Chopping the tail: how preventing superspreading can help to maintain COVID-19 control

Morgan P. Kain[1,2]*, Marissa L. Childs[3]*, Alexander D. Becker[1], Erin A. Mordecai[1]

June 30, 2020

*Denotes equal authorship. Corresponding authors: morganpkain@gmail.com, marissac@stanford.edu
[1]Department of Biology, Stanford University, Stanford, CA, 94305, USA
[2]Natural Capital Project, Woods Institute for the Environment, Stanford University, Stanford, CA 94305, USA
[3]Emmett Interdisciplinary Program in Environment and Resources, Stanford University, Stanford, CA, 94305, USA

Abstract

Disease transmission is notoriously heterogeneous, and SARS-CoV-2 is no exception. A skewed distribution where few individuals or events are responsible for the majority of transmission can result in explosive, superspreading events, which produce rapid and volatile epidemic dynamics, especially early or late in epidemics. Anticipating and preventing superspreading events can produce large reductions in overall transmission rates. Here, we present a compartmental (SEIR) epidemiological model framework for estimating transmission parameters from multiple imperfectly observed data streams, including reported cases, deaths, and mobile phone-based mobility that incorporates individual-level heterogeneity in transmission using previous estimates for SARS-CoV-1 and SARS-CoV-2. We parameterize the model for COVID-19 epidemic dynamics by estimating a time-varying transmission rate that incorporates the impact of non-pharmaceutical intervention strategies that change over time, in five epidemiologically distinct settings—Los Angeles and Santa Clara Counties, California; Seattle (King County), Washington; Atlanta (Dekalb and Fulton Counties), Georgia; and Miami (Miami-Dade County), Florida. We find the effective reproduction number $\mathcal{R}_E$ dropped below 1 rapidly following social distancing orders in mid-March, 2020 and remained there into June in Santa Clara County and Seattle, but climbed above 1 in late May in Los Angeles, Miami, and Atlanta, and has trended upward in all locations since April. With the fitted model, we ask: how does truncating the tail of the individual-level transmission rate distribution affect epidemic dynamics and control? We find interventions that truncate the transmission rate distribution while partially relaxing social distancing are broadly effective, with impacts on epidemic growth on par with the strongest population-wide social distancing observed in April, 2020. Given that social distancing interventions will be needed to maintain epidemic control until a vaccine becomes widely available, "chopping off the tail" to reduce the probability of superspreading events presents a promising option to alleviate the need for extreme general social distancing.

Introduction

In the face of emerging epidemics with limited pharmaceutical options for treatment and prevention of infection, non-pharmaceutical interventions such as social distancing are critical for slowing epidemic growth. Shelter-in-place and other social distancing orders have helped to slow the pace of the COVID-19 pandemic, reducing the effective reproduction number $\mathcal{R}_E$—or the number of secondary infections produced by each infected person—to one or below in most places. In doing so, social distancing has effectively kept most regional healthcare systems operating under maximum capacity. However, after only a few weeks of declining numbers of daily cases due to an $\mathcal{R}_E$ at or below one, most state and county governments in the United States have begun relaxing social distancing orders, citing their major economic impacts. In order to avoid epidemic resurgence, it is vitally important that governments employ long-term strategies that maintain epidemic control as economic reopening commences.

One obstacle to designing effective long-term strategies is a notoriously heterogeneous transmission process. It is now widely recognized that the minority of infections generate the majority of secondary cases, leading to the so-called 20-80 rule in epidemiology (the rule-of-thumb that 20% of people generate 80% of cases)[1]. Work on SARS-CoV-1, measles, and other respiratory viruses suggests that this skew in secondary cases is even larger[2]. This heterogeneity gives rise to events in which a single infected person transmits a disease to dozens or hundreds of people—called superspreading events—which have played an important role in the COVID-19 pandemic[3,4,5,6,7]. Indeed, the frequency of asymptomatic and presymptomatic transmission, potential disconnect between infection and clinical presentation[8], and potential transmission via direct contact, aerosols, and surfaces[9,10] are all features of SARS-CoV-2 that tend to promote superspreading. As local and national governments search for viable exit strategies from shelter-in-place, a critical question is how effective curtailing superspreading events could be in controlling epidemic spread.

Practically, one strategy to help prevent superspreading is to prohibit medium to large indoor gatherings such as exercise classes, sporting events, concerts, and weddings for an extended period after allowing smaller and lower-risk activities to resume. From a modeling standpoint, predicting the effects of this straightforward intervention is difficult for two reasons: 1) local epidemiological dynamics are changing with evolving intervention strategies; and 2) information may not be available to parameterize detailed models of disease spread through heterogeneous populations. Despite these difficulties, it is important to consider some individual-level heterogeneity in transmission because model analyses of mean transmission rates alone may over-estimate the effectiveness of interventions, overlook potentially effective interventions that act on the heterogeneity within populations, overlook potentially explosive resurgences, and poorly predict the final epidemic size[2,7].

Studies of superspreading often empirically estimate secondary case distributions from recorded transmission chains and/or using branching process models[2,5,6,11]. These studies estimate a dispersion parameter, k, that describes the variance in secondary cases based on a Negative Binomial distribution, where smaller values indicate more heterogeneity and skew and large values approach a Poisson distribution. Estimated k values for SARS-CoV-2 remain uncertain, but are thought to range from 0.04 - 0.3[6,7,11,12], similar to the estimate of 0.16 for SARS-CoV-1[2], which we use for this analysis. These empirical and branching process approaches are ideal for characterizing heterogeneity in secondary cases, but not for projecting epidemic trajectories through time, without being further embedded in a compartmental or network modeling framework.

Here, we present a mechanistic susceptible, exposed, infectious, removed (SEIR) model that uses data on cases, deaths, and mobility for parameter estimation, incorporates heterogeneity in transmission rates, and is realistic enough to be useful for scenario exploration but simple enough to be adapted to a wide range of settings. The key innovation in our model is in using the average of Gamma-distributed *individual* transmission rates at each time step, as supported by previous work on secondary case distributions, to generate the distribution of *population-average* transmission rates. This formulation allows us to both generate more realistic variation in trajectories than SEIR models that assume a single average transmission rate, and explore and quantify the impact of altering individual-level transmission distributions on population-level dynamics without more detailed information on contact networks, age structure, or other social information.

The model, with accompanying open-access code, can be used to fit to any county in the U.S. using publicly available data; here we focus on five contrasting epidemiological settings—Seattle (King County), Washington; Los Angeles (Los Angeles County), California; Santa Clara County, California; Atlanta (Dekalb and Fulton Counties), Georgia; and Miami (Miami-Dade County), Florida. For each location we estimate a time-varying effective reproduction number, $\mathcal{R}_E$, which represents the average number of secondary infections produced by each infected person, and is an important (though imperfect[7]) metric of epidemic control. Using each fitted model, we truncate the individual-level transmission rate distribution and stochastically simulate epidemic dynamics into the future, representing a scenario where high-risk events are eliminated but smaller and lower-risk activities are allowed to resume. We investigate the absolute impact of this superspreading prevention strategy on epidemic control, and compare its impact on epidemic dynamics (and $\mathcal{R}_E$) to test-and-isolate and shelter-in-place interventions. Using this comparison we highlight exit strategies from shelter-in-place that are expected to reduce both epidemic growth (i.e., keep $\mathcal{R}_E$ below one) and the probability of explosive resurgence.

Methods

Model Structure

We developed a compartmental model using an SEIR (Susceptible, Exposed, Infectious, Recovered) framework to model COVID-19 transmission, which was first described in Childs et al.[13]. Our model divides the population into the following classes: susceptible (S); exposed but not yet infectious (E); infectious and presymptomatic (I_P), asymptomatic (I_A), mildly symptomatic (I_M), or severely symptomatic (I_S); hospitalized cases that will recover (H_R) or die (H_D); recovered and immune (R); and dead (D). We assume an underlying, unobserved process model of SARS-CoV-2 transmission described by equations 1–10 and shown in Figure S1, where each term $d_{X,Y}$ denotes the transition from compartment X to Y. Transitions between compartments are simulated as binomial ($\mathcal{B}$) or multinomial ($\mathcal{M}$) processes. We use an Euler approximation of the continuous time process with a time step of 4 hours. To produce more realistic latent and infectious periods we divide each infectious class and the exposed period into multiple sub-stages, which results in Erlang distributed periods within stages[14,15]. Specifically, we use three sub-stages for the exposed class, seven sub-stages for the asymptomatic infectious class, two sub-stages for the presymptomatic infectious class, five sub-stages for the mildly symptomatic infectious class, and five sub-stages for the severely symptomatic class. We translate durations into rates for our model with sub-classes and a Euler approximation using the method described in He et al.[16]. Equations 11–18 describe in detail the stochastic rates used to approximate the transition terms in equations 1–10. Parameters are defined in Tables 1, 2, and 3.

$$\frac{dS}{dt} = -d_{S,E} \tag{1}$$
$$\frac{dE}{dt} = d_{S,E} - d_{E,I_A} - d_{E,I_P} \tag{2}$$
$$\frac{dI_A}{dt} = d_{E,I_A} - d_{I_A,R} \tag{3}$$
$$\frac{dI_P}{dt} = d_{E,I_P} - d_{I_P,I_S} - d_{I_P,I_M} \tag{4}$$
$$\frac{dI_M}{dt} = d_{I_P,I_M} - d_{I_M,R} \tag{5}$$
$$\frac{dI_S}{dt} = d_{I_P,I_S} - d_{I_S,H_R} - d_{I_S,H_D} \tag{6}$$
$$\frac{dH_R}{dt} = d_{I_S,H_R} - d_{H_R,R} \tag{7}$$
$$\frac{dH_D}{dt} = d_{I_S,H_D} - d_{H_D,D} \tag{8}$$
$$\frac{dR}{dt} = d_{I_A,R} + d_{I_M,R} + d_{H_R,R} \tag{9}$$
$$\frac{dD}{dt} = d_{H_D,D} \tag{10}$$

By including asymptomatic and presymptomatic individuals, we are able to track "silent spreaders" of the disease, which have been shown to contribute to COVID-19 transmission[17,18]. Mildly symptomatic cases are defined as those people that show symptoms but do not require hospitalization. We assume that all severely symptomatic cases will eventually require hospitalization and that no onward transmission occurs from hospitalized individuals.

$$d_{S,E} \sim \mathcal{B}\left(S, 1 - \exp\left(-\beta_t \frac{\kappa_A I_A + \kappa_P I_P + \kappa_M I_M + \kappa_S I_S}{N} dt\right)\right) \tag{11}$$

$$\begin{pmatrix} d_{E,E} \\ d_{E,I_A} \\ d_{E,I_P} \end{pmatrix} \sim \mathcal{M}\left(E, \begin{pmatrix} \exp(-\gamma dt) \\ \alpha(1-\exp(-\gamma dt)) \\ (1-\alpha)(1-\exp(-\gamma dt)) \end{pmatrix}\right) \tag{12}$$

$$d_{I_A,R} \sim \mathcal{B}(I_A, 1 - \exp(-\lambda_A dt)) \tag{13}$$

$$\begin{pmatrix} d_{I_P,I_P} \\ d_{I_P,I_M} \\ d_{I_P,I_S} \end{pmatrix} \sim \mathcal{M}\left(I_P, \begin{pmatrix} \exp(-\lambda_P dt) \\ \mu(1-\exp(-\lambda_P dt)) \\ (1-\mu)(1-\exp(-\lambda_P dt)) \end{pmatrix}\right) \tag{14}$$

$$d_{I_M,R} \sim \mathcal{B}(I_M, 1 - \exp(-\lambda_M dt)) \tag{15}$$

$$\begin{pmatrix} d_{I_S,I_S} \\ d_{I_S,H_R} \\ d_{I_S,H_D} \end{pmatrix} = \mathcal{M}\left(I_S, \begin{pmatrix} \exp(-\lambda_S dt) \\ \delta(1-\exp(-\lambda_S dt)) \\ (1-\delta)(1-\exp(-\lambda_S dt)) \end{pmatrix}\right) \tag{16}$$

$$d_{H_R,R} \sim \mathcal{B}(H_R, 1 - \exp(-\rho_R dt)) \tag{17}$$

$$d_{H_D,D} \sim \mathcal{B}(H_D, 1 - \exp(-\rho_D dt)) \tag{18}$$

Table 1: Parameter point estimates.

Parameter	*Value*	*Description*	*Estimates and Sources*
$\kappa_P, \kappa_M, \kappa_S$	1	Relative infectiousness of presymptomatic, mild symptomatic, and severe symptomatic	Assumed
γ	3.5 days	Preinfectious period	One meta-analysis[19] found a mean incubation period of 5.8 days, another found a median of 5.1 days[20]. We use a shortened duration because we assume 2 days of presymptomatic transmission
λ_P	2 days	Presymptomatic duration	Range of 1-3 days[21], mean of 3.8 days[22], viral shedding estimated to begin 2.3 days prior to symptom onset[23], many articles find presymptomatic infection is likely but do not estimate duration[24,25,26,27]
λ_A	7 days	Infectious period for asymptomatic infections	Mean seroconversion after 7 days[28]
λ_S	5 days	Time from symptom onset to hospitalizations (severe cases)	Mean of 5.5 days[29], median of 4 days[30], mean of 5.6 days[31]
λ_M	5.0 days	Time from symptom onset to recovery (mild cases)	Infectiousness based on viral shedding estimated to decline substantially within 7 days[23,28] *Note,[23] takes samples from hospitalized patients; we assume similar viral shedding in mild infections
ρ_R	13.3 days	Time from hospitalization to recovery	Mean of 13.3[32], mean of 20.51[31], highly variable by region[33]

Table 2: Parameter range estimates that are not location specific

Parameter	*Lower Bound*	*Upper Bound*	*Description*	*Estimates and Sources*
κ_S	0.4	0.8	Relative infectiousness of asymptomatic infections	0.6[34], few direct estimates, but many examples of asymptomatic transmission potential less than, but potentially close to that of, symptomatic infected individuals[35,36,37]
α	0.3	0.5	Proportion of infections that are asymptomatic	Mean of 43.3%[38], 44%[39]
δ	0.1	0.3	Fatality rate among hospitalizations	Demographic weighted average that will vary by location, see Verity et al.[31], 5%[40]
ρ_D	13 days	20 days	Time from hospitalization to death	Mean of 16 days[41], mean of 17.8 days[31], highly variable by region[33]
$1-\mu$	0.025	0.075	Proportion of symptomatic infections that require hospitalization	Demographic weighted average that will vary by location, see Verity et al.[31]

Table 3: Location-specific parameter range estimates. Population sizes obtained from the US Census Bereau[42]

Parameter	*Santa Clara County, CA*	*Los Angeles County, CA*	*Miami-Dade County, FL*	*King County, WA*	*Fulton+DeKalb, GA*
Epidemic Start Date	01-Jan - 05-Feb	01-Jan - 31-Jan	01-Jan - 16-Mar	01-Jan - 04-Mar	01-Jan - Mar-07
Population Size	1,927,852	10,039,107	2,716,940	2,252,782	1,755,830

The time-varying transmission parameter, β_t, describes the average per capita rate of contact between susceptible and infectious people at time t, multiplied by the per-contact transmission probability. We modeled β_t as a function of human movement using the scaling function:

$$\beta_t = \beta_0 \beta_m^{\theta}, \tag{19}$$

which treats β_t as an exponentially decreasing function of physical distancing (θ; on a scale of 0-1 where 0 is no physical distancing, and 1 is maximum physical distancing). Here, $\beta_0 \beta_m$ is the estimated minimum possible transmission rate given minimal human movement (i.e., maximal physical distancing) and thus contact rate. To model human movement we use SafeGraph's "Shelter in Place Index"[43], which measures the proportion of cell phone devices that are staying home.

To model individual heterogeneity in SARS-CoV-2 transmission rate, we allow individuals to vary over time by modeling an individual's transmission rate in each time step as a Gamma distributed random variable with a dispersion such that the sum of an individual's transmission rates over the duration of their infection approximates a Gamma distributed random variable with dispersion equal to previous Negative Binomial parameterizations for reproductive number SARS-CoV-1 ($k = 0.16$)[44], which closely approximates estimates of overdispersion for SARS-CoV-2[12,45]. Because we model the transmission rates as the multiplication of contact rate and infection probability, this heterogeneity implicitly considers both variation among individuals in infectiousness and contact rate, and can be thought of as modeling superspreading *periods or events*—windows in time when an infected individual has a particularly high transmission rate. To incorporate this variation into an average time step β_t, we model β_t as the average of the transmission of all infected individuals at time t. To do so we apply the property of Gamma distributions that the mean and variance of N samples from a Gamma distribution with defined rate and scale is itself a Gamma distribution with mean equal to that of the original Gamma distribution and variance equal to the variance of the original Gamma divided by N. A full derivation of the equivalence between the individual time step

transmission rate distributions (which we will hereafter refer to with π), the individual infectious period transmission rate distributions, and the population-level transmission rate distribution is available in the Appendix.

We assume that observed deaths are a Negative Binomial random variable with a mean equal to total new deaths accumulated over the observation period (i.e., one day for this analysis), and a dispersion parameter that we estimate. We also assume that daily observed cases are a Negative Binomial random variable, but have a mean equal to the daily number of new symptomatic infections multiplied by a daily detection probability that we estimate from the data. We model daily detection probability as a monotonically increasing logistic function:

$$\frac{\omega}{1+e^{-m(t-\phi)}}, \tag{20}$$

where ω is the maximum fraction of symptomatic cases detected, m is the logistic growth rate, and ϕ gives the location of the inflection point (where the probability of detection equals one half of the maximum detection probability, ω). Because ϕ can be estimated to be in the future, the probability of detection of an infected case in the present can be any value between 0 and ω. We estimate newly observed cases to be a fraction of all new symptomatic infections at time t. Though this ignores testing asymptomatic infections, any detection of asymptomatic infections will be captured as a higher estimated detection fraction of symptomatic infections.

Fitting the Model

We use COVID-19 death and case data from The New York Times, based on reports from state and local health agencies (available at `https://github.com/nytimes/covid-19-data`). Using these data, which are available for all counties in the US, and any form of human movement data that can be scaled to 0-1, our model can be used to fit infection dynamics in any county.

For computational efficiency, we assumed point estimates for some parameters (Table 1) and sampled over uncertainty in others (Tables 2, 3) by drawing 600 sobol sequences, an efficient method for sampling input parameters[46], across a range of plausible values for each in order to form 600 plausible parameter sets. For each parameter set we used the package `pomp`[47] in the statistical programming language `R`[48] to estimate the following parameters: β_0: transmission rate over an entire infection in the absence of social distancing; β_m: estimated transmission given zero human movement; E_0: number of exposed individuals that initiate the epidemic; ω, m, and ϕ: maximum, slope, and inflection point day of the sigmoidal case

detection function; θ_d: Negative Binomial dispersion parameter for deaths; and θ_c: Negative Binomial dispersion parameter for cases. We fit all parameters to daily deaths, cases, and mobility in two steps. First, for each of the 600 parameter sets we used the `mif2` function in `pomp` with random starting conditions, 120 iterations and 2000 particles. We then continued to fit the 60 parameter sets with the highest log likelihoods for an additional 200 iterations using 2000 particles. Each county took approximately nine hours to fit using twenty cores.

We calculated $\mathcal{R}_E$ at each time t as estimated β_t times the median proportion of the population remaining susceptible on each day across 300 simulated epidemics, with simulated epidemics that did not reach at least a total of 100 infected discarded, times the average infectiousness over an infection (as defined by our model structure) using the 10 parameter sets with the largest negative log likelihoods as determined by the second fitting step.

Simulating epidemics under interventions

Any intervention type, intensity, or duration can be modeled using this framework and open-source code (available at https://github.com/morgankain/COVID_interventions) given that it can be written as a function that modifies either human movement or β_t (e.g., social distancing or a pharmaceutical intervention that reduces the probability of infection). Previously we considered the impacts of various social distancing initiatives on epidemic dynamics using a similar model formulation[13]. Here we consider interventions that reduce the skew of the individual time step transmission rate distribution (π), and thus the average time-varying transmission rate β_t; this is our mathematical representation of reducing highly infectious contact periods or events, which for COVID-19 tend to occur in crowded enclosed environments (e.g., church choirs and exercise classes). Specifically, we model truncation of the π distribution by assuming that all samples within the top X% of the π distribution are resampled. To visualize the dynamics of interventions, for each location we simulate 300 epidemics from the maximum likelihood estimate across the 600 parameter sets. The uncertainty band we plot represents the central 95% range of outcomes seen across all stochastic realizations that resulted in epidemics for this parameter set, and thus should not be taken as representation of uncertainty in parameter values or model structure.

Results

Epidemic trajectories

The model produced realistic fits to five contrasting epidemiological settings—King County, Washington; Los Angeles County, California; Santa Clara County, California; Fulton and Dekalb Counties, Georgia; and Miami-Dade County, Florida (hereafter, Seattle, Los Angeles, Santa Clara County, Atlanta, and Miami). Among these locations, we estimated that prior to interventions, $\mathcal{R}_0$ ranged between approximately 2 and 4 (Figure 1). We also estimated that $\mathcal{R}_E$ dropped below one following shelter-in-place orders in all counties, though only briefly in some locations. In particular, in Miami, Los Angeles, and Atlanta $\mathcal{R}_E$ climbed above 1 by mid-May and daily cases and deaths have plateaued or continue to grow into June. Though $\mathcal{R}_E$ remained below 1 into at least early June in Seattle and Santa Clara County, as of June 18 $\mathcal{R}_E$ is $\sim$1 and cases are rising again.

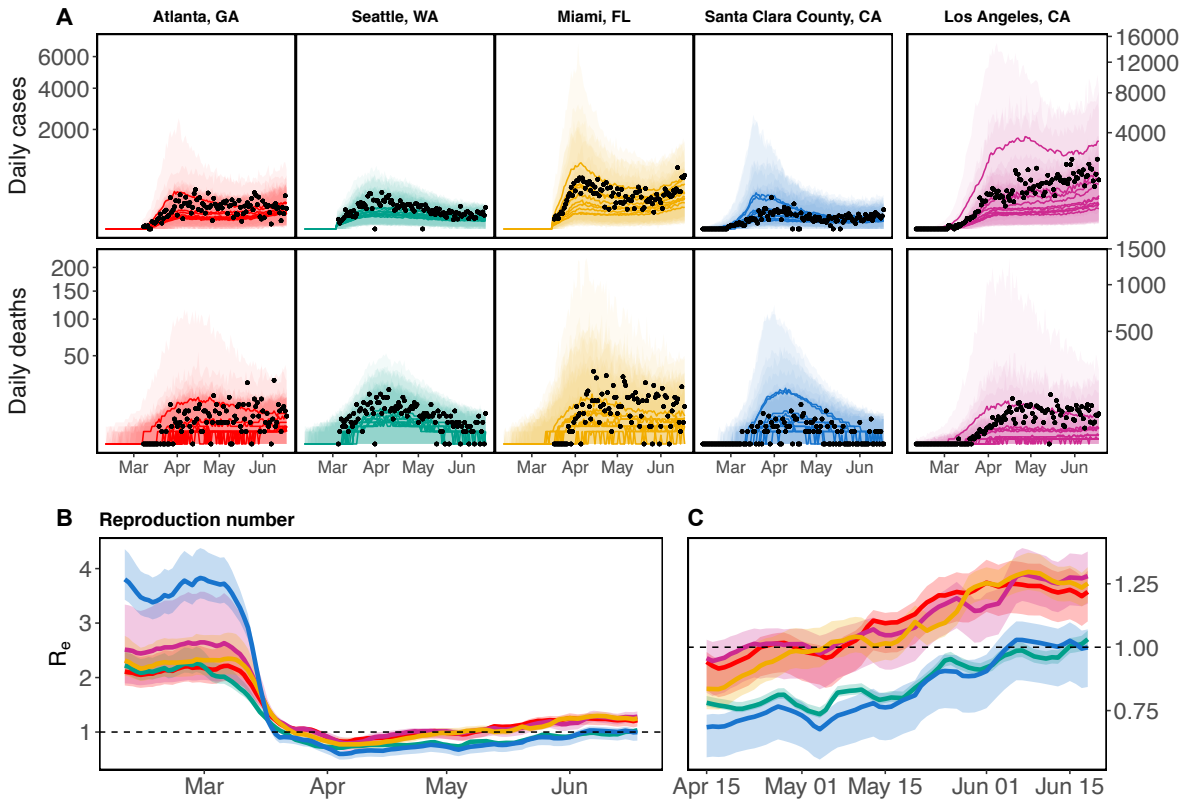


Figure 1: Model estimated daily cases and deaths (A), and reproduction number (B, C) for five locations: Atlanta (red), Seattle (green), Miami (gold), Santa Clara County (blue), and Los Angeles (purple). Los Angeles is displayed on a different y-axis due to differences in magnitude of reported deaths and cases. For each county, we show the 10 model fits with the best log likelihoods. Panel C show the same results pictured in B, but are zoomed in to April 15 - June 18 to better show the dynamics around $\mathcal{R}_E = 1$. Black points are observed daily deaths and reported cases in each county. Solid lines display mean of model simulated trajectories (A) and mean R_e (B, C). Ribbons show the range of estimated R_e (B, C) or 95% CIs over stochastic simulation from each model fit (A). Vertical axes in panel A are square root transformed for visibility.

Interventions

As a basis for comparison, focusing on just two locations—Los Angeles and Seattle—if shelter-in-place were simply lifted, a second peak would be inevitable in the absence of any non-pharmaceutical interventions (Figure 2, blue). However, non-pharmaceutical interventions, including continuing shelter-in-place, infected isolation with intermediate levels of shelter-in-place, or averting superspreading with intermediate levels of shelter-in-place are capable of limiting epidemic growth (Figure 2) and keeping $\mathcal{R}_E$ near or under 1. Here, we consider intermediate levels of shelter-in-place that correspond to mobility levels that are an average of baseline mobility prior to social distancing and final mobility levels observed in the last week of data. Either an infected isolation strategy that reduces to intermediate levels of shelter-in-place and

catches 90% of all mild and severe cases of COVID-19 before they are able to transmit (Figure 2, green), or a truncation strategy that similarly reduces to intermediate levels of shelter-in-place but removes the top 1% of the individual time step transmission rate distribution (π) with 75% efficiency (Figure 2, purple) are able to suppress epidemic growth (and reduce $\mathcal{R}_E$ to below one) in Los Angeles, CA and Seattle, WA.

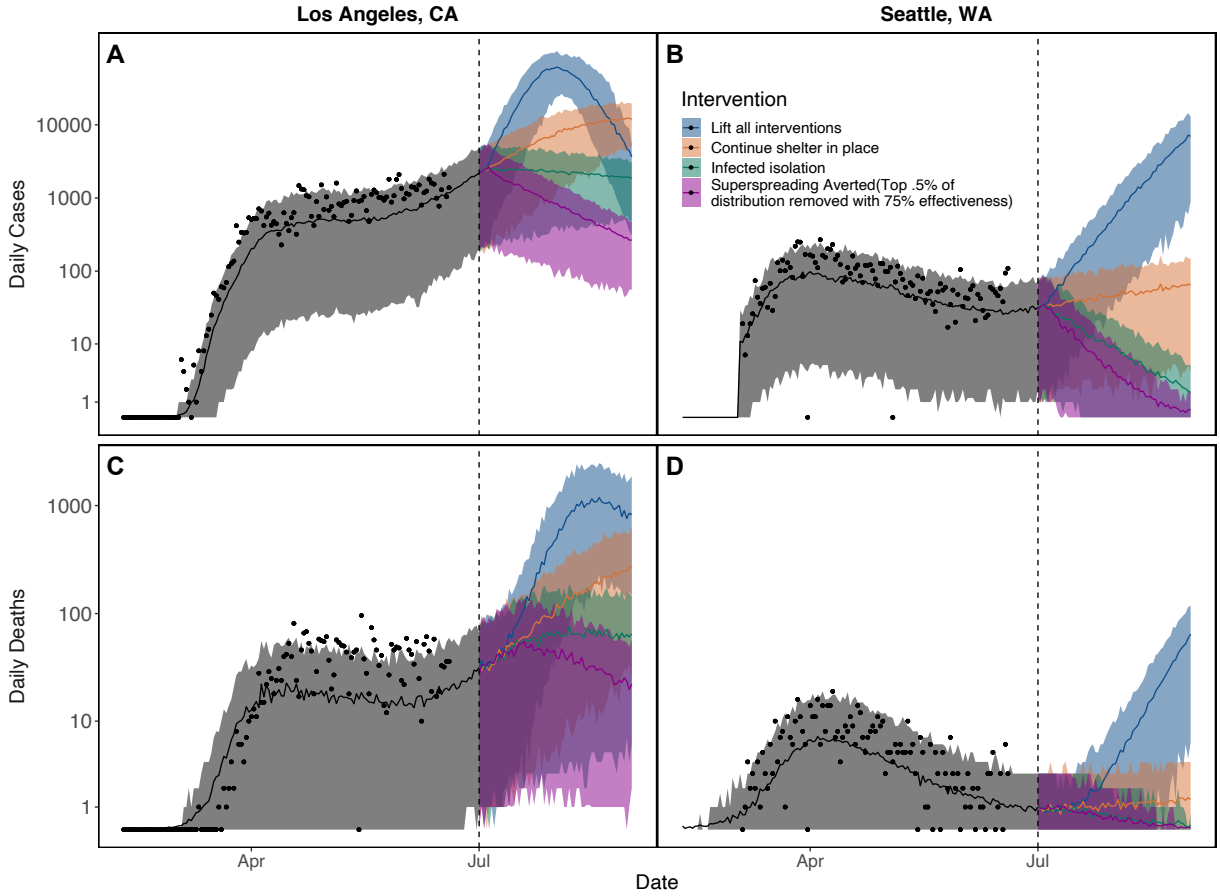


Figure 2: Maintaining shelter-in-place (SIP; orange), test-and-isolate (green), or superspreading aversion (purple) strategies over long periods is necessary to prevent a major epidemic resurgence (blue) in each location where we fit our model (shown here for Los Angeles, CA [A, C] and Seattle, WA [B, D]). However, continuing SIP at current levels (orange) will lead to an increase in daily cases in both Los Angeles (A) and Seattle (B). Daily reported cases are shown in (A) and (B) and daily deaths in (C) and (D). For both shelter-in-place and truncation interventions we assume an intermediate level of mobility (an average of baseline mobility prior to social distancing and final mobility levels observed in the last week of data). Bands show 95% CI on stochastic simulations of daily cases and deaths for the single maximum likelihood parameter set. Dates range from February through September of 2020. Vertical axes are log transformed for visibility.

Curtailing superspreading

Limiting opportunities for superspreading by "chopping off the tail" of the contact rate or infectiousness distributions can be highly effective at epidemic control (Figure 2), driving epidemic growth to be negative and bringing the average number of secondary cases ($\mathcal{R}_E$) below 1. An example truncation intervention is

illustrated in Figure 3: because the individual transmission rate distribution, π, over a 4-hour time period is so skewed (Figure 3A; see appendix for derivation), truncating the upper 0.1% yields a large reduction in the mean and a moderate reduction in the variance of the population-level average transmission rate (Figure 3B, shifting from red to blue distribution). A variety of possible truncation strategies exist, including eliminating varying proportions of π (e.g., upper 0.5%, upper 1%) with varying levels of efficiency (ranging from 50-100%) (Figure S2).

An alternative measure of the impact of averting superspreading (i.e., truncation interventions) is how much general social distancing can be avoided by instead truncating the transmission rate distribution. Prior to social distancing orders, the estimated proportion sheltering in place (SIP, for short) ranged from $\sim$20-22% across our focal locations (Figure 3C, triangles), indicating the baseline level of mobility. If we combine SIP with truncation interventions, a variety of combinations are predicted to provide epidemic control (for example, by reducing transmission rates such that $\mathcal{R}_E$ in a fully susceptible population would be 1; Figure 3). If the truncation intervention is 100% effective, truncating only approximately the upper 0.15% of individual transmission rates, π, (Figure 3A) is effective enough to maintain transmission rates such that $\mathcal{R}_E$ would be 1 in a fully susceptible population, while allowing mobility levels to return to baseline (Figure 3C). Alternatively, if truncation interventions are only half as effective, the same 0.15% truncation intervention would require moderate-strong social distancing (SIP from $\sim$30-45%; Figure 3C). The nonlinear effects of social distancing and truncation on transmission make the combination of interventions needed to maintain epidemic control sensitive to the efficiency and strength of each intervention mechanism.

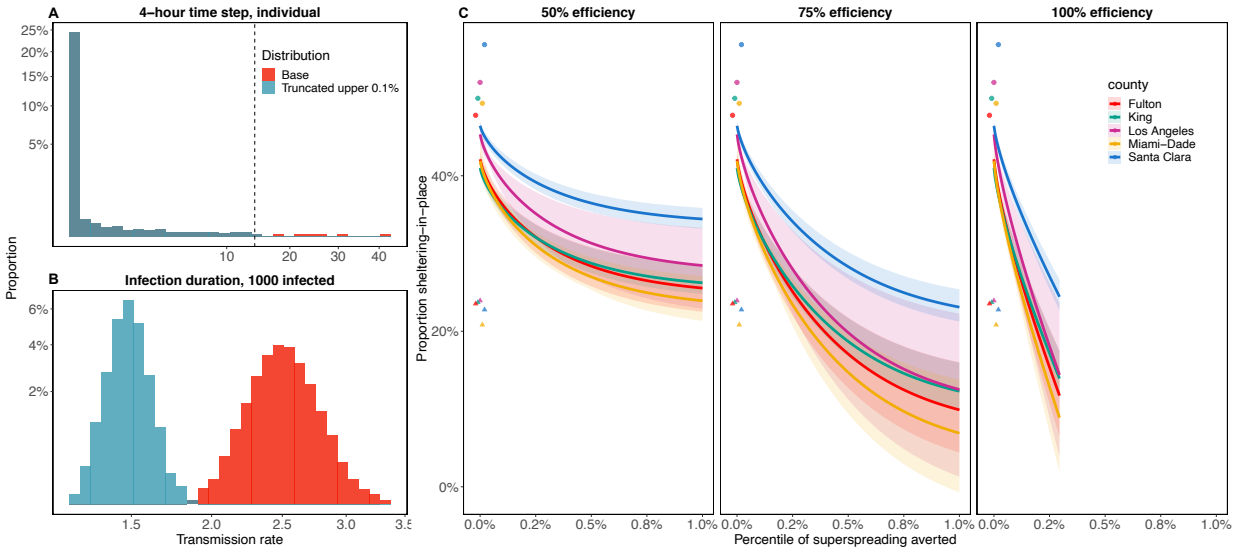


Figure 3: Example of how truncating the individual-level transmission rate distribution, π, (A) affects the population-average transmission rate (B), and combinations of sheltering-in-place (SIP) and truncation strategies that reduce $\mathcal{R}_E$ to one in a fully susceptible population (C). The three panels in C show the combinations of truncation and SIP that produce an $\mathcal{R}_E$ of one for three levels of truncation efficiency. (A) Truncation at the upper 0.1% of π (sampled over a 4-hour time step), in which truncation occurs for all values above the dashed line. (B) Resulting effect on the population-level average infection rate when there are 1000 infected people currently in the population, where the original distribution is in red and the truncated distribution is in blue. The distribution is shown over 10,000 simulations for a population characterized by an individual reproduction number distribution with mean of 2.5 and overdispersion parameter, $k = 0.16$. Horizontal and vertical axes in A and B are square root transformed for visibility. In C, the triangles show baseline SIP in each location and circles show max SIP reached during social distancing. Solid lines indicate the mean over the ten best fits, and the ribbon is the full range of estimates from these fits.

Superspreading and epidemic resurgence

Even if the epidemic is brought almost entirely under control (e.g., to within 1-5 infected individuals remaining in the population), epidemic resurgence remains a possibility if interventions wane, allowing $\mathcal{R}_E$ to increase above one. As we show in (Figure 3), many different combinations of SIP and truncation can be used to produce the same $\mathcal{R}_E$ (in Figure 3, an $\mathcal{R}_E$ of 1); however, epidemic dynamics will vary by combination because of the variation in individual time-step transmission rates, π. If $\mathcal{R}_E$ rises above one because interventions are relaxed, the specific combination of SIP and truncation that remains in place will determine the resulting dynamics. Here we examine how different truncation interventions will affect epidemic extinction probability and the size of epidemic resurgence when it does not go extinct. We compare the full effect of truncation interventions (which influence both the mean of the transmission rate distribution and its shape) to the effects of truncation when $\mathcal{R}_E$ is held constant by scaling SIP, reflecting only truncation effects on the transmission rate distribution shape (variance, skew, etc.).

With few infected individuals and $\mathcal{R}_E > 1$, stochasticity and heterogeneity in β_t can either lead to extinction, moderate resurgence, or explosive resurgence. Keeping interventions in place that remove even a tiny percent of the largest β_t values can help to avoid the more explosive events (Figure 4). Truncation markedly reduces the probability of explosive epidemic resurgence (Figure 4A) both by increasing the extinction probability (Figure 4B) and by reducing the magnitude of resurgent epidemics when they do occur (Figure 4C). While epidemic size was less sensitive to the number initially infected when resurgences do occur (Figure 4C), the stochastic extinction probability was extremely sensitive to the difference between even one, three, or five remaining infections (Figure 4B). Much, but not all, of the benefit of truncation comes from changing the mean transmission rate (and therefore $\mathcal{R}_E$). When $\mathcal{R}_E$ is held constant by adjusting SIP, effects of truncation are more moderate. An increase in efficiency at truncating the top 0.1% of the β_t distribution noticeably decreases the number of infected 42 days after interventions are relaxed (Figure 4D,F). However, because of the need to slightly reduce SIP to hold $\mathcal{R}_E$ constant under truncation, truncation of π marginally decreases the extinction probability of the epidemic, which remains much more sensitive to the number initially infected (Figure 4E). The highly stochastic nature of epidemic growth when cases are rare, combined with the fact that each truncation leaves behind highly skewed distributions regardless of the truncation parameters, results in even 10,000 epidemic simulations producing noisy patterns across intervention scenarios. Similar patterns are seen as more of the π distribution is truncated (Figure S4).

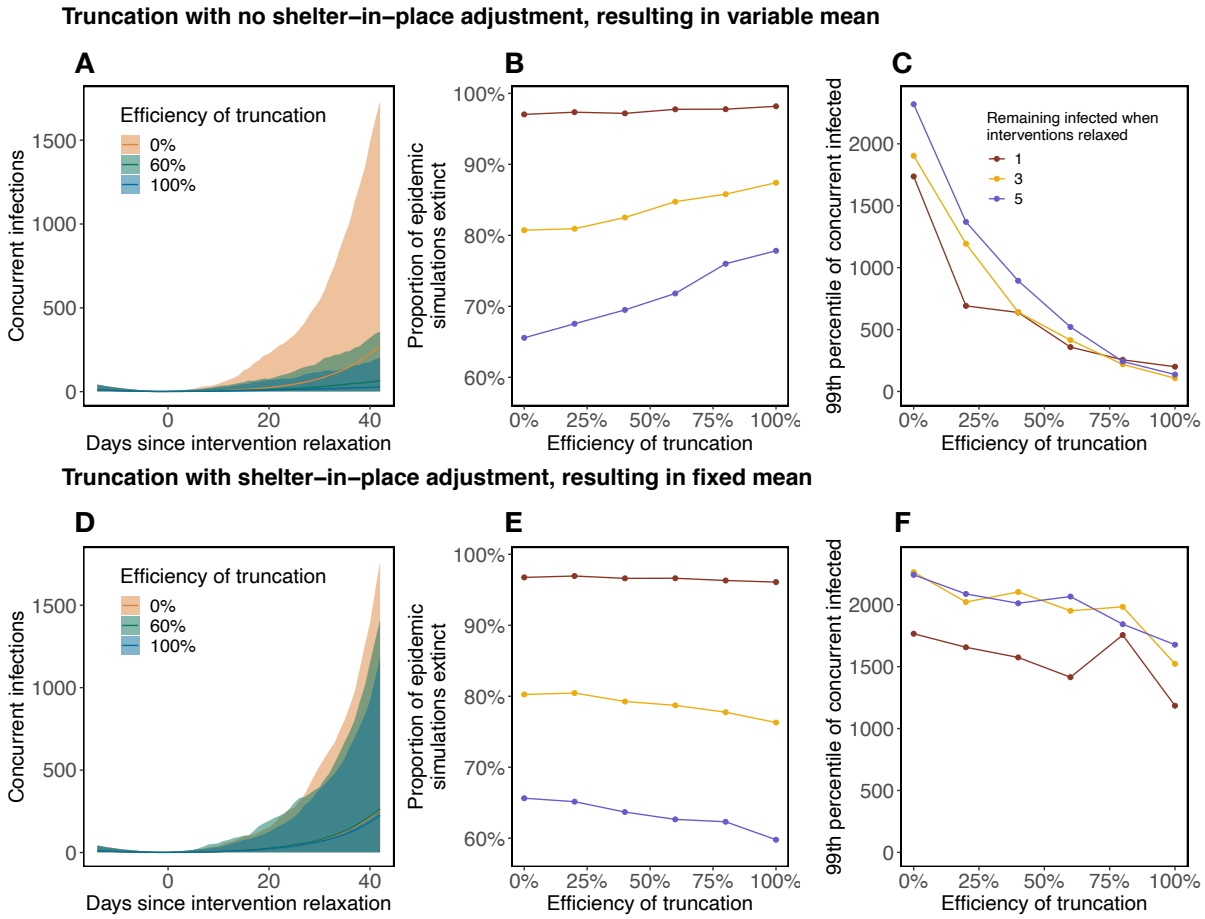


Figure 4: Effects of transmission rate truncation on epidemic die-out and explosive resurgence. With skewed individual variation in transmission rate, relaxing social distancing interventions when infections become rare (allowing $\mathcal{R}_E$ to increase above one) may lead to explosive stochastic epidemic resurgence. Top panels (A-C) show the overall effect of truncation interventions, including effects on both the mean and shape of the transmission rate distribution, and resulting $\mathcal{R}_E$. Bottom panels (D-F) show the effect of truncation when $\mathcal{R}_E$ is held constant by rescaling shelter-in-place at the time of intervention relaxation. Specifically, for a 0% truncation efficiency we simulate epidemic resurgence assuming $\mathcal{R}_0 = 2$, which results in an $\mathcal{R}_E = 2 \cdot S/N$ at the time of resurgence, which will vary by simulation (where S is the number of susceptible individuals and N is the total population size). In panels (A-C) as truncation efficiency increases $\mathcal{R}_E$ decreases; in panels (D-E) we scale shelter-in-place to retain an average $\mathcal{R}_E = 2 \cdot S/N$ across truncations. Simulations are performed with varying efficiencies of truncation of the top 0.1% of the π distribution. Envelopes in (A) and (D) show the central 98% of resurgent simulations (across 10,000 total simulations) for three efficiencies of truncation (0% in orange, 60% in green, 100% in blue). The proportion of epidemic simulations that go extinct within 42 days of intervention relaxation for thresholds of 1 (red), 3 (gold), and 5 (blue) infected individuals is shown in (B) and (E). The upper 99th percentile of concurrent infections 42 days after intervention relaxation in resurgent simulations for the same thresholds is shown in (C) and (F).

Discussion

Understanding local epidemiological dynamics of COVID-19—and the impact of heterogeneity on those dynamics—remains a challenge due to both limited and imperfect data in most regions and ever evolving

interventions and adherence. Reported cases are only a small fraction of all infections, and the proportion of symptomatic cases that are detected remains highly uncertain and variable over space and time. Our approach takes an important step toward capturing locally-specific epidemic dynamics and the impact of heterogeneity across settings by providing a platform (including a mathematical model and open access code) for estimating time-varying transmission rates (β_t) from death, mobility, and imperfectly observed case report data, all of which are publicly available. The model can estimate epidemic dynamics and transmission rates over time across epidemiological settings that vary in population size, demography, and control. By incorporating individual variation in contact rates (or, equivalently, infectiousness) into time step transmission rate distributions, we incorporate some of the known effects of heterogeneity without requiring detailed information on population mixing, structure, social networks, or movement patterns. We find that control measures in March of 2020 rapidly brought the average reproduction number—$\mathcal{R}_E$—from $\sim$2–4 to below 1 in all locations we considered in early April. However, as of June 18, $\mathcal{R}_E$ has once again drifted above one in all of these locations except possibly in Seattle, WA and Santa Clara County, CA, where it remains unclear if it is greater or less than one.

Non-pharmaceutical interventions will be necessary to control COVID-19 in all settings until better pharmaceutical options (in particular, effective vaccines) are widely available. Social distancing in the general population is effective but costly: it is a blunt and imprecise tool. The social and economic necessity of relaxing social distancing demands safe exit strategies based on more precise, targeted interventions to reduce transmission. Testing and isolating symptomatic people, combined with contact tracing, remains the gold standard intervention for limiting onward transmission as social distancing is lifted, but it is expensive and capacity remains limited in many settings. Our model shows that it is possible to target interventions even without precise information on specific population mobility, mixing, and infectiousness patterns, by limiting just the most high-risk activities, such as large gatherings and indoor events that have many close contacts. How much can be gained from these common sense interventions that reduce or eliminate opportunities for superspreading while allowing smaller and safer activities to resume? We find that these truncation interventions, which eliminate the upper percentiles of contact rates in the population, and thereby transmission rates, can be highly effective at maintaining epidemic control (Figure 2), particularly when combined with mild to moderate social distancing (Figure 3). Importantly, even after epidemic control is achieved and case numbers drop very low, "chopping off the tail" can provide powerful insurance against explosive resurgence after social distancing interventions are otherwise lifted (Figure 4).

What does "chopping off the tail" mean in practical terms? Five types of factors tend to promote superspreading: (1) high rates or intensity of contact between people or with surfaces; (2) large aggregations of people; (3) poorly ventilated physical environments, especially indoors[49]; (4) highly infectious individ-

uals; (5) highly susceptible recipient population[4,6,50]. Many settings where SARS-CoV-2 superspreading has occurred—including nursing homes[26], exercise classes, bars and restaurants[49], funerals, churches[51], meat-packing plants[52]—combine multiple risk factors. For example, choir practices combine high densities of people, a high-risk activity (singing)[53], and potentially poorly ventilated indoor spaces; long-term care facilities combine mobile, high-contact caregivers with highly vulnerable residents, often in high-density indoor spaces. Some superspreading events may be easier to eliminate than others. Clearly, healthcare and long-term care facilities serve critical functions despite their high-risk nature, and taking all possible steps for decontamination and personal protection in these facilities is critical to mitigate this risk[50]. On the other end of the spectrum, voluntary, large, indoor events that are mainly for entertainment and could be postponed—gyms, clubs, sporting events, concerts, large lectures—may be the most viable option to reduce superspreading and "chop off the tail" of the contact rate distribution[54]. While these common sense interventions are not novel suggestions[6], and are already part of reopening plans in almost all locations, our work allows a direct comparison of how much general social distancing is avoided by eliminating a fraction of these high-risk events (Figure 3). Truncation strategies are even more desirable in light of their effectiveness at preventing explosive resurgence after controls are otherwise lifted (Figure 4). Mapping actual event types onto the contact rate distribution to determine how particular superspreading reduction policies would affect control remains an important next step. Importantly, associating superspreading with events and locations, rather than specific people, can avoid the stigma sometimes associated with being identified as a superspreader[4].

The impact of truncation interventions is two-fold. First, removing the upper tail of the individual transmission rate distribution reduces the population-level mean, often dramatically (Figure 3A,B). If the mean transmission rate already placed $\mathcal{R}_E$ near 1 (for example, due to other interventions), then additional truncation could be enough to cross this critical threshold. However, most intervention strategies that bring $\mathcal{R}_E$ to 1 already include prohibiting large gatherings, especially indoors, so additional truncation may not be possible within the context of first-wave interventions. However, truncation also acts on the variance and skew of the transmission rate distribution, though these effects are smaller than the effect on the mean (Figure 4D-F compared to A-C). Given that super-spreading events are particularly dangerous when cases are few (in the early or late phases of the epidemic)[2], sustained truncation interventions could be extremely important for preventing explosive stochastic re-emergence when low case numbers allow general social distancing to be lifted (Figure 4). In this scenario, resurgence remains rare (Figure 4B) but possible because individual variation in transmission rates is large; most of the time infectious people transmit to few others, but occasionally someone infects dozens (Figure 4A), quickly overwhelming testing, contact tracing, and isolation efforts. Sustained truncation dramatically reduces the probability of explosive resurgence, and

constrains incipient transmission chains to be smaller and more manageable.

One limitation on understanding the effect of heterogeneity in transmission in particular locations is the challenge of estimating epidemiological parameters from noisy and imperfect data: necessarily a balancing act between model simplicity and complexity. Here, we rely on metrics of heterogeneity previously estimated for SARS-CoV-1 and SARS-CoV-2[2,12,45] instead of estimating them directly from data; we focus our parameter estimation on the mean of the transmission rate distribution. Heterogeneity in contact rates or infectiousness, and the resulting distributional variance and skew, may vary based on local patterns of movement, contact, behavior, and population demography. This heterogeneity can have important consequences: in some cases epidemics with low mean R_0 can actually infect a larger proportion of the population than epidemics with higher mean R_0—as was the case for the 1918 influenza pandemic as compared to the 2014 Ebola outbreak—due to the heterogeneity in transmission rates, as described by higher moments of the secondary case distribution[7]. The true epidemiological parameters in any given location, and the extent of our uncertainty in these parameters, also remain unknown because of the computational challenges of parameter estimation given the limited information contained in noisy case, death, and mobility data. For example, depending on how a particular candidate parameter combination weights the noisiness of cases and deaths and estimates initial conditions, transmission rate estimates can vary substantially (Figure 1). Fully characterizing uncertainty in model structure and parameter values in this context is difficult. Future work that directly estimates case ascertainment rates (e.g., through metrics of percentage of tests that are positive, age distributions of positive tests, epidemiological contact information on cases, and analysis of viral genome sequences[55]), as well as more detailed mobility and contact network information[17] could help to improve the model fit to the full shape of the transmission rate distribution.

First-wave interventions that eliminated large social gatherings and indoor activities and mandated mask-wearing and physical distancing have likely already affected the heterogeneity in transmission rates, by eliminating many of the high-risk events likely to fall into the upper tail of the distribution. It is important to recognize that as social distancing interventions relax, sustaining such truncation interventions may be critical for keeping transmission down to levels manageable through testing, contact tracing, and isolation. This truncation strategy can potentially reduce the social and economic costs of non-pharmaceutical interventions on the general populace, and facilitate sustained adherence by allowing lower-risk activities to resume while insuring against a resurgence. Ultimately, an unmitigated epidemic, whether as a first or second wave, would kill thousands to tens of thousands of people in each of the locations we studied, reinforcing the point that aiming for population herd immunity through naturally acquired infections is not a viable public health strategy. Instead, exit strategies that can sustain epidemic control after shelter-in-place orders end, including truncating the transmission rate distribution, will be necessary until an effective

342 vaccine can be developed and widely distributed.

Data and Code Availability

Data used in this study are available at: https://github.com/nytimes/covid-19-data.
Code used to produce the results in this study are available at: https://github.com/morgankain/COVID_interventions.

Declaration of Interests

We declare no competing interests.

Acknowledgements

We thank the members of the Mordecai Lab at Stanford University for feedback on our model. Funding provided by: the National Science Foundation (DEB-1518681); the National Institutes of Health (National Institute of General Medical Sciences: R35GM133439); the Natural Capital Project; the Helman Scholarship; the Terman Award. Morgan Kain was supported by the Natural Capital Project. Marissa Childs was supported by the Illich-Sadowsky Fellowship through the Stanford Interdisciplinary Graduate Fellowship program at Stanford University.

Author Contributions

MPK helped to design and code the model, conduct model fits and simulations, troubleshoot problems and update the model, write and revise the manuscript.
MLC helped to design and code the model, conduct model fits and simulations, troubleshoot problems and update the model, write and revise the manuscript.
ADB helped to refine the model, conduct model fits and simulations, troubleshoot problems and update the model, and revise the manuscript.
EAM helped to design the model, provide conceptual framing, troubleshoot problems, and write and revise the manuscript.

Appendix

Derivation of the population-level β_t distribution

In heterogeneous populations, the expected number of secondary infections caused by a particular individual (or the individual reproductive number, ν) can be modeled as a negative binomial random variable with mean R_0 and overdispersion parameter k[2,6,56], i.e. $\nu \sim NB(k, \frac{k}{R_0+k})$. This is equivalent to modeling ν as a Poisson random variable whose mean is itself a random gamma variable with shape k and scale R_0/k,

$$\begin{aligned}\nu &\sim \text{Poisson}(\theta)\\ \theta &\sim \Gamma(k, \frac{R_0}{k})\end{aligned}$$

Now let d be the duration of infection for an individual and τ be a time step. Using the fact that $k = \sum_M k/M$, we have

$$\begin{aligned}\theta &\sim \Gamma\left(k, \frac{R_0}{k}\right)\\ &\sim \Gamma\left(\sum_{d/\tau} \frac{k\tau}{d}, \frac{R_0}{k}\right)\\ &\sim \sum_{d/\tau} \Gamma\left(\frac{k\tau}{d}, \frac{R_0}{k}\right).\end{aligned}$$

Thus for a constant duration of infection d, we have the individual infection rate over a time step

$$\pi \sim \Gamma\left(\frac{k\tau}{d}, \frac{R_0}{k}\right).$$

When there are N infected individuals, the average infection rate β_t over a time step is

$$\begin{aligned}\beta_t &= \frac{1}{N}\sum_{i=1}^{N} \pi\\ &\sim \frac{1}{N}\sum_{i=1}^{N} \Gamma\left(\frac{k\tau}{d}, \frac{R_0}{k}\right)\\ &\sim \Gamma\left(\frac{Nk\tau}{d}, \frac{R_0}{Nk}\right),\end{aligned}$$

which will have mean $R_0\tau/d$ and variance $\frac{R_0^2\tau}{Nkd}$. Notably, this behaves well with scalings on R_0 as a function of interventions: Let θ be the amount of physical distancing occurring in the population on a scale of 0-1 where 0 is no physical distancing, and 1 is maximum physical distancing, and f be a function mapping θ to a scaling on R_0. Now $\beta_t \sim f(\theta)\Gamma\left(\frac{Nk\tau}{d}, \frac{R_0}{Nk}\right) = \Gamma\left(\frac{Nk\tau}{d}, \frac{f(\theta)R_0}{Nk}\right)$, which means that properties of the distribution are preserved with $f(\theta)R_0$. Specified as a gamma white noise process $\Gamma_{WN}(\sigma, \mu)$ which has mean μ and variance $\sigma^2\mu$, this is equivalently

$$\beta \sim \Gamma_{WN}\left(\sqrt{\frac{R_0}{Nk}}, \frac{R_0\tau}{d}\right).$$

There are two main differences between the above derivation and our model formulation (note that for the following we assume $f(\theta) = 1$):

1. above, the number of new infections in a time step should be

$$\sum_{i=1}^{N} \text{Poisson}(\pi_i),$$

where each π_i is i.i.d. as $\Gamma\left(\frac{k\tau}{d}, \frac{R_0}{k}\right)$. In our model the number of new infections in a time step is

$$\mathcal{B}(S, 1 - \exp(-\beta_t(C_a I_a/N + C_p I_p/N + C_m I_m/N + C_s I_s/N))),$$

where β_t is the average transmission rate over all individuals infectious during that time step and $\mathcal{B}$ is a binomial process. For large S and small $I_a + I_p + I_m + I_s$, this approximates a Poisson distribution for the number of secondary cases from each infected individual in each time step and the total secondary infections caused by an individual over their infectious period.

2. above, we assume a constant duration of infection. In our model periods are Erlang distributed given our division of stages (e.g. I_a) into n sub stages, each with the same period[14,15]. This marginally increases the variance in the per-infectious period d istribution as we show in Figure S5.

Derivation of the relationship between R, Gamma truncation, and the fraction of individuals sheltering in place

Let d be the average duration of infection, τ be a time step, θ be the proportion of the population sheltering in place, and $X_{p,\eta}$ be a random variable X with right truncation, where truncation occurs at the p-th percentile, with probability η. Suppose that in a given time step we truncate the individual infection rate (π) over a time step at the p-th percentile with probability η. Then the reproduction number is

$$
\begin{aligned}
R &= \mathbf{E}\left[\frac{d}{\tau}\beta_t\right] \\
R &= \mathbf{E}\left[\frac{d}{\tau N}\sum_{i=1}^{N}\pi_{p,\eta}\exp\left(\log(\beta_{min})\theta\right)\right] \\
R &= \frac{\beta_{min}^{\theta} d}{\tau N}\sum_{i=1}^{N}\mathbf{E}[\pi_{p,\eta}] \\
R &= \frac{\beta_{min}^{\theta} d}{\tau}\mathbf{E}[\pi_{p,\eta}] \\
R &= \frac{\beta_{min}^{\theta} d}{\tau}\left(\eta\mathbf{E}[\pi_{p,1}] + (1-\eta)\mathbf{E}[\pi_{1,1}]\right) \\
\beta_{min}^{-\theta} &= \frac{d}{R\tau}\left(\eta\mathbf{E}[\pi_{p,1}] + (1-\eta)\mathbf{E}[\pi_{1,1}]\right) \\
-\theta\log(\beta_{min}) &= log(d) - \log(R) - \log(\tau) + \log\left(\eta\mathbf{E}[\pi_{p,1}] + (1-\eta)\mathbf{E}[\pi_{1,1}]\right) \\
\theta &= \frac{log(d) - \log(R) - \log(\tau) + \log\left(\eta\mathbf{E}[\pi_{p,1}] + (1-\eta)\mathbf{E}[\pi_{1,1}]\right)}{-\log(\beta_{min})}
\end{aligned}
$$

For a truncated gamma distribution with shape a and scale b with upper truncation at u, the expected value is

$$\mathbf{E}[\Gamma(a,b,u)] = \frac{b\left[\Gamma(a+1,0) - \Gamma(a+1,u/b)\right]}{\Gamma(a,0) - \Gamma(a,u/b)},$$

where Γ is the upper incomplete gamma function. See Okasha and Alqanoo (2014) [eq.29][57] for the full derivation.

Letting γ be the lower incomplete gamma function, it follows that

$$
\begin{aligned}
\mathbf{E}[\pi_{p,1}] &= \frac{R_0/k\left[\Gamma(k\tau/d+1,0) - \Gamma(k\tau/d+1,uk/R_0)\right]}{\Gamma(k\tau/d,0) - \Gamma(k\tau/d,uk/R_0)} \\
&= \frac{R_0/k\left[\gamma(k\tau/d+1,uk/R_0)\right]}{\gamma(k\tau/d,uk/R_0)}
\end{aligned}
$$

where u is the p-th percentile of π[57]. Then

$$\theta = \frac{log(d) - \log(R) - \log(\tau) + \log\left(\eta\frac{R_0/k[\gamma(k\tau/d+1,uk/R_0)]}{\gamma(k\tau/d,uk/R_0)} + (1-\eta)\frac{R_0\tau}{d}\right)}{-\log(\beta_{min})}.$$

References

[1] Woolhouse ME, Dye C, Etard JF, Smith T, Charlwood J, Garnett G, et al. Heterogeneities in the transmission of infectious agents: implications for the design of control programs. Proceedings of the National Academy of Sciences. 1997;94(1):338–342.

[2] Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. Nature. 2005;438(7066):355–359.

[3] Adam D, Wu P, Wong J, Lau E, Tsang T, Cauchemez S, et al. Clustering and superspreading potential of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections in Hong Kong. Research Square. 2020; Available from: https://doi.org/10.21203/rs.3.rs-29548/v1.

[4] Kumar S, Jha S, Rai SK. Significance of super spreader events in COVID-19. Indian Journal of Public Health. 2020;64(6):139.

[5] Liu Y, Eggo RM, Kucharski AJ. Secondary attack rate and superspreading events for SARS-CoV-2. The Lancet. 2020;395(10227):e47.

[6] Althouse BM, Wenger EA, Miller JC, Scarpino SV, Allard A, Hé-Dufresne L, et al.. Stochasticity and heterogeneity in the transmission dynamics of SARS-CoV-2; 2020. Available from: https://covid.idmod.org/data/Stochasticity_heterogeneity_transmission_dynamics_SARS-CoV-2.pdf.

[7] Hébert-Dufresne L, Althouse BM, Scarpino SV, Allard A. Beyond R0: Heterogeneity in secondary infections and probabilistic epidemic forecasting. 2020; Available from: https://www.medrxiv.org/content/10.1101/2020.02.10.20021725v2.

[8] Davies NG, Klepac P, Liu Y, Prem K, Jit M, Eggo RM, et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. medRxiv. 2020; Available from: https://www.medrxiv.org/content/medrxiv/early/2020/03/27/2020.03.24.20043018.

[9] World Health Organization. Modes of transmission of virus causing COVID-19: implications for IPC precaution recommendations: scientific brief, 27 March 2020. World Health Organization; 2020. Available from: https://apps.who.int/iris/bitstream/handle/10665/331601/WHO-2019-nCoV-Sci_Brief-Transmission_modes-2020.1-eng.pdf.

[10] Guo ZD, Wang ZY, Zhang SF, Li X, Li L, Li C, et al. Aerosol and surface distribution of severe

acute respiratory syndrome coronavirus 2 in hospital wards, Wuhan, China, 2020. Emerg Infect Dis. 2020;26(7):10–3201.

[11] Zhang Y, Li Y, Wang L, Li M, Zhou X. Evaluating transmission heterogeneity and super-spreading event of COVID-19 in a metropolis of China. International Journal of Environmental Research and Public Health. 2020;17(10):3705.

[12] Endo A, Abbott S, Kucharski AJ, Funk S, et al. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. Wellcome Open Research. 2020;5(67):67.

[13] Childs ML, Kain MP, Kirk D, Harris M, Couper L, Nova N, et al. The impact of long-term non-pharmaceutical interventions on COVID-19 epidemic dynamics and control. medRxiv. 2020; Available from: https://www.medrxiv.org/content/medrxiv/early/2020/05/06/2020.05.03.20089078.

[14] Anderson D, Watson R. On the spread of a disease with gamma distributed latent and infectious periods. Biometrika. 1980;67(1):191–198.

[15] Lloyd AL. Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and dynamics. Theoretical population biology. 2001;60(1):59–71.

[16] He D, Ionides EL, King AA. Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. Journal of the Royal Society Interface. 2010;7(43):271–283.

[17] Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Döner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science. 2020;368(6491).

[18] Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). Science. 2020;368(6490):489–493.

[19] McAloon CG, Collins A, Hunt K, Barber A, Byrne A, Butler F, et al. The incubation period of COVID-19: A rapid systematic review and meta-analysis of observational research. medRxiv. 2020; Available from: https://www.medrxiv.org/content/medrxiv/early/2020/04/28/2020.04.24.20073957.

[20] Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Annals of internal medicine. 2020;172(9):577–582.

[21] Wei WE, Li Z, Chiew CJ, Yong SE, Toh MP, Lee VJ. Presymptomatic Transmission of SARS-CoV-2—Singapore, January 23–March 16, 2020. CDC: (MMWR) Morbidity and Mortality Weekly Report. 2020;69(14):411.

[22] Zhang W, Cheng W, Luo L, Ma Y, Xu C, Qin P, et al. Secondary Transmission of Coronavirus Disease from Presymptomatic Persons, China. Emerging Infectious Diseases. 2020;26(8).

[23] He X, Lau EH, Wu P, Deng X, Wang J, Hao X, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nature medicine. 2020;26(5):672–675.

[24] Zhang W. Estimating the presymptomatic transmission of COVID19 using incubation period and serial interval data. medRxiv. 2020; Available from: https://www.medrxiv.org/content/medrxiv/early/2020/04/06/2020.04.02.20051318.

[25] Gao W, Li L. Advances on presymptomatic or asymptomatic carrier transmission of COVID-19. Zhonghua Liu Xing Bing Xue Za Zhi. 2020;41:485–488.

[26] Arons MM, Hatfield KM, Reddy SC, Kimball A, James A, Jacobs JR, et al. Presymptomatic SARS-CoV-2 infections and transmission in a skilled nursing facility. New England journal of medicine. 2020;382:2081–2090.

[27] Lee S, Meyler P, Mozel M, Tauh T, Merchant R. Asymptomatic carriage and transmission of SARS-CoV-2: What do we know? Canadian Journal of Anaesthesia. 2020;p. 1–7.

[28] Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, et al. Virological assessment of hospitalized patients with COVID-2019. Nature. 2020;581(7809):465–469.

[29] Sanche S, Lin YT, Chonggang Xu, Ethan Romero-Severson, Nick Hengartner, Ruian Ke. High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2. Emerging Infectious Diseases. 2020;26(7).

[30] To KKW, Tsang OTY, Leung WS, Tam AR, Wu TC, Lung DC, et al. Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. The Lancet Infectious Diseases. 2020;20:565–574.

[31] Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. The Lancet Infectious Diseases. 2020;20:569–677.

[32] Tindale L, Coombe M, Stockdale J, Garlock E, Lau WYV. Transmission interval estimates suggest pre-symptomatic spread of COVID-19. medRxiv. 2020; Available from: https://www.medrxiv.org/content/10.1101/2020.03.03.20029983v1.

[33] Rees EM, Nightingale ES, Jafari Y, Waterlow NR, Clifford S, Pearson CA, et al. COVID-19 length of hospital stay: a systematic review and data synthesis. medRxiv. 2020; Available from: https://www.medrxiv.org/content/10.1101/2020.04.30.20084780v3.

[34] Perkins A, Cavany SM, Moore SM, Oidtman RJ, Lerch A, Poterek M. Estimating unobserved SARS-CoV-2 infections in the United States. medRxiv. 2020; Available from: https://www.medrxiv.org/content/10.1101/2020.03.15.20036582v2.

[35] He D, Zhao S, Lin Q, Zhuang Z, Cao P, Wang MH, et al. The relative transmissibility of asymptomatic cases among close contacts. International Journal of Infectious Diseases. 2020;63(5):706–711.

[36] Ye F, Xu S, Rong Z, Xu R, Liu X, Deng P, et al. Delivery of infection from asymptomatic carriers of COVID-19 in a familial cluster. International Journal of Infectious Diseases. 2020;94:133–138.

[37] Hu Z, Song C, Xu C, Jin G, Chen Y, Xu X, et al. Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. Science China Life Sciences. 2020;63(5):706–711.

[38] Lavezzo E, Franchin E, Ciavarella C, Cuomo-Dannenburg G, Barzon L, Vecchio CD, et al. Suppression of COVID-19 outbreak in the municipality of Vo, Italy. medRxiv. 2020; Available at: https://www.medrxiv.org/content/10.1101/2020.04.17.20053157v1.

[39] Gudbjartsson DF, Helgason A, Jonsson H, Magnusson OT, Melsted P, Norddahl GL, et al. Spread of SARS-CoV-2 in the Icelandic population. New England Journal of Medicine. 2020;382:2302–2315.

[40] Li Lq, Huang T, Wang Yq, Wang Zp, Liang Y, Huang Tb, et al. COVID-19 patients' clinical characteristics, discharge rate, and fatality rate of meta-analysis. Journal of medical virology. 2020;92(6):577–583.

[41] Gaythorpe K, Imai N, Cuomo-Dannenburg G, Baguelin M, Bhatia S, Boonyasiri A, et al. Report 8: Symptom progression of COVID-19. Imperial College London; 2020. Available from: http://spiral.imperial.ac.uk/handle/10044/1/77344.

[42] County Population Totals 2010-2019. United States Census Bureau; 2019. Available at: https://www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-total.html.

[43] Shelter in Place Index: The Impact of Coronavirus on Human Movement. SafeGraph; 2020. Available at: https://www.safegraph.com/dashboard/covid19-shelter-in-place?s=US&d=06-04-2020&t=counties&m=index.

[44] Lloyd-Smith JO, Schreiber SJ, Getz WM. Moving beyond averages: Individual-level variation in disease transmission. In: Mathematical Studies on Human Disease Dynamics: Emerging Paradigms and Challenges: AMS-IMS-SIAM Joint Summer Research Conference on Modeling the Dynamics of Human Diseases: Emerging Paradigms and Challenges, July 17-21, 2005, Snowbird, Utah. vol. 410. American Mathematical Soc.; 2006. p. 235.

[45] Grantz K, Metcalf CJE, Lessler J. Dispersion vs. Control; 2020. Available at: https://hopkinsidd.github.io/nCoV-Sandbox/DispersionExploration.html.

[46] Burhenne S, Jacob D, Henze GP. Sampling based on Sobol'sequences for Monte Carlo techniques applied to building simulations. In: Proc. Int. Conf. Build. Simulat; 2011. p. 1816–1823.

[47] King AA, Nguyen D, Ionides EL. Statistical Inference for Partially Observed Markov Processes via the R Package pomp. Journal of Statistical Software. 2016;69(12):1–43.

[48] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2020. Available from: https://www.R-project.org/.

[49] Lu J, Gu J, Li K, Xu C, Su W, Lai Z, et al. COVID-19 outbreak associated with air conditioning in restaurant, Guangzhou, China, 2020. Emerging infectious diseases. 2020;26(7).

[50] Frieden TR, Lee CT. Identifying and interrupting superspreading events—implications for control of severe acute respiratory syndrome coronavirus 2. CDC Stacks. 2020; Available from: https://stacks.cdc.gov/view/cdc/88905.

[51] Shim E, Tariq A, Choi W, Lee Y, Chowell G. Transmission potential and severity of COVID-19 in South Korea. International Journal of Infectious Diseases. 2020;93:339–344.

[52] Dyal JW. COVID-19 Among Workers in Meat and Poultry Processing Facilities—19 States, April 2020. CDC: (MMWR) Morbidity and Mortality Weekly Report. 2020;69. Available from: https://www.cdc.gov/mmwr/volumes/69/wr/mm6918e3.htm.

[53] Stadnytskyi V, Bax CE, Bax A, Anfinrud P. The airborne lifetime of small speech droplets and their potential importance in SARS-CoV-2 transmission. Proceedings of the National Academy of Sciences. 2020;117(22):11875–11877.

[54] Benzell SG, Collis A, Nicolaides C. Rationing social contact during the COVID-19 pandemic: Transmission risk and social benefits of US locations. Proceedings of the National Academy of Sciences. 2020; Available from: https://www.pnas.org/content/early/2020/06/09/2008025117.

[55] Miller D, Martin MA, Harel N, Kustin T, Tirosh O, Meir M, et al. Full genome viral sequences inform patterns of SARS-CoV-2 spread into and within Israel. medRxiv. 2020; Available from: https://www.medrxiv.org/content/medrxiv/early/2020/05/22/2020.05.21.20104521.

[56] Endo A, for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group C, Abbott S, Kucharski AJ, Funk S. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. Wellcome Open Research. 2020;5.

[57] Okasha MK, Alqanoo I. Inference on the doubly truncated gamma distribution for lifetime data. Int J Math Stat Invent. 2014;2:1–17.

Supplemental Material for "Chopping the tail: how preventing super-spreading can help to maintain COVID-19 control"

Morgan P. Kain[1,2]*, Marissa L. Childs[3]*, Alexander D. Becker[1], Erin A. Mordecai[1]

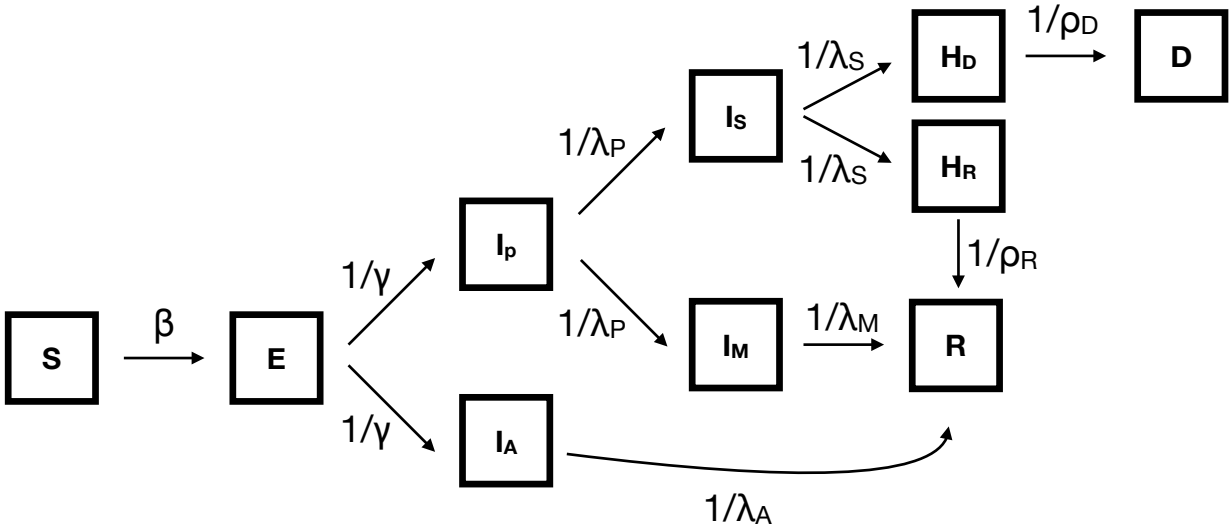
*Denotes equal authorship. Corresponding authors: morganpkain@gmail.com, marissac@stanford.edu

[1]Department of Biology, Stanford University, Stanford, CA, 94305, USA

[2]Natural Capital Project, Woods Institute for the Environment, Stanford University, Stanford, CA 94305, USA

[3]Emmett Interdisciplinary Program in Environment and Resources, Stanford University, Stanford, CA, 94305, USA

Figures



State variables

- **S** = Susceptible
- **E** = Exposed
- **I_P** = Pre-symptomatic
- **I_A** = Asymptomatic
- **I_S** = Symptomatic, severe case
- **I_M** = Symptomatic, mild case
- **H_D** = Hospitalized, eventual death
- **H_R** = Hospitalized, eventual recovery
- **R** = Recovered
- **D** = Dead

Transition Rates

- **β** = Transmission rate
- **γ** = Preinfectious period
- **λ_A** = Asymptomatic infectious period
- **λ_P** = Presymptomatic infectious period
- **λ_S** = Severe infectious period until hospitalization
- **λ_M** = Mild infectious period
- **ρ_D** = Hospitalization period until death
- **ρ_R** = Hospitalizations period until recovery

Figure S1: Epidemiological model box diagram

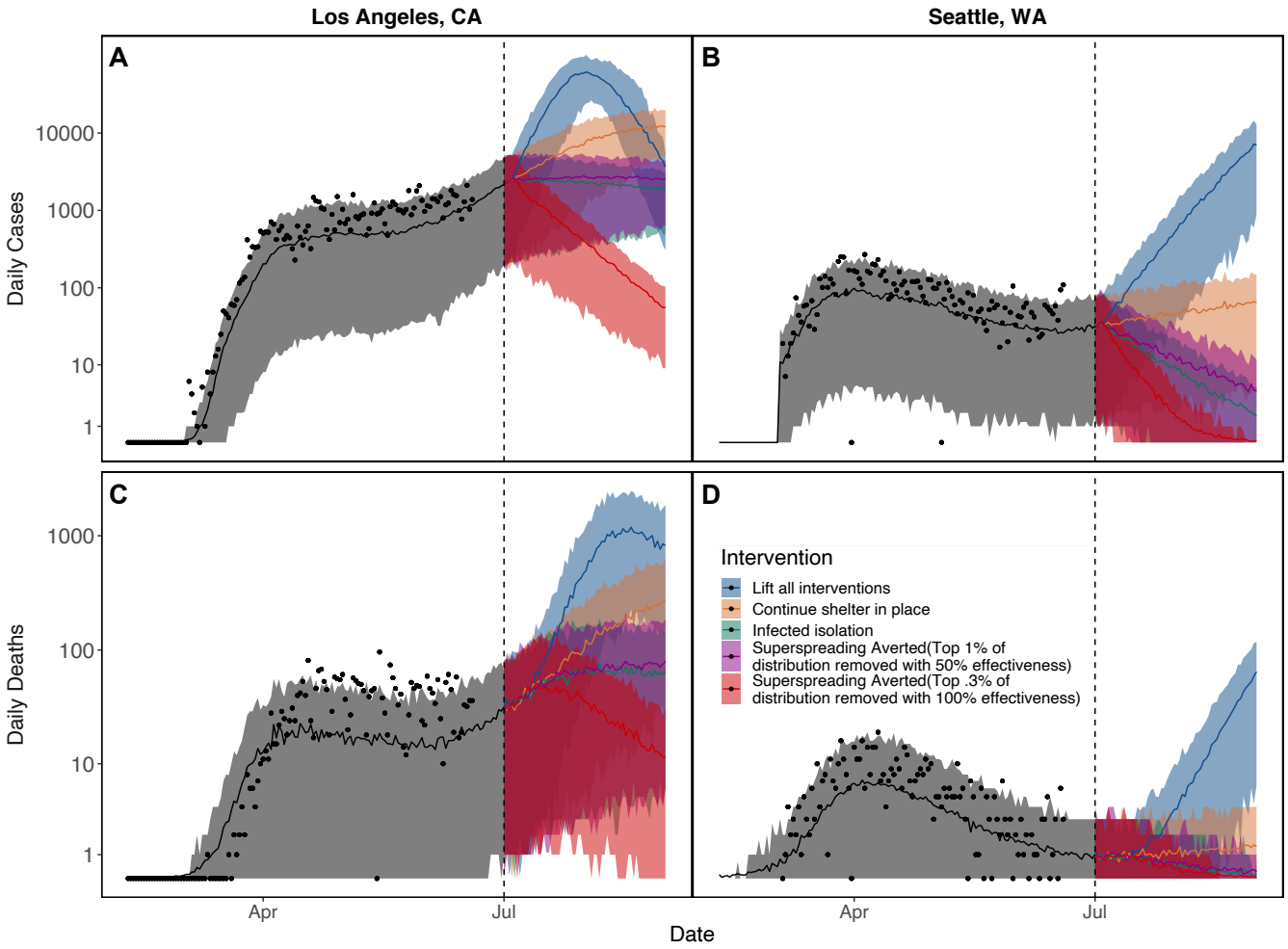


Figure S2: Many other truncation interventions are viable alternatives to the top 0.5% with 75% efficiency presented in the main text including: truncating the top 1% with 50% efficiency (purple) and top 0.3% with 100% efficiency (red). Bands show 95% CI on stochastic simulations of daily cases and deaths for the single maximum likelihood estimate. Dates range from February 2020 to October 2020.

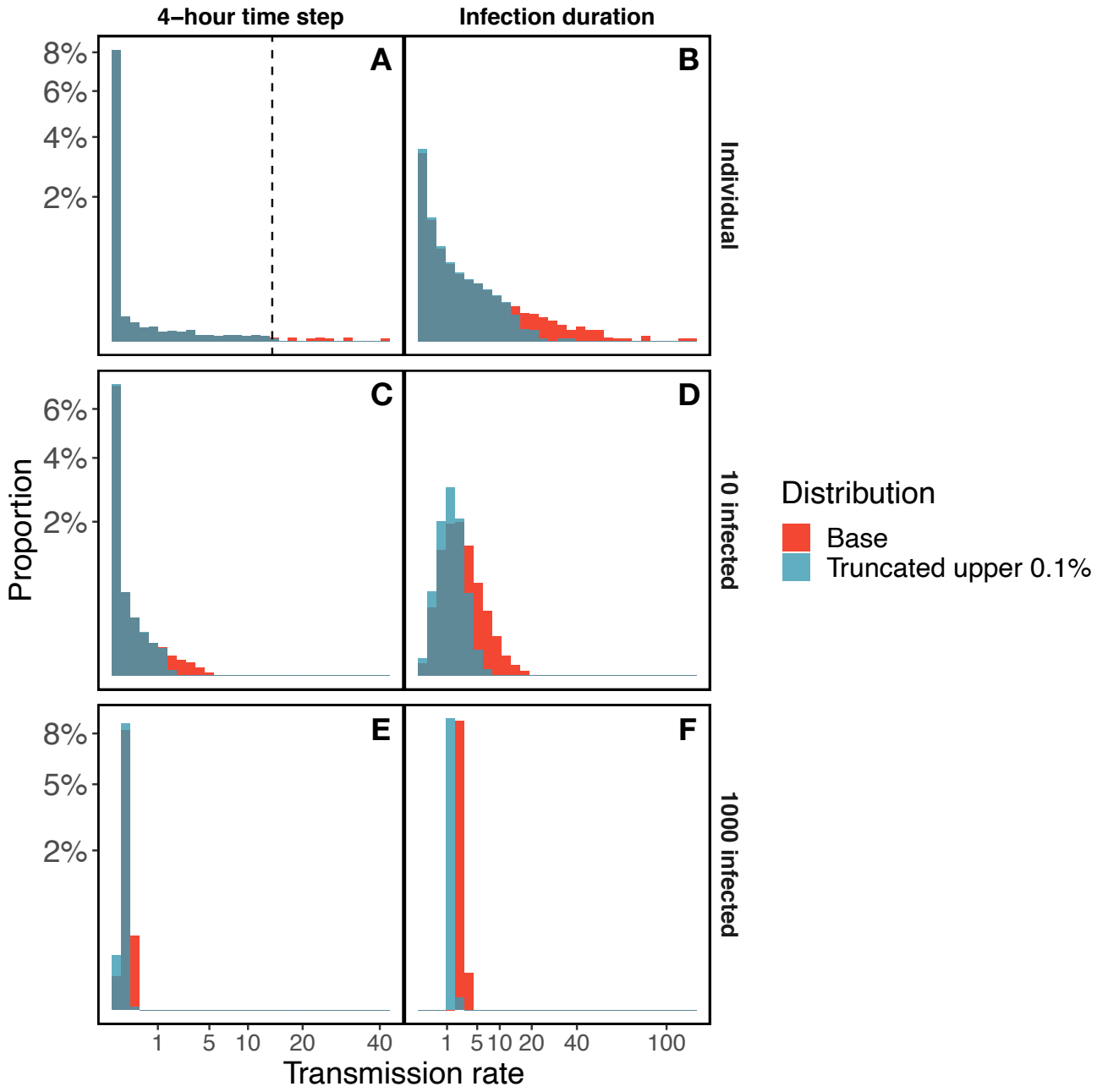


Figure S3: Expanded view of truncating the upper 0.1% of the individual level time step transmission rate distribution (π) at a four-hour time step (A). This truncation leads to a reduction of the mean and variance for an individual's infectious period reproduction potential (B). As the number of infected individuals in the population increases from 10 (C, D) to 1000 (E, F), the variance in decreases in both the population-level average transmission rate during each 4-hour period (C, E) and over the lifetime of those infected (D, F).

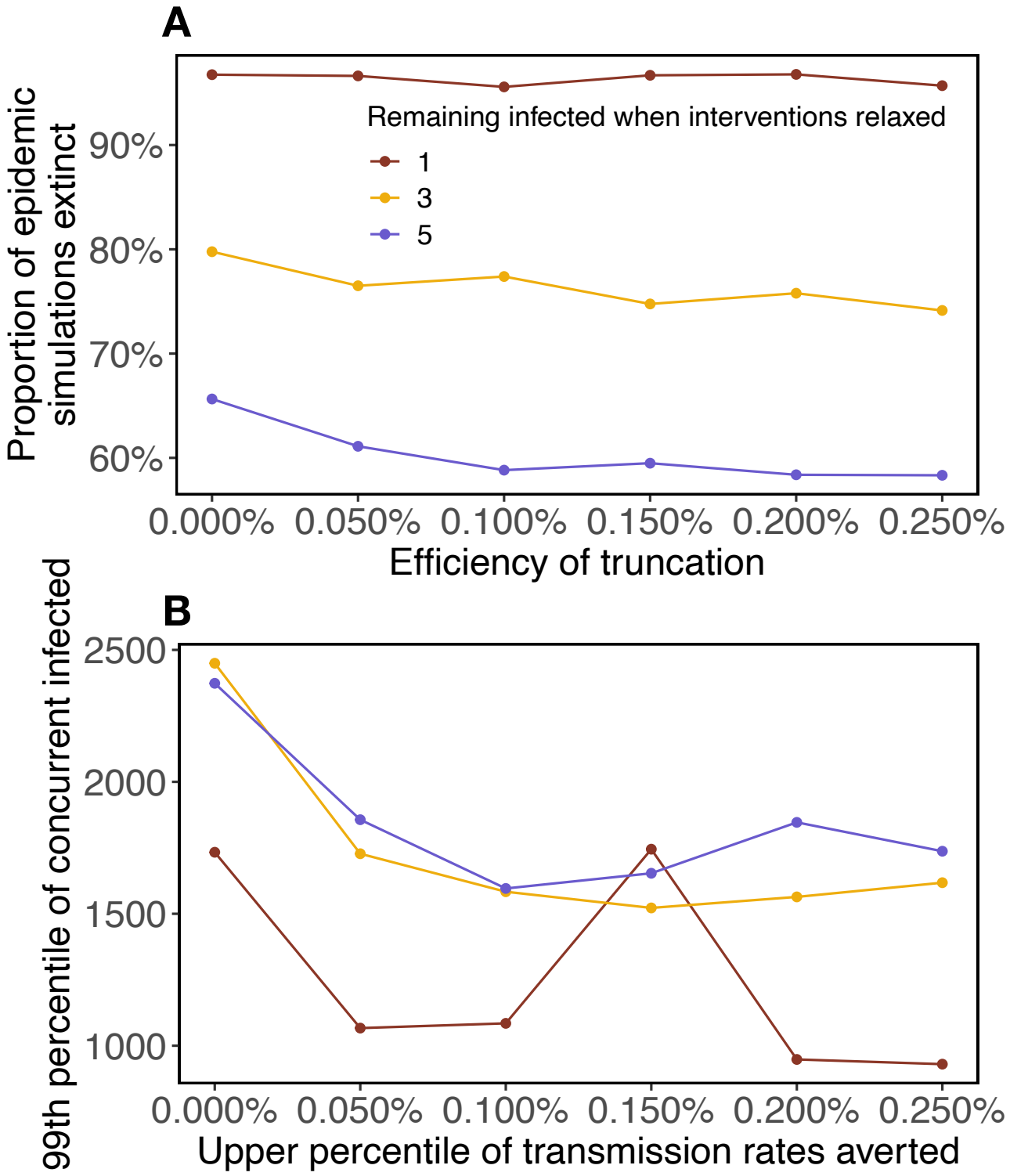


Figure S4: The proportion of epidemic simulations that went extinct (A) and the upper 99th percentile of the number concurrent infected after 42 days (B) for the resurgent simulations among 5000 total simulations for increasing truncation proportions of π. Shelter-in-place is scaled so that transmission rate at the time of intervention relaxation is identical across intervention scenarios and would result in $\mathcal{R}_E = 2$ in a fully susceptible population.

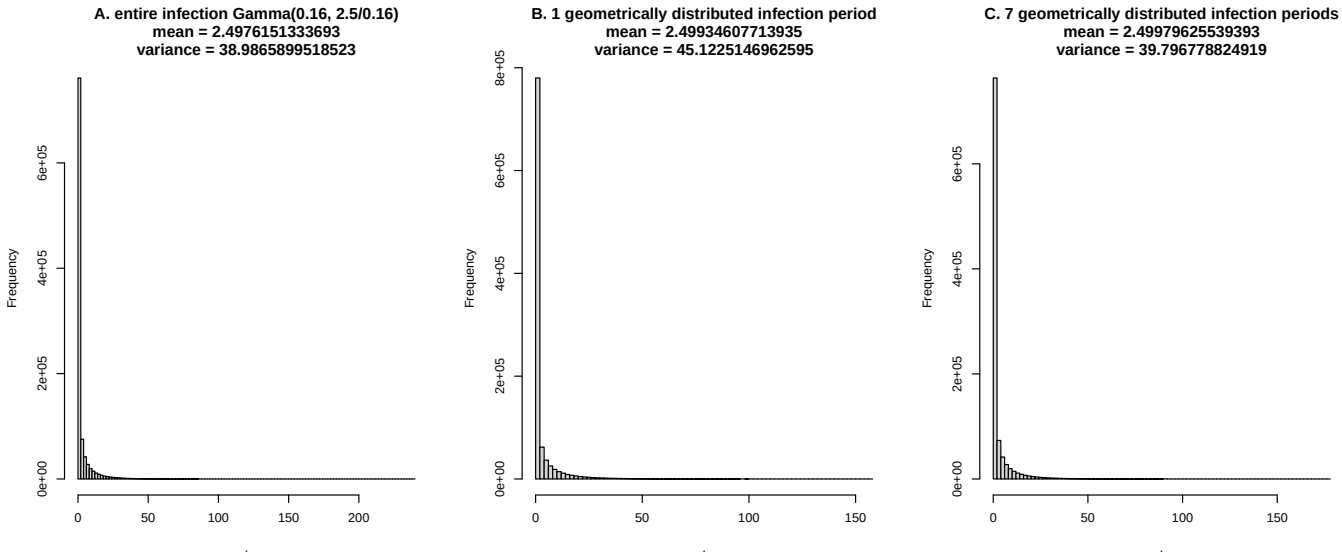


Figure S5: The distribution of individual lifetime reproduction (R) when modeled as a Gamma distribution with a mean of 2.5 and a scale of 0.16 (A). This distribution implicitly assumes a constant infectious duration. Using a geometrically distributed infectious period with only one period ("box"), and a time period of 4 hours result in an increase in the variance of the individual reproductive distribution relative to assuming a constant infectious period (B). Breaking the infectious period into 7 sub-stages (boxes) reduces the variance, though the variance remains marginally higher than when assuming a constant infectious period (C).