptibles, S_k, contribute. These focal susceptibles can be infected when attending any gathering no larger than their risk level, $n \leq k$, by any attending infected with risk levels at least as large as the respective gathering size, $m \geq n$. In contrast, infections among susceptibles from other subpopulations do not contribute to new focal infections and thus they function similarly

to the recovered population. Overall, the transmission term is a weighted average over all gathering sizes according to a distribution, P_n, set by society (e.g., by lockdowns) [4]. The risk-SIR model includes transmission heterogeneity as subpopulations with higher risk level attend a larger fraction of gatherings, homophily as subpopulations with higher risk levels preferentially interact at larger gatherings, and super-spreading as multiple infections can occur at large gatherings.

The risk-SIR yields single-peaked epidemics like the classical SIR, but with underlying sequential peaks among decreasing risk level subpopulations (Figure 2b). Differences in growth rates cause these different peaks, namely that susceptibles in higher risk subpopulations attend larger gatherings on top of attending the same gatherings at the same rate as those in lower risk subpopulations. Similarly, infections in higher risk level subpopulations lead to more secondary infections. The basic reproduction number of an infected individual with a specified risk level, k, follows a recurrence relation:

$$\begin{aligned} R_0(k) &= R_0(k-1) + \mathbb{P}_k p_\beta \left(1 - \left(1 - \frac{1}{N}\right)^k\right) \\ &\approx R_0(k-1) + \mathbb{P}_k p_\beta \frac{k(k-1)}{N} \end{aligned}$$

such that a disproportionate number of secondary infections occur at larger gatherings as enumerated by k. The largest fraction of the secondary infections occur within the same risk subpopulation or in larger risk subpopulations as these susceptibles can attend all the same gathering as the infected, whereas susceptibles in smaller risk subpopulation do not attend the larger gatherings. Together, these dynamics cause the initial growth of the epidemic to be driven by the largest risk subpopulations, in line with prior work [25].

Impact on Surveillance

The sequential epidemic peaks within risk subpopulations have multiple impacts for disease surveillance. First, the cases initially accumulate within subpopulations at most risk, highlighting the importance of caution when interpreting and designing serosurveys to account for differential risk of exposure. Any underlying surveillance metrics that vary between subpopulations will now be dynamic. For example, we use infection fatality rates (IFR) to predict the threat of an emerging outbreak. If virulence depends on comorbidities that vary in frequency between each subpopulation then the IFR will change over the successive peaks of the underlying outbreaks in different subpopulations. Figure 2c illustrates IFR dynamics for two general cases. First, when individuals with comorbidities choose to avoid larger gatherings, i.e., subpopulation IFR negatively correlates with risk level (IFR low risk), then the IFR increases over time because at risk populations are infected relatively more frequently at later times. This would also reduce case ascertainment at the outset of an epidemic, potentially

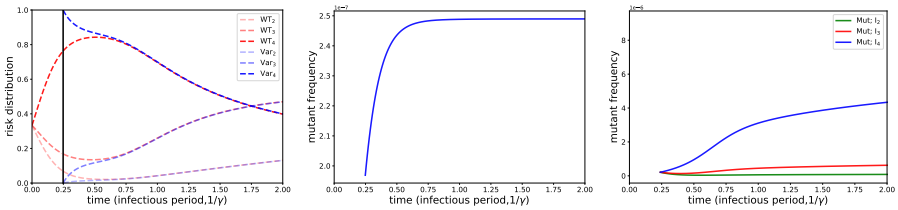

Figure 3: Variant frequency dynamics depend on which risk subpopulation it emerges within. A) Risk distribution of an epidemic (red) and a neutral variant (blue) that emerges within the highest risk subpopulation. B) The associated frequency dynamics of the emerging neutral variant. C) Frequency dynamics of a variant with 50% greater transmissibility relative to the background epidemic emerging in different risk subpopulations

biasing for lower estimates of R_0. Conversely, if individuals with comorbidities attend larger gatherings, i.e., subpopulation IFR positively correlates with risk level, then the IFR decreases over time (IFR high risk). This latter example is particular relevant for essential workers during lockdowns and communicable childhood diseases within school settings. IFR dynamics emphasize the danger in reporting aggregate statistics without accounting for disproportionate disease spread among subpopulations, such as when tracking emerging variants.

An epidemic grows at a rate determined by the socialization of all infected individuals, i.e., it's risk distribution. Meanwhile, a variant emerges within a subset of these infected individuals. Thus, even a neutral variant with no increased transmissibility relative to the background it emerges in is expected to increase in frequency if it is transmitted by individuals more social than average across the epidemic. Figure 3a shows the risk distribution of an epidemic and a neutral variant emerging within the most social subpopulation at a given time. The resulting variant frequency dynamics, shown in Figure 3b, mimic logistic growth–approaching an equilibrium frequency prior to sweeping. This dynamic emphasizes the role that social context plays in determining the observed growth rate of a variant. For example, a variant of concern with increased transmissibility relative to wild type will ultimately grow in frequency, but if it emerges within a less social subpopulation then the observed increase in frequency will be delayed (figure 3c) increasing its chance to stochastically become extinct. The same plot with a logarithmic y-axis shows a clear decreases in frequency of the more transmissible variants that emerge in less social subpopulations, which increases the chance for stochastic extinction (supplementary figure 1).

Since both increased transmissibility and increased contact opportunities contribute to the apparent fitness of a variant, identifying fast growing lineages during genomic surveillance imposes a selection bias towards identifying lineages preferentially spreading among more social individuals. To see this, we simulated the individual transmission and recovery events of a risk-SIR model using the Gillespie algorithm and tracked all possible lineages [12]. Figure 4a

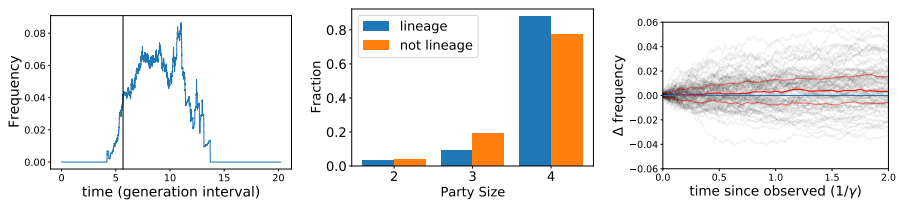

Figure 4: Selection bias of fast growing clades a) An example of a fast growing clade that continues to grow beyond the point it was identified (black vertical line) and its b) risk distribution among infected individuals as compared to all others infected at the moment of censusing. c) Frequency dynamics of the fastest growing, emergent clades following a shared censusing point (transparent gray) across 100 simulations. The median is shown in red and the upper and lower quartiles are shown in pink.

shows the frequency dynamics of the fastest growing lineage (having reached a given size at a given time - dotted vertical line). This lineage is neutral but appears to be growing more rapidly than its peers, and continues to increase in frequency. This occurs because the risk distribution of the variant at the point of sampling, shown in Figure 4b, contains relatively more social individuals relative to the rest of the epidemic and, in turn, an expected increase in frequency relative to the epidemic. Indeed, repeating this simulation shows that the fastest growing lineages are on average biased towards an increased growth after censusing, as shown in Figure 4c. This means a fast-growing lineage that initially spreads among more social individuals is expected to continue to appear more transmissible until its risk distribution relaxes to that of the entire epidemic.

Society modulated surveillance bias

Who interacts with whom in a society affects the magnitude of the above identified surveillance biases. For example, the amount of homophily, i.e., assortative mixing, within a community sets the time scale over which such lineages will continue to spuriously increase in frequency. Intuitively, homophily increases the probability a skewed risk distribution remains skewed as onward infections primarily occur within riskier subpopulation. As the risk-SIR model has a fixed social structure and homophily, we can better understand the relationship between homophily and skewed risk distribution retention by analyzing a simplified SIR type model that disentangles transmission heterogeneity and homophily:

$$\frac{dI_H}{dt} = S_H(\beta_{HH} I_H + \beta_{HL} I_L) - \gamma I_H$$
$$\frac{dI_L}{dt} = S_H(\beta_{HL} I_H + \beta_{LL} I_L) - \gamma I_L$$

where H denotes highly social individuals and L denotes less social individuals. This is a minimal model including transmission heterogeneity when the subpopulations differ in their basic reproduction number, i.e., $R_{0,H} = \frac{\beta_{HH}S_H+\beta_{LH}S_L}{\gamma} > \frac{\beta_{LL}S_L+\beta_{LH}S_H}{\gamma} = R_{0,L}$, while smaller values of cross-transmission, β_{HL}, correspond to societies with greater homophily. For this model, the risk distribution simplifies to the relative fraction of the more social subpopulation $f_H = \frac{I_H}{I_H+I_L}$. We can solve for the equilibrium risk distribution, f_H^*, early on in an epidemic during exponential growth by fixing the susceptibles (S_H and S_L) to constants and solving the respective dynamics, $\frac{df_H}{dt} = 0$. The equilibrium risk distribution depends on the structure of society, with increasing homophily reflected as a higher proportion of highly social individuals (see supplementary figure 2).

Randomness in individual transmission and recovery events (i.e., demographic stochasticity) causes the risk distribution to drift, even when at equilibrium. We can solve for the dynamics of the probability distribution over risk distributions as induced by demographic stochasticity via a so-called "master equation" parametrized by the average rates of the individual events [39]. Figure 5a shows probability distributions after 500 population changes (transmission or recovery) for epidemics varying in homophily and initiated with 20 infected individuals distributed according to their respective equilibrium risk distribution. Note, we vary homophily alone and keep the same levels of transmission heterogeneity across models by fixing the subpopulation basic reproduction numbers, $R_{0,H}$ and $R_{0,L}$. Demographic stochasticity induces a spread that eventually dissipates over time (supplementary figure 3). This means emerging variants are likely to deviate further from the equilibrium growth rate than the rest of the epidemic. This motivates quantifying how often a variant will grow faster versus slower than the equilibrium growth rate, $\rho = \frac{\mathbb{P}(f_H > f_H^*)}{\mathbb{P}(f_H > f_H^*)+\mathbb{P}(f_H < f_H^*)}$. Figure 5b compares the relative probability of increased sociality, ρ, dynamics after initiating infected populations at their respective equilibrium value, f_H^*, for different amounts of homophily. Emerging variants in societies with increased homophily have larger bias towards increased growth rates relative to the overall epidemic. Figure 5c shows the bias in risk distribution for epidemics initiating either in the more social (I_H, dashed lines) or the less social (I_L, solid lines) subpopulations. The subpopulation in which a variant emerges sets an initial risk distribution bias and societies with increased homophily retain this bias longer. However, the risk distribution bias ultimately approaches the same biased quasi-equilibrium value as seen above when started at the equilibrium.

Preferential stochastic extinction of slower growing epidemics exacerbates the growth bias induced by homophily. An epidemic goes stochastically extinct when its last remaining infected recovers prior to onwards transmission. Intuitively, epidemics are more likely to go stochastically extinct when the last remaining infected is among a less social subpopulation. This further increases the growth bias beyond that induced by homophily alone (see Supplementary Figure 4). Hence, we expect emerging variants to increase in frequency by preferentially spreading among more social individuals, particularly when they emerge in societies with high levels of homophily.

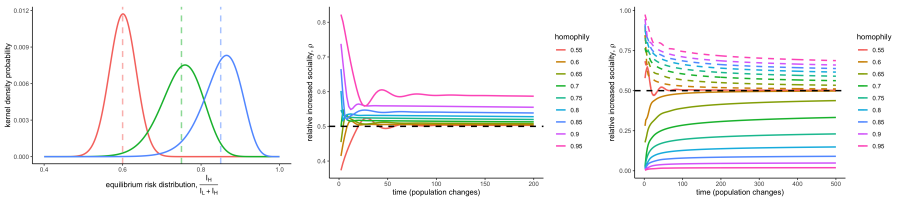

Figure 5: Homophily biases risk distributions a) Smoothed probability distributions of the risk distribution of epidemics initiated with 20 individuals at the respective equilibrium distribution after 500 population changes (total recovery or transmission events). Vertical dashed lines specify respective equilibrium risk distributions. B) Relative probability of a more social risk distribution ($f_H > f_H^*$) than a less social risk distribution ($f_H < f_H^*$) for epidemics initiated with 20 infected individuals distributed according to the equilibrium risk distribution (f_H^*) denoted as homophily. C) Relative probability of a more social risk distribution ($f_H > f_H^*$) than a less social risk distribution ($f_H < f_H^*$) for epidemics initiated with 1 infected individuals either in the more social subpopulation (dashed lines) or the less social subpopulation (solid lines).

Discussion

These models offer a simple and flexible approach to capture the independent role socialization plays in epidemic spread relative to the inherent epidemiologic properties of the pathogen. In particular, the distribution of gathering sizes determines rates of epidemic spread with the most social subpopulations disproportionately driving growth. Differences in degree of socialization, which leads to transmission heterogeneity and may vary over the course of an epidemic, bias surveillance in multiple ways that can influence early estimates of key epidemiologic parameters and the detection of emerging variants. Social structure modulates these with the amount of homophily setting the time scale and magnitude over which they change. While we analyze dynamics where socialization rates are static, we expect these to change in response to both exogenous factors like non-pharmaceutical interventions and endogenous factors like individual avoidance when cases are high. Indeed, it is all but certain that avoidance leads to a reduction in case counts prior to that expected from depletion of susceptibles alone as predicted from a classic SIR model, as evidenced by variant dynamics during the SARS-CoV-2 pandemic.

Exactly how long the apparent fitness bias persists depends on both epidemiological and social factors. We numerically showed bias remains for hundreds of infectious generations, which for a pathogen with a generation interval of 3-4 days and R0=3 (e.g., emerging SARS-CoV-2) gives retention times in the order of months [11]. This time frame is consistent with the incomplete, logistic growth of certain lineages that was observed in populations subjected to intense genomic surveillance [17], although we caution against over-interpreting as other mechanisms can lead to similar dynamics.

Repeat introductions during a pandemic complicate comparing variant dynamics between communities. Each introduction to a community can be interpreted as a single trajectory sampled from the probability distribution of risk distributions. Thus, the overall variant transmissibility within the community is a weighted average of each trajectory over time, as repeat introductions happen at different times. Furthermore, any variant fitness bias depends strongly on the subpopulation in which it emerges. Hence, if individuals willing to travel during a pandemic are more likely to engage in behavior that risks exposure in general then we expect introductions to preferentially occur amongst subpopulations willing to attend larger gatherings and, in turn, further contribute to an increased apparent fitness. In this case, we expect the largest fitness bias for imported variants within communities with little community transmission, i.e., those with recent introductions. However, it remains to be seen whether this social importation bias is significant compared to the inherent differences in epidemic growth rate between communities due to structural differences in socialization.

The sequential epidemic spread through less social subpopulations affects the emergence of future variants, particularly those capable of immune escape. For example, multiple variants emerged during the SARS-CoV-2 pandemic that were are able to transmit better among individuals with prior infections, as compared to variants well targeted by antibodies previously generated [5]. Our modeling predicts that more infections occur within more social subpopulations and that these infections tend to occur early in the epidemic. On one hand, if the earliest infections tend to happen among the most social, then immune escape increases the relative fitness of an emerging variant by effectively adding more susceptibles to the more social subpopulations. On the other hand, if antibody waning causes those previously infected to become susceptible again, then the most social individuals convert earlier due to prior infections. In that case, immune escape provides less of a fitness benefit, as recovered individuals are preferentially retained in less social populations that were previously infected at later times. Overall, the relative fitness benefit of immune escape depends on the distribution of recovered individuals among social subpopulations as determined by the rate of antibody waning.

Understanding the dynamics of socialization rates themselves is the next step to better model real epidemics and the risk-SIR model provides a framework for these future studies. However, identifying empirical proxies for socialization to inform these models is challenging due to the limited scope of measurements. Mobility data from tracked cell phones offer one possibility, with prior work showing lockdowns reducing socialization during the initial stages of the SARS-CoV-2 pandemic [43, 21]. However, such data ignore individuals without cell phones, which may include key social subpopulations like young schoolchildren. Additionally, even inferring interactions from mobility data alone may be fraught as it is agnostic to epidemiological details such as where and when masking may be in use, or spurious connections between phones when proximate but between barriers. One possible route to estimate varying contact patterns over time is methods from population genetics, leveraging the extraordinary

amounts of genetic data that are increasingly routinely collected [34].

This work, along with others, illustrates how simple models can capture the way social context determines and complicates epidemic dynamics, without explicit consideration of host networks [42]. To begin to quantify emergent complex parameters such as the reproduction number R_0 can be compared among relevant contexts. Prior work has shown the value of these key parameters depends on social context through the distribution of secondary contacts [28, 16]. Nevertheless, secondary infection rates and case count growth rates are easily measurable and can be incorporated into this modeling framework, allowing us to contextualize discussions within specific social contexts instead of a single mean value which can understate the capacity for transmission in certain populations. Understanding social contexts associated with rapid growth may allow early warning of for future spread in different, neighboring social contexts [31], and will be a key component of improved epidemic forecasting that accounts for changing behavior. Determining these social "routes of transmission" could be key for broad infection control, where resources can be prioritized towards mitigating epidemics within and between communities with increased growth.

Materials & Methods

Transmissibility of mutant lineages

A mutant lineage emerges in a single infected individual with the difference in transmissibility and recovery rate setting the initial growth rate. Within a neutral model, Only tranmissibility differs between risky subpopulation where we can derive the difference in growth rate by comparing sampling process for parties. Focusing on infected inidividuals between adjacent risk groups (i.e., n+1 and n). Note, individuals from each focal group have the same probability of attending parties hosted by individuals with risk levels up to the lower risk level. Hence, differences of transmissibility arise in three cases: when either focal group hosts but the other focal group does not attend and when focal individuals attend party invitations from either riskier members or the focal mutant lineages themselves.

The transmissbility arises from the expected number of infections due to a focal lineage at a party. We must account for the probability of a specific party composition, and then the expected number of infections attributable to the focal lineage at the party and not other infectors. We attribute infections to a single, random infector when a susceptible is multiply infected. The expected

number of infections depends on which member of the riskier population hosts:

$$\mathbb{I}_S(n+1) = \frac{n}{N_{\geq n+1}} \frac{S_{n+1}}{N} \sum_{s,i,r}^{n-1} \text{Mult}(s,i,r,n-1)(s+1)\overline{p_\beta}(i),$$

$$\mathbb{I}_I(n+1) = \frac{n}{N_{\geq n+1}} \frac{I_{n+1}}{N} \sum_{s,i,r}^{n-1} \text{Mult}(s,i,r,n-1)s\overline{p_\beta}(i+1),$$

$$\mathbb{I}_R(n+1) = \frac{n}{N_{\geq n+1}} \frac{R_{n+1}}{N} \sum_{s,i,r}^{n-1} \text{Mult}(s,i,r,n-1)s\overline{p_\beta}(i),$$

$$\overline{p_\beta}(i) = p_\beta \sum_{k=0}^{i} \left(\frac{1}{1+k}\right) \binom{i}{k} p_\beta^k (1-p_\beta)^{i-k} = \frac{1-(1-p_\beta)^{i+1}}{i+1}$$

where $\mathbb{I}_X(n+1)$ is the expected number of focal infections when a X_{n+1} hosts. The first term on the right hand side is the probability that the focal lineage joins the party, $\lim_{N\to\infty}\left(1-\left(1-\frac{1}{N_{\geq n+1}}\right)^n\right) = \frac{n}{N_{\geq n+1}}$. The contribution of each party is weighted by its multinomial probability of sampling $n-1$ individuals since the host and the focal infected are already accounted for. The last term reflects the binomial number of attending susceptibles infected by the focal lineage, where we also properly attribute the source infection in the case of multiple infectors on a single susceptible. This rescales the focal lineage probability of infection by the number other attending infectors, $p_\beta \to \overline{p_\beta}(i)$. Note, we perform sampling with replacement which is accurate as $N \to \infty$.

There is also a difference in transmissibilty between risk groups due to different numbers of attendees at parties hosted by each focal lineage. The exact terms follow similar sampling to that above:

$$\mathbb{I}_F(n+1) = \frac{1}{N} \sum_{s,i,r}^{n} \text{Mult}(s,i,r,n)s\overline{p_\beta}(i)$$

where F refers to focal. Note, we ignore focal lineages attending parties hosted by other focal lineages of other risk groups, since these terms disappear as $N \to \infty$. The full difference in transmissibility follows:

$$B(n+1) - B(n) =$$
$$\mathbb{I}_S(n+1) + \mathbb{I}_I(n+1) + \mathbb{I}_R(n+1) + \mathbb{I}_F(n+1) - \mathbb{I}_F(n)$$

Acknowledgements

This project has been funded (in part) by contract 200-2016-91779 with the Centers for Disease Control and Prevention. Disclaimer: The findings, conclusions, and views expressed are those of the author(s) and do not necessarily represent the official position of the Centers for Disease Control and Prevention (CDC).

References

[1] Sonia Altizer, Charles L. Nunn, Peter H. Thrall, John L. Gittleman, Janis Antonovics, Andrew A. Cunningham, Andrew P. Dobson, Vanessa Ezenwa, Kate E. Jones, Amy B. Pedersen, Mary Poss, and Juliet R.C. Pulliam. Social Organization and Parasite Risk in Mammals: Integrating Theory and Empirical Studies. *Annual Review of Ecology, Evolution, and Systematics*, 34(1):517–547, 11 2003.

[2] Ronan F. Arthur, James H. Jones, Matthew H. Bonds, Yoav Ram, and Marcus W. Feldman. Adaptive social contact rates induce complex dynamics during epidemics. *PLOS Computational Biology*, 17(2):e1008639, feb 10 2021.

[3] Shweta Bansal, Jonathan Read, Babak Pourbohloul, and Lauren Ancel Meyers. The dynamic nature of contact networks in infectious disease epidemiology. *Journal of Biological Dynamics*, 4(5):478–489, 9 2010.

[4] Christopher B. Boyer, Eva Rumpler, Stephen M. Kissler, and Marc Lipsitch. Infectious disease dynamics and restrictions on social gathering size. *Epidemics*, 40:100620, 9 2022.

[5] Mary Bushman, Rebecca Kahn, Bradford P. Taylor, Marc Lipsitch, and William P. Hanage. Population impact of SARS-CoV-2 variants with enhanced transmissibility and/or partial immune escape. *Cell*, 184(26):6229–6242.e18, 12 2021.

[6] Aroon Chande, Seolha Lee, Mallory Harris, Quan Nguyen, Stephen J. Beckett, Troy Hilley, Clio Andris, and Joshua S. Weitz. Real-time, interactive website for US-county-level COVID-19 event risk assessment. *Nature Human Behaviour*, 4(12):1313–1319, nov 9 2020.

[7] David A. Edwards, Dennis Ausiello, Jonathan Salzman, Tom Devlin, Robert Langer, Brandon J. Beddingfield, Alyssa C. Fears, Lara A. Doyle-Meyers, Rachel K. Redmann, Stephanie Z. Killeen, Nicholas J. Maness, and Chad J. Roy. Exhaled aerosol increases with COVID-19 infection, age, and obesity. *Proceedings of the National Academy of Sciences*, 118(8), feb 9 2021.

[8] Ceyhun Eksin, Jeff S. Shamma, and Joshua S. Weitz. Disease dynamics in a stochastic network game: a little empathy goes a long way in averting outbreaks. *Scientific Reports*, 7(1), mar 14 2017.

[9] Akira Endo, Hiroaki Murayama, Sam Abbott, Ruwan Ratnayake, Carl A. B. Pearson, W. John Edmunds, Elizabeth Fearon, and Sebastian Funk. Heavy-tailed sexual contact networks and monkeypox epidemiology in the global outbreak, 2022. *Science*, 378(6615):90–94, oct 7 2022.

[10] Christophe Fraser, T. Déirdre Hollingsworth, Ruth Chapman, Frank de Wolf, and William P. Hanage. Variation in HIV-1 set-point viral load: Epidemiological analysis and an evolutionary hypothesis. *Proceedings of the National Academy of Sciences*, 104(44):17441–17446, oct 30 2007.

[11] Tapiwa Ganyani, Cécile Kremer, Dongxuan Chen, Andrea Torneri, Christel Faes, Jacco Wallinga, and Niel Hens. Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. *Eurosurveillance*, 25(17), apr 30 2020.

[12] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 12 1977.

[13] Stephanie S. Godfrey, C. Michael Bull, Richard James, and Kris Murray. Network structure and parasite transmission in a group living lizard, the gidgee skink, Egernia stokesii. *Behavioral Ecology and Sociobiology*, 63(7):1045–1056, apr 2 2009.

[14] Smita Gopinath, Sarah Carden, and Denise Monack. Shedding light on Salmonella carriers. *Trends in Microbiology*, 20(7):320–327, 7 2012.

[15] Mallory J. Harris, Kimberly J. Cardenas, and Erin A. Mordecai. Social divisions and risk perception drive divergent epidemics and large later waves. *Evolutionary Human Sciences*, 5, 2023.

[16] Laurent Hébert-Dufresne, Benjamin M. Althouse, Samuel V. Scarpino, and Antoine Allard. BeyondR_0: heterogeneity in secondary infections and probabilistic epidemic forecasting. *Journal of The Royal Society Interface*, 17(172):20200393, 11 2020.

[17] Emma B. Hodcroft, Moira Zuber, Sarah Nadeau, Timothy G. Vaughan, Katharine H. D. Crawford, Christian L. Althaus, Martina L. Reichmuth, John E. Bowen, Alexandra C. Walls, Davide Corti, Jesse D. Bloom, David Veesler, David Mateo, Alberto Hernando, Iñaki Comas, Fernando González-Candelas, Fernando González-Candelas, Galo Adrian Goig, Álvaro Chiner-Oms, Irving Cancino-Muñoz, Mariana Gabriela López, Manuela Torres-Puente, Inmaculada Gomez-Navarro, Santiago Jiménez-Serrano, Lidia Ruiz-Roldán, María Alma Bracho, Neris García-González, Llúcia Martínez-Priego, Inmaculada Galán-Vendrell, Paula Ruiz-Hueso, Griselda De Marco, Maria Loreto Ferrús, Sandra Carbó-Ramírez, Giuseppe D'Auria, Mireia Coscollá, Paula Ruiz-Rodríguez, Francisco Javier Roig-Sena, Isabel Sanmartín, Daniel Garcia-Souto, Ana Pequeno-Valtierra, Jose M. C. Tubio, Jorge Rodríguez-Castro, Nuria Rabella, Ferrán Navarro, Elisenda Miró, Manuel Rodríguez-Iglesias, Fátima Galán-Sanchez, Salud Rodriguez-Pallares, María de Toro, María Bea Escudero, José Manuel Azcona-Gutiérrez, Miriam Blasco Alberdi, Alfredo Mayor, Alberto L. García-Basteiro, Gemma Moncunill, Carlota Dobaño, Pau Cisteró, Darío García-de Viedma, Laura Pérez-Lago, Marta Herranz, Jon Sicilia, Pilar Catalán-Alonso, Patricia Muñoz, Cristina Muñoz-Cuevas, Guadalupe

Rodríguez-Rodríguez, Juan Alberola-Enguidanos, Jose Miguel Nogueira, Juan José Camarena, Antonio Rezusta, Alexander Tristancho-Baró, Ana Milagro, Nieves Felisa Martínez-Cameo, Yolanda Gracia-Grataloup, Elisa Martró, Antoni E. Bordoy, Anna Not, Adrián Antuori-Torres, Rafael Benito, Sonia Algarate, Jessica Bueno, Jose Luis del Pozo, Jose Antonio Boga, Cristián Castelló-Abietar, Susana Rojo-Alba, Marta Elena Alvarez-Argüelles, Santiago Melon, Maitane Aranzamendi-Zaldumbide, Andrea Vergara-Gómez, Jovita Fernández-Pinero, Miguel J. Martínez, Jordi Vila, Elisa Rubio, Aida Peiró-Mestres, Jessica Navero-Castillejos, David Posada, Diana Valverde, Nuria Estévez-Gómez, Iria Fernandez-Silva, Loretta de Chiara, Pilar Gallego-García, Nair Varela, Rosario Moreno, Maria Dolores Tirado, Ulises Gomez-Pinedo, Mónica Gozalo-Margüello, Maria Eliecer-Cano, José Manuel Méndez-Legaza, Jesus Rodríguez-Lozano, María Siller, Daniel Pablo-Marcos, Antonio Oliver, Jordi Reina, Carla López-Causapé, Andrés Canut-Blasco, Silvia Hernáez-Crespo, Maria Luz A. Cordón, María-Concepción Lecároz-Agara, Carmen Gómez-González, Amaia Aguirre-Quiñonero, José Israel López-Mirones, Marina Fernández-Torres, Maria Rosario Almela-Ferrer, Nieves Gonzalo-Jiménez, Maria Montserrat Ruiz-García, Antonio Galiana, Judith Sanchez-Almendro, Gustavo Cilla, Milagrosa Montes, Luis Piñeiro, Ane Sorarrain, José María Marimón, Maria Dolores Gomez-Ruiz, José Luis López-Hontangas, Eva M. González Barberá, José María Navarro-Marí, Irene Pedrosa-Corral, Sara Sanbonmatsu-Gámez, Carmen Pérez-González, Francisco Chamizo-López, Ana Bordes-Benítez, David Navarro, Eliseo Albert, Ignacio Torres, Isabel Gascón, Cristina Juana Torregrosa-Hetland, Eva Pastor-Boix, Paloma Cascales-Ramos, Begoña Fuster-Escrivá, Concepción Gimeno-Cardona, María Dolores Ocete, Rafael Medina-Gonzalez, Julia González-Cantó, Olalla Martínez-Macias, Begoña Palop-Borrás, Inmaculada de Toro, Maria Concepción Mediavilla-Gradolph, Mercedes Pérez-Ruiz, Óscar González-Recio, Mónica Gutiérrez-Rivas, Encarnación Simarro-Córdoba, Julia Lozano-Serra, Lorena Robles-Fonseca, Adolfo de Salazar, Laura Viñuela-González, Natalia Chueca, Federico García, Cristina Gómez-Camarasa, Ana Carvajal, Raul de la Puente, Vicente Martín-Sánchez, Juan-Miguel Fregeneda-Grandes, Antonio José Molina, Héctor Argüello, Tania Fernández-Villa, Maria Amparo Farga-Martí, Victoria Domínguez-Márquez, José Javier Costa-Alcalde, Rocío Trastoy, Gema Barbeito-Castiñeiras, Amparo Coira, María Luisa Pérez-del Molino, Antonio Aguilera, Anna M. Planas, Alex Soriano, Israel Fernandez-Cádenas, Jordi Pérez-Tur, Maria Ángeles Marcos, Antonio Moreno-Docón, Esther Viedma, Jesús Mingorance, Juan Carlos Galán-Montemayor, Mónica Parra-Grande, Tanja Stadler, and Richard A. Neher. Spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *Nature*, 595(7869):707–712, jun 7 2021.

[18] Matt Keeling. The implications of network structure for epidemic dynamics. *Theoretical Population Biology*, 67(1):1–8, 2 2005.

[19] Matt J Keeling and Ken T.D Eames. Networks and epidemic models. *Journal of The Royal Society Interface*, 2(4):295–307, jun 20 2005.

[20] W. O. Kermack and A. G McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115(772):700–721, 8 1927.

[21] Nishant Kishore, Rebecca Kahn, Pamela P. Martinez, Pablo M. De Salazar, Ayesha S. Mahmud, and Caroline O. Buckee. Lockdowns result in changes in human mobility which may impact the epidemiologic dynamics of SARS-CoV-2. *Scientific Reports*, 11(1), mar 26 2021.

[22] Trevor D. Lawley, Donna M. Bouley, Yana E. Hoy, Christine Gerke, David A. Relman, and Denise M. Monack. Host Transmission of *Salmonella enterica* Serovar Typhimurium Is Controlled by Virulence Factors and Indigenous Intestinal Microbiota. *Infection and Immunity*, 76(1):403–416, 1 2008.

[23] James O Lloyd-Smith, Sebastian J Schreiber, P Ekkehard Kopp, and Wayne M Getz. Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066):355–359, 2005.

[24] L. MATTHEWS, I. J. McKENDRICK, H. TERNENT, G. J. GUNN, B. SYNGE, and M. E. J. WOOLHOUSE. Super-shedding cattle and the transmission dynamics of*Escherichia coli*O157. *Epidemiology and Infection*, 134(1):131–142, jun 3 2005.

[25] Robert M. May and Roy M. Anderson. Transmission dynamics of HIV infection. *Nature*, 326(6109):137–142, mar 1 1987.

[26] Ernst Mayr. *Systematics and the Origin of Species*. New York: Columbia University Press., 1942.

[27] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27(1):415–444, 8 2001.

[28] Lauren Ancel Meyers, Babak Pourbohloul, M.E.J. Newman, Danuta M. Skowronski, and Robert C. Brunham. Network theory and SARS: predicting outbreak diversity. *Journal of Theoretical Biology*, 232(1):71–81, 1 2005.

[29] M. E. J. Newman. Assortative Mixing in Networks. *Physical Review Letters*, 89(20), oct 28 2002.

[30] Nicola Perra, Duygu Balcan, Bruno Gonçalves, and Alessandro Vespignani. Towards a Characterization of Behavior-Disease Models. *PLoS ONE*, 6(8):e23084, aug 3 2011.

[31] Brittany A Petros, Jacquelyn Turcinovic, Nicole L Welch, Laura F White, Eric D Kolaczyk, Matthew R Bauer, Michael Cleary, Sabrina T Dobbins, Lynn Doucette-Stamm, Mitch Gore, Parvathy Nair, Tien G Nguyen, Scott Rose, Bradford P Taylor, Daniel Tsang, Erik Wendlandt, Michele Hope, Judy T Platt, Karen R Jacobson, Tara Bouton, Seyho Yune, Jared R Auclair, Lena Landaverde, Catherine M Klapperich, Davidson H Hamer, William P Hanage, Bronwyn L MacInnis, Pardis C Sabeti, John H Connor, and Michael Springer. Early Introduction and Rise of the Omicron Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Variant in Highly Vaccinated University Populations. *Clinical Infectious Diseases*, 76(3):e400–e408, jul 1 2022.

[32] Thomas C. Quinn, Maria J. Wawer, Nelson Sewankambo, David Serwadda, Chuanjun Li, Fred Wabwire-Mangen, Mary O. Meehan, Thomas Lutalo, and Ronald H. Gray. Viral Load and Heterosexual Transmission of Human Immunodeficiency Virus Type 1. *New England Journal of Medicine*, 342(13):921–929, mar 30 2000.

[33] Pejman Rohani, Xue Zhong, and Aaron A. King. Contact Network Structure Explains the Changing Epidemiology of Pertussis. *Science*, 330(6006):982–985, nov 12 2010.

[34] Samuel V. Scarpino, Atila Iamarino, Chad Wells, Dan Yamin, Martial Ndeffo-Mbah, Natasha S. Wenzel, Spencer J. Fox, Tolbert Nyenswah, Frederick L. Altice, Alison P. Galvani, Lauren Ancel Meyers, and Jeffrey P. Townsend. Epidemiological and Viral Genomic Sequence Analysis of the 2014 Ebola Outbreak Reveals Clustered Transmission. *Clinical Infectious Diseases*, 60(7):1079–1082, 2015.

[35] Richard A. Stein. Super-spreaders in infectious diseases. *International Journal of Infectious Diseases*, 15(8):e510–e513, 8 2011.

[36] Hein Stigum, W. Falck, and P. Magnus. The core group revisited: The effect of partner mixing and migration on the spread of gonorrhea, chlamydia, and HIV. *Mathematical Biosciences*, 120(1):1–23, 3 1994.

[37] Charles A. Taylor, Christopher Boulos, and Douglas Almond. Livestock plants and COVID-19 transmission. *Proceedings of the National Academy of Sciences*, 117(50):31706–31715, nov 19 2020.

[38] Alan R. Templeton. The reality and importance of founder speciation in evolution. *BioEssays*, 30(5):470–479, 2008.

[39] N.G. van Kampen. *Stochastic Processes in Physics and Chemistry (Third Edition)*. Elsevier, 2007.

[40] Erik Volz and Lauren Ancel Meyers. Susceptible–infected–recovered epidemics in dynamic contact networks. *Proceedings of the Royal Society B: Biological Sciences*, 274(1628):2925–2934, sep 18 2007.

[41] Erik M. Volz, Joel C. Miller, Alison Galvani, and Lauren Ancel Meyers. Effects of Heterogeneous and Clustered Contact Patterns on Infectious Disease Dynamics. *PLoS Computational Biology*, 7(6):e1002042, jun 2 2011.

[42] Joshua S. Weitz, Sang Woo Park, Ceyhun Eksin, and Jonathan Dushoff. Awareness-driven behavior changes can shift the shape of epidemics away from peaks and toward plateaus, shoulders, and oscillations. *Proceedings of the National Academy of Sciences*, 117(51):32764–32771, 12 2020.

[43] Amy Wesolowski, Taimur Qureshi, Maciej F. Boni, Pål Roe Sundsøy, Michael A. Johansson, Syed Basit Rasheed, Kenth Engø-Monsen, and Caroline O. Buckee. Impact of human mobility on the emergence of dengue epidemics in Pakistan. *Proceedings of the National Academy of Sciences*, 112(38):11887–11892, sep 8 2015.

[44] Mark EJ Woolhouse, C Dye, J-F Etard, T Smith, JD Charlwood, GP Garnett, P Hagan, JLK Hii, PD Ndhlovu, RJ Quinnell, et al. Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proceedings of the National Academy of Sciences*, 94(1):338–342, 1997.

[45] JAMES A. YORKE, HERBERT W. HETHCOTE, and ANNETT NOLD. Dynamics and Control of the Transmission of Gonorrhea. *Sexually Transmitted Diseases*, 5(2):51–56, 4 1978.