## S1 Appendix

## An integrated risk and epidemiological model to estimate risk-stratified COVID-19 outcomes for Los Angeles County: March 1, 2020 - March 1, 2021

Abigail L. Horn $^1$ , Lai Jiang $^1$ , Faith Washburn $^2$ , Emil Hvitfeldt $^1$ , Kayla de la Haye $^1$ , William Nicholas $^2,$ Paul Simon<sup>2</sup>, Maryann Pentz<sup>1</sup>, Wendy Cozen<sup>1</sup>, Neeraj  $\mathbf{Sood}^{3,4}$ , & David V. Conti<sup>1</sup>

<sup>1</sup> Department of Preventive Medicine Keck School of Medicine University of Southern California

<sup>2</sup> Los Angeles County Department of Public Health 313 N Figueroa St, Los Angeles, CA 90012

<sup>3</sup> Sol Price School of Public Policy University of Southern California

<sup>4</sup> Schaeffer Center for Health Policy and Economics University of Southern California

### **Contents**





# <span id="page-2-0"></span>Part I SEIR+RISK Stochastic Epidemic Model

### <span id="page-2-1"></span>**1 Epidemic model variables and parameters**

### <span id="page-2-2"></span>**1.1 State variables**



Table 1. State variables.

### <span id="page-2-6"></span><span id="page-2-3"></span>**1.2 Fixed parameters (Transition times)**



Table 2. Transition time parameters, values, and sources. These parameters are modeled as fixed values.

### <span id="page-2-7"></span><span id="page-2-4"></span>**1.3 Estimated parameters (transition branching probabilities)**



<span id="page-2-5"></span>Table 3. Transition branching probabilities. These parameters are estimated.

#### **1.4 Deterministic differential equation model**

The "skeleton" [\[5\]](#page-31-4) of the compartmental model is described by the following set of coupled ordinary differential equations (ODE) describing the transitions of individuals between the 8 compartments across time:

$$
dS/dt = -\beta_t S(I + A) \tag{1}
$$

$$
dE/dt = \beta_t S(I + A) - \frac{1}{d_{EI}} E
$$
\n
$$
(2)
$$

$$
dA/dt = \frac{1-r_t}{d_{EI}}E - \frac{1}{d_{IR}}A
$$
  
\n
$$
dI/dt = \frac{r_t}{d_{EI}}E - (\frac{\alpha_t}{d_{IH}} + \frac{1-\alpha_t}{d_{IR}})I
$$
\n(3)

$$
dH/dt = \frac{\alpha_t}{d_{IH}} I - \left(\frac{\kappa_t}{d_{HQ}} + \frac{1-\kappa_t}{d_{HR}}\right)H\tag{5}
$$

$$
dQ/dt = \frac{\kappa_i}{d_{HQ}}H - \left(\frac{\delta_i}{d_{QP}} + \frac{1-\delta_i}{d_{QR}}\right)Q\tag{6}
$$

$$
dD/dt = \frac{\delta_i}{d_{QD}}Q\tag{7}
$$

$$
dR/dt = \frac{1-\alpha_i}{d_{IR}}I + \frac{1-\kappa_i}{d_{HR}}H + \frac{1-\delta_i}{d_{QR}}Q + \frac{1}{d_{IR}}A
$$
\n
$$
(8)
$$

With the total population size:

$$
P = S + E + A + I + H + Q + D + R \tag{9}
$$

#### <span id="page-3-0"></span>**1.5 Stochastic dynamical state model**

We develop a stochastic discrete-time model of the dynamical state system, which provides an effective framework for parameter estimation and for generating confidence bounds reflecting stochasticity of the disease process in model projections.

We model the number of individuals leaving any of the classes by all available routes over a particular time interval by a set of coupled multinomial counting processes,  $N_{X_i\to X_j}(t)$  for all transition pairs  $(i, j)$ , with random transition rates. Transitions of individuals from one to the next stage of the disease are seen as stochastic movements between the corresponding population compartments at transition rates with random fluctuations. At each period an individual either stays or moves on to the next compartment. It has been proven that in the limit as the time interval ∆*t* → 0, the following discrete-time approximation process fulfills the properties defining a continuous time Markov process [\[6\]](#page-31-5). We do not model this continuous-time Markov process explicitly and operate only with its discrete approximation.

In general in this process, the random variable for the number of individuals leaving class  $X_i$  and going to  $X_j$  and  $X_k$  over time interval  $[t, t + \Delta t)$ ,  $(\Delta N_{X_i \to X_j}(t), \Delta N_{X_i \to X_k}(t))$ , has the multinomial distribution

$$
(\Delta N_{X_i \to X_j}(t), \Delta N_{X_i \to X_k}(t)) \sim \text{Multinomial}(X_i(t); p_{X_i \to X_j}(t), p_{X_i \to X_k}(t)),
$$

for states *X<sup>j</sup>* and *X<sup>k</sup>* reachable from state *X<sup>i</sup>* . The multinomial distributions result from summation over individual independent and identical Bernoulli trials for all members of each compartment [\[7,](#page-31-6) [8\]](#page-31-7). If there is only one possible transition direction out of  $X_i$ , then the distribution above is given by a binomial distribution.

The amount of time spent in a compartment is described by a Poisson process. The time length that an individual spends in a compartment is thus exponentially distributed with some compartment-specific rate  $\lambda_i(t) = \lambda_{i}(t) + \lambda_{i}(t)$ , where  $\lambda_{i}(t)$  and  $\lambda_{ik}(t)$  are the rates to go from  $X_i$  to  $X_j$  and  $X_k$ , respectively. The probability of extending the stay by a further period of length  $\Delta t$  is exp( $\lambda_i(t)\Delta t$ ) and the probability of leaving is  $1 - \exp(\lambda_i(t)\Delta t)$  [\[7\]](#page-31-6). Given there is a transition out of  $X_i$ , the conditional probability it is to  $X_j$  is  $\frac{\lambda_{ij}}{\lambda_{ij}(t)+\lambda_{ik}(t)}$ . Therefore the probability of leaving  $X_i$  and going to  $X_j$  in  $[t, t+\Delta t)$ ,  $p_{X_i\rightarrow X_j}$ , is found as

$$
p_{X_i \to X_j} = (1 - \exp\left(-(\lambda_{ij}(t) + \lambda_{ij}(t))\Delta t\right)) \frac{\lambda_{ij}}{\lambda_{ij}(t) + \lambda_{ik}(t)}.
$$

In the following we proceed to define the increments over  $[t, t + \Delta t)$  of all counting process,  $\Delta N_{X_i \to X_j}(t)$ , defining the transitions between all states in the process.

### <span id="page-3-1"></span>*1.5.1 Counting processes for transitions between states*

**New latent infections (***S* **to** *E* **transitions)** New infections occur from direct or indirect interactions between susceptible (*S*) and infected (*I* or *A*) individuals, and arise first in the exposed (*E*) compartment. The rate at which susceptible individuals become infected is described by the *force of infection*, Λ, which combines the rate of infections per day per infected individual  $(\beta_t)$  across all infected individuals by multiplying with the overall fraction of infected individuals at time *t*,  $\frac{I(t)+A(t)}{P}$  $\frac{H(H)}{P}$ , where *P* is equal to the total population size;  $\Lambda(t) = \beta_t \frac{I(t) + A(t)}{P}$  $\frac{+A(I)}{P}$ .

We assume that the corresponding probability at time *t* that a susceptible individual becomes infected and leaves the *E* compartment in one time unit  $\Delta t$  is equal to  $p_{S\rightarrow}(t) = 1 - e^{-\beta_t \frac{I(t) + A(t)}{P} \Delta t}$ .

Then, the overall number of individuals becoming infected and moving from the *S* to the *E* compartment in one time step,  $\Delta N_{S\rightarrow E}$ , is distributed as a binomial distribution with one draw for each of *S* individuals, each with probability  $p_{S\rightarrow E}(t)$ , i.e.

$$
\Delta N_{S\rightarrow E}(t) \sim \text{Binomial}(S(t), p_{S\rightarrow E}(t)).
$$

**Exposed, latent infections (***E***) becoming infectious (***I* **and** *A***)** Exposed and latent individuals will become infectious and transition to the observed infected state (*I*) or the unobserved infected (*A*) state.

The total rate of departures from *E* is equal to the sum of the rate of departures to *I*,  $\frac{r_t}{d_{EI}}$ , and the rate of departures to *A*,  $\frac{1-r_t}{d_{EI}}$ , which is simply  $\frac{1}{d_{EI}}$ . In each time step  $\Delta t$ , each exposed individual has a time-independent probability of becoming infectious based on this rate,  $p_{E\to}(t) = 1 - e^{-\frac{1}{d_{EI}}\Delta t}$ . The relative probability that a departure out of *E* is to *I* is equal to  $r_t$ , and to *A* is 1−*r<sup>t</sup>* . With these probabilities we can find the number of individuals leaving *E* and going to *I* and *A* over time interval  $[t, t + \Delta t)$ ,  $(\Delta N_{E \rightarrow I}(t), \Delta N_{E \rightarrow A}(t))$ , as

 $(\Delta N_{E\rightarrow I}(t), \Delta N_{E\rightarrow A}(t)) \sim \text{Multinomial}(E(t), (r_t)p_{E\rightarrow I}(t), (1-r_t)p_{E\rightarrow I}(t)).$ 

**Infection (***I***) to Hospitalization (***H***) or Recovery (***R***)** We model transitions for infectious individuals to recovery (*R*) or hospitalization (*H*).

The total rate of departures out of *I* is equal to the sum of the rate of departures to *H*,  $\frac{\alpha_t}{d_H}$ , and the rate of departures to *R*,  $\frac{1-\alpha_i}{d_{IR}}$ . The probability that an infected individual transitions out of *I* in each time step is then found as  $p_{I\rightarrow} = 1 - e^{-(\frac{\alpha_i}{d_{IH}} + \frac{1-\alpha_i}{d_{IR}})\Delta t}$ . The relative probability that a departure is to *H* is equal to  $\frac{\frac{\alpha_H}{d_H}}{\frac{d_H}{d_H} + \frac{1-\alpha_H}{d_R}}$ , and to *R* is  $\frac{\frac{1-\alpha_t}{d_{IR}}}{\frac{\alpha_t}{d_{IH}} + \frac{1-\alpha_t}{d_{IR}}}$ . The number of infected individuals that become hospitalized in each time step, ∆*NI*→*H*, and that recover, ∆*NI*→*R*, are then described by the multinomial distribution,

$$
(\Delta N_{I\rightarrow H}(t), \Delta N_{I\rightarrow R}(t)) \sim \text{Multinomial}(I(t), \frac{\frac{\alpha_I}{d_{IH}}}{\frac{\alpha_I}{d_{IH}} + \frac{1-\alpha_I}{d_{IR}}} p_{I\rightarrow}, \frac{\frac{1-\alpha_I}{d_{IR}}}{\frac{\alpha_I}{d_{IH}} + \frac{1-\alpha_I}{d_{IR}}} p_{I\rightarrow}).
$$

**Transitions out of hospital (***H***) and ICU (***Q***)** We model transitions out of hospital *H* and out of the ICU *Q* in the same way as transitions out of *I*, each of which go to two compartments.

The probability of transition in one time step out of *H* is

$$
p_{H\rightarrow}(t)=1-e^{-(\frac{\kappa_t}{d_{HQ}}+\frac{1-\kappa_t}{d_{HR}})\Delta t}.
$$

The random variable for the number of hospitalized individuals that go to the ICU,  $\Delta N_{H\to Q}(t)$ , and that recover,  $N_{H\to R}(t)$ , is

$$
(\Delta N_{H\rightarrow Q}(t),\Delta N_{H\rightarrow R}(t))\sim \mathrm{Multinomial}(H(t),\frac{\frac{\delta H_O}{d_{HQ}}}{\frac{\kappa_I}{d_{HQ}}+\frac{1-\kappa}{d_{HR}}}p_{H\rightarrow}(t),\frac{\frac{1-\kappa_I}{d_{HR}}}{\frac{\kappa_I}{d_{HQ}}+\frac{1-\kappa_I}{d_{HR}}}p_{H\rightarrow}(t)).
$$

The total rate of transitions out of the ICU, to death (*D*) or recovery, is

$$
p_{Q\rightarrow}(t) = 1 - e^{-(\frac{\delta_t}{d_{QD}} + \frac{1-\delta_t}{d_{QR}})\Delta t}
$$

The random variable for the number of individuals in ICU that die,  $\Delta N_{Q\to D}(t)$ , and that recover,  $N_{Q\to R}(t)$ , is found as

$$
(\Delta N_{Q\to D}(t), \Delta N_{Q\to R}(t)) \sim \text{Multinomial}(Q(t), \frac{\frac{\delta_t}{d_{QD}}}{\frac{\delta_t}{d_{QR}} + \frac{1-\delta_t}{d_{QR}}}p_{Q\to}(t), \frac{\frac{1-\delta_t}{d_{QR}}}{\frac{\delta_t}{d_{QD}} + \frac{1-\delta_t}{d_{QR}}}p_{Q\to}(t)).
$$

**Transitions out of unobserved infected state (***A***)** We assume all unobserved infected cases will recover, at the rate  $\frac{1}{d_{IR}}$ . We model the probability of recovery in one time step as  $p_{A\to}(t) = 1 - e^{-\frac{1}{d_R}\Delta t}$ . The number of recoveries moving from *A* to *R* in each time step is then found as

<span id="page-4-0"></span>
$$
\Delta N_{A\to R}(t) \sim \text{Binomial}(A(t), p_{A\to}(t)).
$$

#### *1.5.2 System of equations for state variables and advancing the model*

The counting process random variables are coupled to the state variables via the following identities:

$$
\Delta S(t) = -\Delta N_{S \to E}(t)
$$
  
\n
$$
\Delta E(t) = \Delta N_{S \to E}(t) - \Delta N_{E \to I}(t) - \Delta N_{E \to A}(t)
$$
  
\n
$$
\Delta I(t) = \Delta N_{E \to I}(t) - \Delta N_{I \to H}(t) - \Delta N_{I \to R}(t)
$$
  
\n
$$
\Delta A(t) = \Delta N_{E \to A}(t) - \Delta N_{A \to R}(t)
$$
  
\n
$$
\Delta H(t) = \Delta N_{I \to H}(t) - \Delta N_{H \to Q}(t) - \Delta N_{H \to R}(t)
$$
  
\n
$$
\Delta Q(t) = \Delta N_{H \to Q}(t) - \Delta N_{Q \to D}(t) - \Delta N_{Q \to R}(t)
$$
  
\n
$$
\Delta D(t) = \Delta N_{Q \to D}(t)
$$
  
\n
$$
\Delta R(t) = \Delta N_{I \to R}(t) + \Delta N_{H \to R}(t) + \Delta N_{Q \to R}(t).
$$

To simulate from this system we employ an Euler numerical scheme for Markov process compartment models with stochastic rates [\[6\]](#page-31-5). We sample the transitions from the counting processes ∆*NXi*→*X<sup>j</sup>* (*t*), and update each state variable, representing the number of individuals in each compartment at time *t*, to reflect these counts.

#### <span id="page-5-0"></span>*1.5.3 Cumulative and daily counts*

The dynamics of the stochastic model are advanced by updating state variables representing the "current census" of individuals in each compartment *X<sup>i</sup>* . Observed COVID-19 infection data, however, is available as daily or cumulative counts. We therefore also keep track of auxiliary count variables representing the daily new counts or cumulative infection counts (observed data described in Section [2.2\)](#page-9-1).

For cumulative counts we have, based on the counting processes for transitions between compartments already defined:

$$
I_{cum}(t+1) = I_{cum}(t) + \Delta N_{E \to I}(t)
$$

where the *Icum* represents the cumulative version of each state variable *I*. Daily count *Inew* can be worked out from the cumulative counts as  $I_{new}(t+1) = I_{cum}(t+1) - I_{cum}(t)$ .

### <span id="page-5-1"></span>**1.6** Solving for the basic reproduction number,  $R_0$

The basic reproductive number, *R*0, is a dimensionless parameter defined as the average number of secondary infections produced by a single infected individual during that individual's entire period of infection, in a completely susceptible population and no interventions to reduce the spread [\[9\]](#page-31-8). While not a parameter of the system of state variables, *R*0 is a parameter of interest because it provides a measure of the likelihood that an epidemic will occur within a population.

In a deterministic model, *R*0 represents an epidemic threshold for which values of *R*0 < 1 indicate a lack of disease spread, and values of  $R0 > 1$  are consistent with epidemic spread. For a stochastic model, if  $R0 > 1$  the probability that an epidemic will develop is less than 1, and if  $R0 < 1$ , extinction of the infection will occur with a probability less than 1. On average, however, no epidemic will occur if  $R0 \le 1$  [\[10\]](#page-31-9). Therefore, estimates of R0 provide clearly communicable insight regarding the degree of intensity of interventions required to achieve control [\[9\]](#page-31-8).

For a simple SIR model, *R*0 is a simple expression equal to the product of the number of infections per time for an infected individual, i.e. the transmission rate  $\beta$ , and the average duration of the infectious period, e.g.  $d_I$ , that is,

$$
R0 = \frac{\text{infections}}{\text{time}} \times \frac{\text{time}}{\text{infection}} = \beta d_I.
$$

Because our model involves multiple infectious classes with different rates of infection duration, the average number of individuals an infected individual infects is a more complex expression. We use the *Next Generation Matrix* (NGM) [\[11\]](#page-31-10) approach to find the expression for *R*0. The NGM approach involves computing *R*0 by linearizing the ODE model in Section [1.4](#page-2-5) around the infection-free steady state and identifying conditions that guarantee increases in infected states [\[11\]](#page-31-10). We focus on the *infection subsystem*, i.e. the subset of ODEs that describe the production of new infections and changes in the states of already existing infections only, which is

$$
dE/dt = \beta_0 S(I + A) - \frac{1}{d_{EI}} E
$$
  
\n
$$
dA/dt = \frac{1-r_0}{d_{EI}} E - \frac{1}{d_{IR}} A
$$
  
\n
$$
dI/dt = \frac{r_0}{d_{EI}} E - (\frac{\alpha_0}{d_{IH}} + \frac{1-\alpha_0}{d_{IR}})I,
$$

where  $\beta_0$ ,  $r_0$ , and  $\alpha_0$  represent the initial estimated values of these parameters.

Setting  $X = (\frac{dE}{dt}, \frac{d\vec{A}}{dt}, \frac{dI}{dt})^T$ , we write the linearized infection subsystem in the form

$$
\frac{dX}{dt} = (T + \Sigma)X,
$$

where *T* corresponds to all epidemiological events leading to new infections, and Σ corresponds to all epidemiological events relating to transitions in or out of these infected states. We write the new infection matrix *T* as

$$
T = \left(\begin{array}{ccc} 0 & \beta_0 & \beta_0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right)
$$

and the new transmission matrix  $\Sigma$  as

$$
\Sigma = \begin{pmatrix} -\frac{1}{d_{EI}} & 0 & 0 \\ \frac{r_0}{d_{EI}} & -(\frac{\alpha_0}{d_{IH}} + \frac{1-\alpha_0}{d_{IR}}) & 0 \\ \frac{1-r_0}{d_{EI}} & 0 & -\frac{1}{d_{IR}} \end{pmatrix}.
$$

Diekmann et al. [\[11\]](#page-31-10) show that *R*0 can be found as the dominant eigenvalue of the next generation matrix *K<sup>L</sup>* = −*T*Σ −1 , or equivalently, the dominant eigenvalue of  $K = E^T \Sigma^{-1} E$ , where *E* is an auxiliary matrix whose columns consist of unit vectors relating to the non-zero rows of *T*. *E*, which in this case is equal to  $E^{\dagger} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$ , is included to reduce dimensionality.

We find  $\Sigma^{-1}$  as

$$
\Sigma^{-1} = \begin{pmatrix} d_{EI} & 0 & 0 \\ \frac{-r_0}{\frac{a_0}{d_{IH}} + \frac{1 - \alpha_0}{d_{IR}}} & \frac{1}{\frac{a_0}{d_{IH}} + \frac{1 - \alpha_0}{d_{IR}}} & 0 \\ -(1 - r_0)d_{IR} & 0 & d_{IR} \end{pmatrix}.
$$

We have  $ETT = \begin{bmatrix} 0 & \beta_0 & \beta_0 \end{bmatrix}$ , and  $\Sigma^{-1}E = \begin{bmatrix} d_{EI} & \frac{-r_0}{\alpha_0 + 1 - \alpha_0} \\ \frac{d_{IH} + d_{IR}}{\alpha_{IR}} & \frac{1}{\alpha_{IR}} \end{bmatrix}$  $-(1-r_0)d_{IR}$ <sup>T</sup>. We can then find the next generation matrix as,

$$
K = -E^{\mathsf{T}} T \Sigma^{-1} E = \beta_0 \left[ \frac{-r_0}{\frac{\alpha_0}{d_{IH}} + \frac{1 - \alpha_0}{d_{IR}}} + (1 - r_0) d_{IR} \right].
$$

*R*0 is the dominant eigenvalue of *K*, which is simply equal to *K* in this case. Therefore,

$$
R0 = \beta_0 \left[ \frac{r_0}{\frac{\alpha_0}{d_{IH}} + \frac{1-\alpha_0}{d_{IR}}} + (1-r_0)d_{IR} \right].
$$

We note that the NGM approach for deriving *R*0 is based on the deterministic ODE equations, while we use this parameter with the stochastic model. We justify deriving it from the deterministic model because for large population sizes *N*, the mean of the stochastic equations will converge on the deterministic model equations [\[10,](#page-31-9) [12\]](#page-31-11).

Given that *R*0 is the value of the reproductive number at the beginning of the pandemic period before modifications, we define the time-varying reproductive number as *R<sup>t</sup>* .

Additionally, the reproduction number will decrease as natural immunity increases, reflecting the fact that as the proportion of susceptible individuals decreases  $(S(t)/P)$ , disease transmission slows [\[13\]](#page-31-12). We therefore define the time-varying effective reproductive number, accounting for both the community transmission reduction factor and the decreasing proportion of susceptible individuals, as  $R_{eff,t} = R_t \frac{S(t)}{N}$  $\frac{(t)}{N}$ .

### <span id="page-6-0"></span>**2 Parameter Estimation**

Transmission rate parameters we model as fixed values taken from the literature (Table [2\)](#page-2-6). The remaining five model parameters,  $\{\beta_t, r_t, \alpha_t, \kappa_t, \delta_t\}$  (Table [3\)](#page-2-7), must be estimated with data. To this set of five parameters we add an additional parameter,  $t_0$ , representing the starting time of the epidemic in LAC; that is, the very first case of COVID-19, which may have been unobserved. This parameter links the epidemic model to the observed count data based on calendar dates. In addition, because

prior information from existing studies on the reproductive number *R<sup>t</sup>* is more readily available than for β*<sup>t</sup>* , we estimate *R<sup>t</sup>* from data and then use equation [1.6](#page-5-1) to obtain  $\beta_t$  from  $R_t$ . Therefore the estimated parameter set is  $\theta = \{R_t, r_t, \alpha_t, \kappa_t, \delta_t, t_0\}$ .

We use a two-step process to define unimodal posterior distributions and achieve model convergence, first performing a broad grid search to define a narrow range for each parameter, and second using Approximate Bayesian Computation (ABC) techniques with prior distributions set by the ranges defined by the grid search to estimate the final posterior distribution for each parameter. External data sources were used to specify the limits of the parameter space for the grid search. In the following we describe the data sources used to inform the grid search space for each parameter, and the resulting prior parameter distributions chosen to use in ABC.

#### <span id="page-7-0"></span>**2.1 Specifying prior parameter distributions**

#### <span id="page-7-1"></span>*2.1.1 Basic Reproduction Number, Rt*

We base the prior distribution for *R*0 on values for *R*0 estimated from other published studies on COVID-19. A review of *R*0 values estimated for COVID-19 across studies involving different locations, time periods, and modeling approaches reported values in the range of [1.40-6.49], with a mean of 3.28 [\[14\]](#page-31-13). Values for *R*0 in SEIR models are also higher than those in *SIR* models, as has been demonstrated by previous researchers [\[15\]](#page-31-14); for example, Bertozzi et al. [\[16\]](#page-31-15) demonstrate an estimated *R*0 for California of 2.4 using a SIR model and 4.9 using an SEIR model. After performing a broad grid search across a parameter space from [3,5], we model a prior distribution for *R*0 in ABC as a normal distribution centered around 3.5, with a standard deviation chosen such that 95% of the area of the distribution is within approximately  $\pm 0.25$  of the mean. The modeled prior distribution for *R*0 is therefore:

 $R0 \sim N(3.5, 0.2)$ 

### <span id="page-7-2"></span>*2.1.2 Time-varying Reproduction Number after Interventions, Rt*

Geolocation trace data from smartphones, i.e. mobility data, can be used estimate changes in distances travelled, commuting, time spent outside of home, visits to specific types of venues, and encounter rates in the community. These data sources, which represent contact rates in the population, can be used as a proxy to estimate changes in contact rate and equivalently in β*<sup>t</sup>* and  $R_t$  over time [\[17\]](#page-31-16).

We use mobility data to inform the magnitude and the timing of inflection points in the factor reduction in *R<sup>t</sup>* . We reference mobility data for LAC from a dashboard provided by Unacast, a company that builds and maintains anonymized mobility datasets based on geolocation data aggregated from millions of users across the U.S. [\[18\]](#page-31-17). Unacast provides three metrics: the *Change in Average Mobility (Based on Distance Traveled)*, the *Change in Non-Essential Visits*, and the *Difference in Encounter Density*. We reference the metrics *Change in Average Mobility (Based on Distance Traveled)* and *Difference in Encounter Density* (Figure [1\)](#page-8-0) because these both inform the changes and timing of changes in contact rates, in different ways [\[17\]](#page-31-16).

Across these two metrics we observe the following trends in  $\mu(t)$ : (i) a sharp initial descent from the original contact rate beginning around March 12 2020 (a few days before county-wide quarantine policy was announced on March 15), and leveling out around March 28; (ii) a steady plateau during the initial period of the lockdown from around March 28 - April 26; (iii) a steady incline in contact rates from April 27 - May 15 (May 8th being the end of the lockdown); (iv) a lower plateau beginning August 15 (when schools opened), and (v) an increase beginning October 15.

As noted in the main text, an increase beginning in October and throughout the third wave of the epidemic in LAC is not visible on the mobility metrics, but is required for the model to achieve accurate fits to data; this likely reflects the fact that the infection rate is dependent on both the contact rate and the probability of transmission given contact and may include decreasing NPI adherence, more time spend indoors, and other conditions that may increase transmissiblility of the virus. Thus, for dates after October 15, 2020, we stopped using mobility data to inform the grid search space for *R<sup>t</sup>* and instead set this equal to  $1 < R_t < R0$ , for dates corresponding to increasing infection trends.

Based on the results from the grid search step, we develop a distribution with the change points in time given in Table [4.](#page-8-1) A piecewise linear function is interpolated between dates.

#### <span id="page-7-3"></span>*2.1.3 Fraction of observed infections rt*

<span id="page-7-4"></span>Grid search was not performed for  $r_t$ ; this was only estimated within the ABC framework. The prior distribution for  $r_t$  is estimated on April 15 (*r*1), and August 15, 2020 (*r*2). We interpolate a linear change between the values on the two dates, and assume that  $r_t$  before April 15 is equal to  $r_1$  and after August 15 is equal to  $r_2$ , i.e. the distribution shown in Table [5.](#page-9-3) The prior for  $r_1$  is allowed to vary over an interval truncated by minimum and maximum values calculated by seroprevalence studies performed by the CDC in March - early May, which were found to be as few as 34.7 times as many infections as observed cases, and as many as 3.2 times as many cases observed as estimated, in the study regions [\[19\]](#page-31-18). We allow the second date *r<sup>t</sup>* to vary over a much wider interval since we do not have prior information corresponding to this time period.

<span id="page-8-0"></span>

<span id="page-8-1"></span>Figure 1. The *Change in Average Mobility (Based on Distance Traveled)* and the *Difference in Encounter Density* metric for Los Angeles County from March 1 - Nov 15, 2020 from the the Unacast dashboard [\[18\]](#page-31-17).

prior $r_t$
$\sim N(3.65, 0.2)$
$\sim N(3.65, 0.2)$
$\sim N(0.77, 0.11)$
$\sim N(0.77, 0.11)$
$\sim N(1.2, 0.12)$
$\sim N(1.2, 0.12)$
$\sim N(0.77, 0.11)$
$\sim N(0.77, 0.11)$
$\sim N(2.01, 0.13)$
$\sim N(2.01, 0.13)$
$\sim N(2.51, 0.14)$
$\sim N(2.51, 0.14)$
$\sim N(1.3, 0.12)$
$\sim N(1.3, 0.12)$
$\sim N(1.3, 0.12)$

**Table 4.** Prior distribution for the time-varying reproductive number  $r_t$ .

date	r,
2020-03-01	$\sim U(0.20, 0.1)$
2020-04-15	$\sim U(0.20, 0.1)$
2020-08-15	$\sim U(0.50, 0.1)$
2021-03-01	$\sim U(0.50, 0.1)$

<span id="page-9-4"></span><span id="page-9-3"></span>Table 5. Prior distribution for the fraction of observed cases over all infections, *rt* .



Table 6. Prior distributions for the time-varying probabilities of each disease stage over time: probability of hospitalization given infection,  $\alpha_i$ ; probability of ICU admission given hospitalization,  $\kappa_i$ ; probability of death given ICU admission,  $\delta_i$ .

#### *2.1.4 Probabilities of disease stage progression*

Grid search ranges for the parameters representing the probabilities of disease stage progression,  $\alpha_t$ ,  $\kappa_t$ , and  $\delta_t$ , were informed by the ratios of the observed numbers of infections, hospitalizations, and deaths in LAC. Specifically, key change points in these ratios informed our modeling of change points in the distribution of the three parameters over time. At each change point (the date column in table [6\)](#page-9-4) we re-estimated the three parameters by first performing the broad grid search step over a wide range for each parameter ([.05,.9]). After choosing a mode returned by the grid search posterior distribution for each parameter at each time point, we set the prior parameter distributions as specified in table [6.](#page-9-4)

### <span id="page-9-0"></span>**2.1.5 Starting time,**  $t_0$

The first observed case of COVID-19 in LAC is registered as occurring on January 25 2020, and the epidemic began to grow exponentially in early March 2020 [\[20\]](#page-31-19). It is likely that the epidemic in LAC had multiple starting points. We therefore model *t*<sup>0</sup> as a uniform distribution with the unobserved first initial case beginning anywhere between January 1 and February 15 2020.

#### <span id="page-9-1"></span>**2.2 Variable observations used for parameter estimation**

The model was fit to the daily and cumulative count of observed infections and deaths, and current numbers in-hospital and in-ICU, coming from the GitHub page of the Los Angeles Times (LA Times) Data and Graphics Department [\[21\]](#page-31-20). The infection and death data is sourced from reports logged by LA Times reporters and editors based on reports from the LAC Department of Public Health. The in-hospital and in-ICU data was sourced by the LA Times directly from the California Department of Public Health's Open Data Portal [\[22\]](#page-31-21). We use the total of both confirmed and suspected COVID-19 patients in hospital or ICU.

#### <span id="page-9-2"></span>*2.2.1 Count variables used in model estimation*

We combine the new and cumulative counts of infections and deaths and current numbers in-hospital and in-ICU into a single computed summary statistic to use in ABC,  $Summ(\Phi) = \{I_{new}(t), I_{cum}(t), D_{new}(t), D_{cum}(t), H(t), Q(t)\}\$ , where  $\Phi$  represents the observed data. We truncate the data such that we do not include infection observations from early in the epidemic, when observations are more likely to be incomplete and unreliable. Specifically, we compute the summary statistic, *Summ*(Φ), for <span id="page-10-0"></span>the data  $\Phi$ : *Summ*( $\Phi$ ) = { $I_{cum}(t > March 15)$ ,  $H(t > March 20)$ ,  $Q(t > March 20)$ ,  $D_{cum}(t > March 25)$ }.

### **2.3 Running the model**

### <span id="page-10-1"></span>*2.3.1 Uncertainty quantification*

We use ABC to estimate all model parameters simultaneously, producing joint posterior probability estimates over all parameters. In forward simulations of our model, we simulate trajectories with parameter values coming from this joint posterior distribution, rather than a single value of a parameter. Uncertainty is also contributed by the stochastic differential equations, which model stochasticity in numbers of transitions between compartments.

To produce uncertainty estimates, we estimate credible intervals (CIs) for all model variables and estimated parameters by quantifying uncertainty from two sources: variability due to joint estimated parameter values, and variability due to the stochastic variability between model runs with the same parameters. We aggregate simulations from 100 jointly estimated parameter sets. For each parameter set, we simulate 20 realizations of the stochastic epidemic model, resulting in 2000 total realizations. We pool together all simulations and report their median and 2.5th/97.5th percentiles.

### <span id="page-10-2"></span>*2.3.2 Model implementation*

The model was implemented in R (version 3.6.3).

**Epidemic model initial values** In our experiments, model time  $t = 0$  is set to the ABC-estimated time for  $t_0$ , which is linked to the calendar date of around January 15, 2020, though this will vary across the estimated parameter sets. We set *S*(0) is equal to 10 million, approximately the population of LAC;  $I(0) = 1$ ,  $E(0) = 10$ ; and all other compartments initially empty.

**Parameter estimation implementation details** We apply the R package EasyABC [\[23\]](#page-31-22) and the algorithm *Marjoram*, which implements the algorithm of Marjoram et al. (2003) [\[24\]](#page-31-23) with improvement steps by Wegmann et al. (2009) [\[25\]](#page-31-24). The arguments that need to be specified are the prior parameter distributions, the original data Φ, the epidemic model to simulate data Φ<sup>∗</sup> from, and the function that computes summary statistics *Summ*(Φ). The additional parameters that need to be specified are the number *n* of simulations to perform in the calibration step; the number of parameter values to accept *N*, the *tolerance quantile* η that helps to determine the tolerance level ξ; and the *scale factor* ρ that helps to determine the size of the proposal range.

We set  $n = 100,000$  burn-in simulations,  $N = 1000$  sets of jointly estimated parameter values to save, and use the default values for tolerance quantile of  $\eta = 0.01$  and scale factor of  $\rho = 1$ . We test the model for convergence by running the parameter estimation procedure multiple times with different seeds and verifying similarity across variable outputs. Prior parameter distributions were adjusted such that convergence is achieved.

### <span id="page-10-3"></span>**2.4 Calculating population-wide** *CFR<sup>t</sup>* **and** *IFR<sup>t</sup>*

For each simulated model realization, we find the *CFR<sup>t</sup>* and *IFR<sup>t</sup>* as the estimated number of deaths (*D*) for each profile over estimated observed infections (*I*) for each profile, and number of deaths over total infections  $(I + A)$ , respectively. Repeating across the 2000 model realizations achieves the 95% CI.

### <span id="page-10-4"></span>**3 Epidemic model results**

### <span id="page-10-5"></span>**3.1 Estimated timeseries of numbers in infection states**

This section provides close-ups of the model-estimated timeseries of numbers in infection states relative to key dates and COVID-19 policy decisions in LAC. Available data is shown for new infections, new deaths, and current numbers in-hospital and in-ICU. Model-estimated current vs. total infections, although without data, are also shown. For each figure the the model-estimated median curves are plotted along with the 50th% (dark shading) and 95% CI (light shading).

### <span id="page-10-6"></span>**3.2 Parameter estimates**

Estimates of each of the five model parameters, and the population-wide *CFR<sup>t</sup>* and *IFR<sup>t</sup>* , at two week intervals between March 1, 2020 and March 1, 2021 are shown in Table [7.](#page-15-0)



Figure 2. New infections, *Inew*.



Figure 3. Current observed infections, *I*, and total infections, *I* +*A*.



Figure 4. Current in-hospital, *H*, and in-ICU, *Q*. The first of the two horizontal dotted lines shows the approximate hospital capacity limit, while the second shows the ICU capacity limit.



Figure 5. New deaths, *D*.

<span id="page-15-0"></span>

date	$R_t$	$R_t$ Effective	$r_{\rm t}$	$\pmb{\alpha}_t$	$\kappa_t$	$\delta_{\scriptscriptstyle{t}}$	$CFR_t$ %	$IFR_t$ %	
3/1/20	3.69(3.6, 3.82)	3.69(3.6, 3.82)	0.19(0.12, 0.26)	0.349(0.322, 0.38)	0.327(0.309, 0.348)	0.506(0.452, 0.565)	1.21 (0.21, 2.42)	0.24(0.03, 0.54)	
3/15/20	3.13(3.04, 3.24)	3.12 (3.02,3.24)	0.19(0.12, 0.26)	0.349(0.322, 0.38)	$\overline{0.327(0.309, 0.348)}$	0.506(0.452, 0.565)	1.21(0.77, 1.62)	0.24(0.08, 0.43)	
4/1/20	0.88(0.77, 0.95)	0.87(0.76, 0.95)	0.19(0.12, 0.26)	0.349(0.322, 0.38)	0.327(0.309, 0.348)	0.506(0.452, 0.565)	2.31(1.77, 2.67)	0.46(0.17, 0.75)	
4/15/20	0.88(0.77, 0.95)	0.87(0.75, 0.95)	0.19(0.12, 0.26)	0.349(0.322, 0.38)	0.327(0.309, 0.348)	0.506(0.452, 0.565)	$\overline{4.18}$ (3.27,4.74)	0.83(0.31, 1.36)	
5/1/20	$\overline{0.97}$ $(0.84, 1.05)$	0.95(0.8, 1.04)	0.23(0.15, 0.31)	0.349(0.322, 0.38)	0.327(0.309, 0.348)	0.506(0.452, 0.565)	5.56(4.35,6.3)	1.1(0.41, 1.81)	
5/15/20	1.26(1.06, 1.39)	1.23(1.01, 1.37)	0.27(0.18, 0.36)	0.153(0.118, 0.19)	0.196(0.143, 0.234)	0.526(0.503, 0.543)	5.86 (4.61,6.62)	1.16(0.43, 1.91)	
6/1/20	1.26(1.06, 1.39)	1.21(0.99, 1.37)	0.31(0.21, 0.41)	0.153(0.118, 0.19)	0.196(0.143, 0.234)	0.526(0.503, 0.543)	4.97 (3.86,5.64)	0.99(0.36, 1.64)	
6/15/20	1.26(1.06, 1.39)	1.2(0.97, 1.36)	0.34(0.23, 0.45)	0.153(0.118, 0.19)	0.196(0.143, 0.234)	0.526(0.503, 0.543)	4.18(3.2, 4.81)	0.84(0.3, 1.4)	
7/1/20	1.26(1.06, 1.39)	1.18(0.94, 1.36)	0.39(0.26, 0.5)	0.184(0.142, 0.228)	0.246(0.179, 0.293)	0.579(0.553, 0.597)	3.51(2.69, 4.18)	0.71(0.26, 1.19)	
7/15/20	1.14(0.97, 1.25)	1.05(0.83, 1.22)	0.42(0.28, 0.54)	$\overline{0.184(0.142,0.228)}$	0.246(0.179, 0.293)	0.579(0.553, 0.597)	3.2(2.46, 3.89)	0.64(0.24, 1.09)	
8/1/20	1(0.86, 1.09)	0.9(0.7, 1.05)	$\overline{0.46}$ (0.31,0.6)	0.153(0.118, 0.19)	0.196(0.143, 0.234)	0.526(0.503, 0.543)	3(2.3,3.68)	0.6(0.23, 1.03)	
8/15/20	0.88(0.77, 0.95)	0.78(0.61, 0.91)	0.5(0.34, 0.64)	0.153(0.118, 0.19)	0.196(0.143, 0.234)	0.526(0.503, 0.543)	2.86(2.19, 3.54)	0.58(0.22, 0.99)	
9/1/20	0.88(0.77, 0.95)	$\overline{0.77}$ $(0.6, 0.91)$	0.5(0.34, 0.64)	0.153(0.118, 0.19)	0.196(0.143, 0.234)	0.526(0.503, 0.543)	2.79(2.13, 3.45)	0.56(0.22, 0.97)	
9/15/20	0.88(0.77, 0.95)	0.76(0.58, 0.91)	$\overline{0.5}$ (0.34,0.64)	0.153(0.118, 0.19)	0.196(0.143, 0.234)	0.526(0.503, 0.543)	2.76(2.1, 3.42)	0.55(0.22, 0.97)	
10/1/20	0.88(0.77, 0.95)	0.76(0.58,0.9)	0.5(0.34, 0.64)	0.153(0.118, 0.19)	0.196(0.143, 0.234)	0.526(0.503, 0.543)	2.73 (2.08,3.38)	0.55(0.22,0.96)	
10/15/20	2.03(1.87, 2.22)	1.73(1.37,2.1)	0.5(0.34, 0.64)	0.1(0.077, 0.124)	0.157(0.114, 0.187)	0.737(0.704, 0.76)	2.74 (2.08,3.39)	0.55(0.22,0.96)	
11/1/20	2.03(1.87, 2.22)	1.69(1.34, 2.08)	$\overline{0.5}$ (0.34,0.64)	0.1(0.077, 0.124)	0.157(0.114, 0.187)	0.737(0.704, 0.76)	2.6(1.96, 3.24)	0.52(0.21, 0.92)	
11/15/20	2.03(1.87, 2.22)	1.63(1.29,2.03)	0.5(0.34, 0.64)	0.1(0.077, 0.124)	0.157(0.114, 0.187)	0.737(0.704, 0.76)	2.36(1.73,3.02)	0.47(0.19, 0.85)	
12/1/20	2.03(1.87, 2.22)	1.51(1.18, 1.91)	0.5(0.34, 0.64)	0.1(0.077, 0.124)	0.157(0.114, 0.187)	0.737(0.704, 0.76)	2.01(1.34, 2.73)	0.4(0.16, 0.74)	
12/15/20	2.03(1.87, 2.22)	1.33(1.01, 1.74)	$\overline{0.5(0.34, 0.64)}$	0.1(0.077, 0.124)	0.157(0.114, 0.187)	0.737(0.704, 0.76)	1.72(1.08, 2.44)	0.34(0.14, 0.64)	
1/1/21	2.03(1.87, 2.22)	1.05(0.72, 1.42)	0.5(0.34, 0.64)	0.1(0.077, 0.124)	$\overline{0.157(0.114, 0.187)}$	0.737(0.704, 0.76)	1.51(0.97, 2.2)	0.29(0.13, 0.54)	
1/15/21	1.4(1.29, 1.53)	0.57(0.35,0.8)	0.5(0.34, 0.64)	0.1(0.077, 0.124)	0.157(0.114, 0.187)	0.737(0.704, 0.76)	1.42(0.98, 2.01)	0.28(0.13, 0.49)	
2/1/21	$\overline{1.32(1.21,1.44)}$	0.45(0.27, 0.67)	0.5(0.34, 0.64)	0.1(0.077, 0.124)	0.157(0.114, 0.187)	0.737(0.704, 0.76)	1.5(1.12, 1.98)	0.29(0.14, 0.5)	
2/15/21	1.32(1.21, 1.44)	$\overline{0.42}$ (0.25,0.63)	0.5(0.34, 0.64)	0.1(0.077, 0.124)	0.157(0.114, 0.187)	0.737(0.704, 0.76)	1.59 (1.23,2.02)	0.31(0.15, 0.53)	
3/1/21	1.32(1.21, 1.44)	0.41(0.25, 0.62)	0.5(0.34, 0.64)	0.1(0.077, 0.124)	0.157(0.114, 0.187)	0.737(0.704, 0.76)	1.65(1.29, 2.06)	0.32(0.16, 0.55)	

Table 7. Median and 95% confidence interval (CI) of estimated model parameters and quantities over time. *CFR<sup>t</sup>* and *IFR<sup>t</sup>* are shown as percentages.

#### <span id="page-16-0"></span>**3.3 Posterior parameter densities**

This section visualizes the density plots for the prior distribution specified for, and the estimated posterior distribution returned by, ABC parameter estimation. A single value is shown for each parameter, corresponding to the value it takes on over a specific time interval. For each parameter, the bottom subfigure shows the posterior density over its full range, and the top subfigure shows the prior density over that same range. Therefore, if the prior density has a wider distribution or a different range than the posterior, the prior density will appear truncated in the figure. The prior distributions shown were informed by the results of the broad grid search that scanned across a much wider range of each parameter space (these ranges are specified in Section 2.1).

Convergence is not reached for the broad grid search step, with multi-modal distributions returned for each parameter (not shown). By specifying a narrow prior distribution around a mode chosen from the broad grid search distribution, convergence is reached and a dominant single mode identified in the final posterior density returned by the ABC step.

Although the prior distributions used in the ABC step are narrow (with 95% of a prior parameter's value lying within  $\pm 25\%$ of the mean of the chosen mode from broad grid search), they are not too narrow to allow the posterior distributions to take a different shape. All posterior distributions differ slightly from the priors; the mean of each posterior is not exactly aligned with the mean of the prior, and the standard deviations becomes narrower.



Figure 6. Basic reproductive number *R*0: value at beginning of the epidemic before interventions.



Figure 7. Fraction of infections that are observed, *rt* : value during May 2020 – March, 2021. Alpha(t) March - April 2020 Prior Distribution



**Figure 8.** Probability of hospitalization given infection,  $\alpha_t$ : value during March – April, 2020.







**Figure 10.** Probability of death given ICU admission,  $\delta_t$ , during  $t \in \text{March } - \text{April}, 2020$ .

**17[/31](#page-32-0)**

# <span id="page-19-0"></span>Part II Risk Model

This section provides details on the data inputs and calculations used in each of the 6 steps of the risk model described in the main text.

The notation used in this section is summarized in table [8](#page-19-3) (in the order that they appear in the following text). Although not noted in the table, if a variable contains a subscript *<sup>t</sup>* , this means it is time-varying.

<span id="page-19-3"></span>

Table 8. Notation used within the risk model.

### <span id="page-19-1"></span>**4 Modeled risk factors**

The risk factors,  $p \in P$ , included in the risk model analysis are:

- Age
- Body mass index (BMI)
- Smoking
- Any comorbidity: diabetes, hypertension, chronic obstructive pulmonary disease (COPD), hepatitis B, coronary heart disease, stroke, cancer and chronic kidney disease.

<span id="page-19-2"></span>Age was categorized within four groups:  $0 - 18$ ,  $19 - 49$ ,  $50 - 64$ ,  $65 - 79$ , and  $80 +$ . We modeled BMI as an ordinal variable and assume an additive effect of BMI on the three models. BMI was categorized in three groups according to obesity classes: Class 1 (no obesity)  $BMI < 30 \frac{kg}{m^2}$ ; Class 2 (obesity),  $30 \leq BMI \leq 40 \frac{kg}{m^2}$ ; Class 3 (severe obesity),  $BMI > 40 \frac{kg}{m^2}$ . Any comorbidity and smoking were modeled as binary variables. Note that risk factors are age, BMI, smoking and comorbidities, but age has 5 categories and BMI has 3 categories so  $p = \{1,..,10\}$ .

### **5 Step 2: Conditional RR for BMI, smoking, and comorbidities**

We estimate the conditional relative risk (RR) effects corresponding to the risk factors BMI, smoking, and comorbidity conditional on age for each of the three risk models using marginal effects estimates available from reported studies and a method called the joint analysis of marginal summary statistics (JAM) [\[26\]](#page-31-25). JAM uses two pieces of information: (i) the marginal effect estimates between risk factors and outcomes for each model, and (ii) a reference correlation structure between the risk factors. For information informing (i) we obtain the log marginal RR between individual risk factors *p* and COVID-19 infection severity for each model *m*,  $\psi_{p,m}^{Marg}$ , from peer-reviewed clinical studies on patients with laboratory-confirmed COVID-19 [\[27,](#page-31-26) [28\]](#page-32-2). For (ii), we obtain the reference correlation structure, Σ, using data from The National Health and Nutrition Examination Survey (NHANES) from 2017-2018 [\[29\]](#page-32-3).

### <span id="page-20-0"></span>**5.1 Marginal relative risks between risk factors from published literature**

We extracted the marginal RRs for each risk factor from clinical studies on patients with laboratory-confirmed COVID-19. Ordinal age, smoker, and any comorbidity we extracted from [\[27\]](#page-31-26), a study reporting outcomes for 1099 patients from 552 hospitals in 30 provinces, autonomous regions, and municipalities in mainland China. The marginal RR of ordinal BMI were extracted from [\[28\]](#page-32-2), with 4103 COVID-19 patients with laboratory-confirmed COVID-19 treated at a single academic health system in New York City. The left column of the main text Table 2 in the main text displays the marginal RRs extracted from the literature (95% confidence interval),  $\psi_{p,m}^{Marg}$ , for each risk factor and each of the three models hospitalization given illness,  $(H|I)$ , ICU admission given hospitalization,  $(Q|H)$  and death given ICU admission,  $(D|Q)$ .

#### <span id="page-20-1"></span>**5.2 Correlation structure between risk factors**

We obtain the correlation structure, Σ, between the risk factors *p* using data from The National Health and Nutrition Examination Survey (NHANES) [\[29\]](#page-32-3). NHANES is a survey research program conducted by the National Center for Health Statistics to assess the health and nutritional status of adults and children in the United States, and to track changes over time. We use the NHANES cohort of 2017-2018. To make the correlation matrix representative of the LAC population, we calculated it separately for each race/ethnicity and weight the correlation matrix by the distribution of the race/ethnicity in LAC.

### <span id="page-20-2"></span>**5.3 Using JAM to estimate conditional RR**

We use the JAM method [\[26\]](#page-31-25) to calculate the conditional log relative risk (RR) for each risk factor,  $\psi_p^{Cond}$ , from the log of the *marginal RR*  $\psi_p^{Marg}$  *and the correlation structure between risk factors Σ. The <i>reference group* was set as patients aged 0 – 18,  $BMI \leq 30 \frac{kg}{m^2}$ , non-smoker, and no comorbidities.

Using the marginal summary statistics from (i), specifically the marginal log relative risks  $\psi_{p,m}^{Marg}$  for risk factor  $p$  and model *m*, JAM obtains conditional log relative risks  $\psi_{p,m}^{Cond}$  for each factor. To accomplish this JAM first expresses the relationship between an outcome *m*, such as hospitalization given infection, ICU admission given hospitalization, and death given ICU admission, and the risk factors  $p \in P$  as a normal linear model,  $m \sim N({\bf P}\psi,\tau^2{\bf I})$ . For such a model the conditional or adjusted estimates of effect are given by  $\hat{\mathbf{\psi}} = (\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}'\mathbf{m}$ .

Heuristically, to fit this model without access to individual-level data we substitute an estimate of P'P based on an estimate of this matrix using the correlation  $\Sigma$  between the risk factors from external NHANES data as specified in (ii). P'm defines the mean value of the outcome for each of the corresponding values of the risk factor.

Technically, to adapt the linear model to use reference data, we first multiply all factors in the linear model by  $P'$  and define a *P*-length vector  $z := P'm$  to give  $z \sim N(P'P\psi, \tau^2P'P)$ . Since P'P is inherently Hermitian and, if it is positive definite, we perform a Cholesky decomposition and simplify the likelihood by first defining  $P'P = L'L$  and  $zL := L'^{-1}z$ . Thus, we have  $L'^{-1}z \sim MVN_P(L\psi, \tau^2 I_P)$ . For this model the conditional estimates are then  $\psi_p^{Cond} = (L'L)^{-1}z$ . z is constructed using marginal summary statistics and the frequencies for each risk factor. For a single binary risk factor,  $z := [(\hat{n}_0 \hat{n}_1)/(\hat{n}_0 + \hat{n}_1) \cdot \psi_p^{Marg}$ , where  $\hat{n}_0$  and  $\hat{n}_1$  are the estimated counts of each group defined in the risk factor by  $P = 0$  and  $P = 1$ , respectively. To get a plug-in for P'P we use the following: P'P  $\approx B^{1/2} \Sigma B^{1/2}$ , where B gives the diagonal elements of P'P defined using the frequencies of each risk factor and the sample size of the reference data.  $\Sigma$  is the estimated correlation matrix for the risk factors from the reference data. More details of the statistical methodology is in [\[26\]](#page-31-25).

The resulting conditional relative risks (i.e.,  $\exp \hat{\psi}_{p,m}^{Cond}$ ) calculated by JAM for each risk factor and each of the three models are shown in the right column of Table 2 in the main text.

### <span id="page-20-3"></span>**6 Step 3: Prevalence of each risk profile in the infected population**

We estimate the time-varying frequency of each risk profile in the infected population,  $f_{t,q,I}$ . First, we use available LAC data the prevalence of the individual risk factors within the overall LAC population [\[30\]](#page-32-4) [\[31\]](#page-32-5) to estimate the frequency of each risk profile within the overall population. Second, we use available data on the prevalence of each age group in illnesses [\[32\]](#page-32-6)

together with our estimate of the prevalence of each risk factor in the overall LAC population to estimate the frequency of each profile within the infected population on each date. Illness timeseries data by age group is used because this is the only individual risk factor with observed infection prevalence data in LAC.

### <span id="page-21-0"></span>**6.1 Frequency of risk profiles in overall LAC population**

We estimate the frequency of the risk profiles  $q$  in the overall LAC population,  $l_q$ , by simulating a sample population based on the prevalence of each individual risk factor in LAC and the weighted correlation structure between the risk factors obtained from NHANES data,  $\Sigma$  (Section [5.2\)](#page-20-1).

### <span id="page-21-1"></span>*6.1.1 Data sources*

The prevalence of age comes from the American Community Survey via the <tidycensus> R package [\[33\]](#page-32-7). The prevalence of obesity, comorbidities besides cancer, and smoking are taken from the Los Angeles County Health Survey (LACHS), study year 2018 [\[30\]](#page-32-4). To construct the obesity variable we find the  $BMI \leq 30 \frac{kg}{m^2}$  and  $BMI > 30 \frac{kg}{m^2}$  classes from LACHS. We then divide the *BMI* >  $30 \frac{kg}{m^2}$  into the  $30 \leq BMI < 40 \frac{kg}{m^2}$  and  $BMI > 40 \frac{kg}{m^2}$  classes based on the relative ratios between these classes on average across the U.S., in data from the Behavioral Risk Factor Surveillance System study year 2010 [\[34\]](#page-32-8). The prevalence of cancer comes from the California Health Information Survey (CHIS) [\[31\]](#page-32-5).

The prevalence of each of the 10 risk factors are combined in the vector  $\mathbf{l}_p$ .

### <span id="page-21-2"></span>*6.1.2 Estimation*

To calculate the frequency of the risk profiles *q* in the overall LAC population,  $l_q$ , we first generate a simulated population  $\chi$  by sampling from a multivariate normal, χ ∼ *N*(*x*;l*p*,Σ), where *x* is the number of samples, l*<sup>p</sup>* is the vector of the prevalence of each individual risk factor in LAC, and  $\Sigma$  is the correlation structure between the risk factors as described in Section [5.2.](#page-20-1) Each sample from  $\chi$  represents a specific risk profile, sampled in proportion to the probability of the combination of risk factors co-occurring. Prevalence of the *any comorbidity* category is constructed by sampling profiles expanded across all comorbidities, and creating an aggregated binary *any comorbidity* variable that takes value 1 if any of the comorbidities are included in the sampled profile.

We then calculate the vector of the frequencies of each risk profile  $q$  in the overall LAC population,  $I_q$ , as its relative frequency in the simulated population  $\chi$ .

### <span id="page-21-3"></span>**6.2 Frequency of risk profiles in infected population**

We obtain the frequency of each age group over illnesses,  $f_{t,p',I_{obs}}$  from the LA Times data [\[21\]](#page-31-20). To estimate the frequency of each risk profile  $q$  in the infected population,  $f_{t,q,I}$ , the frequency of each age group over infections is stratified across the risk profiles according to the relative frequency of each profile within each age group in the overall LAC population.

We find  $f_{i,q,I}$ , the vector of the frequency of each risk profile in the infected population, as follows.

### <span id="page-21-4"></span>*6.2.1 Profile design matrix*

First we define a risk profile design matrix **R**, a *q x p* matrix, that includes indicator variables for the presence or absence of each risk factor *p* (columns) within each risk profile *q* (row), encoding the linear combination of all (plausible) linear combinations of the risk factors:

$$
\mathbf{R} = \begin{bmatrix} R_{11} & \dots & R_{q1} \\ \dots & \dots & \dots \\ R_{1p} & \dots & R_{qp} \end{bmatrix}.
$$

We also define  $\mathbf{R}_{p'}$  to be a profile design matrix encoding age group risk factors only, i.e., a  $q \times 5$  matrix representing the first five columns of R.

### <span id="page-21-5"></span>*6.2.2 Calculation*

We find the vector of the frequency of the risk profiles in the infected population,  $f_{t,q,I}$ , using  $I_q$ ,  $f_{t,p',I_{obs}}$ , and  $R_{p'}$ .

First, let  $I_{p'}$  be a vector representing the frequency of each age group in the (estimated) overall population, found as  $\mathbf{l}_{p'} = \mathbf{l}_q^{\mathsf{T}} \mathbf{R}_{p'}.$ 

Then we find  $f_{t,q,I}$  as

<span id="page-21-6"></span>
$$
\mathbf{f}_{t,q,I} = \mathbf{l}_q \cdot \frac{\mathbf{R}_{p'} \mathbf{f}_{t,p',I_{obs}}}{\mathbf{R}_{p'} \mathbf{l}_{p'}}
$$

.

### **7 Step 4: Risk-profile-stratified probabilities of disease stage progression**

### <span id="page-22-0"></span>**7.1 Overview**

We estimate the time-varying disease stage probabilities stratified across all risk profiles,  $q \in Q$ , for the three models  $m = 1$ :  $P_t(H|I)$ ,  $m = 2: P_t(Q|H)$ , and  $m = 3: P_t(D|Q)$ , resulting in the probability vectors  $P_t(H|I)$ ,  $P_t(Q|H)$ , and  $P_t(H|I)$ , respectively. This is done by combining in a logistic model all linear combinations of the *p* risk factors specified in a mean-centered design matrix,  $X_m$ , and their corresponding conditional log-RR obtained from JAM,  $\hat{\psi}_{p,m}^{Cond}$  (Section [5\)](#page-19-2), with intercepts that are set to the logit of the estimated probabilities from the epidemic model for  $\hat{\alpha}_t$ ,  $\hat{\kappa}_t$ ,  $\hat{\delta}_t$  (Section [2.1.4\)](#page-7-4), respectively. The matrix  $X_m$  is a profile design matrix that has been mean-centered on the frequency of each risk profile in the incoming population relevant to each model *m*; namely, the infected (*I*) population for model 1, the hospitalized (*H*) population for model 2, and the in-ICU (*Q*) population for model 3. The specification of the frequency of each risk profile in the infected (*I*) population is described in Section [6.](#page-20-3) The frequency of each risk profile in the hospitalized population,  $f_{t,q,H}$  and the in-ICU population,  $f_{t,q,Q}$ , are calculated recursively from the estimated frequency of each profile in the incoming population to each stage of disease, described in Section [7.2.3.](#page-22-4) Patients age  $19-49$  with  $BMI < 30 \frac{kg}{m^2}$ , non-smoker, and no comorbidities are the reference profile.

### <span id="page-22-1"></span>**7.2 Inputs**

### <span id="page-22-2"></span>7.2.1 Logit transformed population-average probabilities of each stage of disease  $\hat{\alpha}_t$ ,  $\hat{\kappa}_t$ ,  $\hat{\delta}_t$

In Section [2.1.4](#page-7-4) we estimate the population-average probability that individuals in LAC who acquire infection are admitted to hospital,  $\hat{\alpha}_t$ , who are in hospital require admittance to the ICU,  $\hat{\kappa}_t$ , and who are in ICU will die,  $\hat{\delta}_t$ . We logit-transform these probabilities as

$$
logit(\hat{\alpha}_t) = log(\frac{\hat{\alpha}_t}{1 - \hat{\alpha}_t})
$$

$$
logit(\hat{\kappa}_t) = log(\frac{\hat{\kappa}_t}{1 - \hat{\kappa}_t})
$$

$$
logit(\hat{\delta}_t) = log(\frac{\hat{\delta}_t}{1 - \hat{\delta}_t}).
$$

### <span id="page-22-3"></span>*7.2.2 <code>Vector</code> of conditional log risk estimates for model*  $m$ *,*  $\hat{\Psi}_{m}^{Cond}$

In Section [5](#page-19-2) we described the conditional log risk effects estimates, i.e.,  $\hat{\psi}_{p,m}^{Cond}$  for risk factor p and model m. We bring the conditional log risk effect estimates for model *m* together in to the vector  $\hat{\psi}_{m}^{Cond}$ *m* .

### <span id="page-22-4"></span>*7.2.3 Frequencies of each incoming population* f*t*,*q*,*in*

In Section [6](#page-20-3) we estimated the frequency of each risk profile  $q$  in the infected population,  $f_{t,q,I}$  (a vector). The frequency of each risk profile in the hospitalized population,  $f_{t,q,H}$  and the in-ICU population,  $f_{t,q,Q}$ , are calculated recursively from the estimated frequency of each profile in the incoming population to each stage of disease, as the normalized product of the frequency in the incoming stage of disease and the probability of advancing to the subsequent stage, normalized over all risk profiles.

Specifically, the frequency of each risk profile in the hospitalized population,  $f_{t,q,H}$ , is calculated from the estimated  $P_t(H|I)$ and  $f_{t,q,I}$  as:

$$
\mathbf{f}_{t,q,H} = \frac{\widehat{\mathbf{P}_{t}(\mathbf{H}|\mathbf{I})} \cdot \mathbf{f}_{t,q,\mathbf{I}}}{\widehat{\mathbf{P}_{t}(\mathbf{H}|\mathbf{I})}^{\mathsf{T}} \mathbf{f}_{t,q,\mathbf{I}}},
$$

and the frequency of each risk profile in the in-ICU population,  $\mathbf{f}_{t,q,O}$ , is calculated from the estimated  $P_t(Q|H)$  and  $\mathbf{f}_{t,q,H}$  as:

$$
f_{t,q,Q} = \frac{\widehat{P_{t}(Q|H)} \cdot f_{t,q,H}}{\widehat{P_{t}(Q|H)}^{\text{T}} f_{t,q,H}}.
$$

We also calculate the frequency of each profile in the deceased population,  $f_{t,q,D}$ , which is needed in Step 5 and 6 although not for calculating the risk-profile-stratified probabilities of illness progression. f*t*,*q*,*<sup>D</sup>* is found as the normalized product of the frequency in the incoming stage of disease, the in-ICU population, and the probability of advancing from the ICU to the deceased population, normalized over all risk profiles, i.e.:

$$
\mathbf{f}_{t,q,D} = \frac{\widehat{\mathbf{Pr(D|Q)}_{q,t} \cdot \mathbf{f}_{t,q,Q}}}{\widehat{\mathbf{Pr(D|Q)}_{q,t} \cdot \mathbf{f}_{t,q,Q}}}.
$$

In the following we will require the marginal frequency of each risk factor *p* in each incoming population, f*t*,*p*,*<sup>I</sup>* , which we find by marginalizing the risk factors over the risk profiles *q* for each population as

$$
\mathbf{f}_{t,p,in} = \mathbf{f}_{t,q,in}^{\mathsf{T}} \mathbf{R}.\tag{10}
$$

#### <span id="page-23-0"></span>*7.2.4 Mean-centered design matrix* X*m, and risk profile design matrix* R

The mean-centered design matrix for model *m*, X*m*, is a profile design matrix that has been mean-centered on the frequency of each risk profile in the incoming population  $m_{in}$  relevant to each model  $m$ ,  $\mathbf{f}_{p,m_{in}}$ ; namely, the infected (*I*) population for model 1, the hospitalized (*H*) population for model 2, and the in-ICU (*Q*) population for model 3.

We find the mean-centered design matrices as:

$$
\mathbf{X}_1 = \mathbf{R} - \mathbf{f}_{t,p,I}
$$

$$
\mathbf{X}_2 = \mathbf{R} - \mathbf{f}_{t,p,H}
$$

$$
\mathbf{X}_3 = \mathbf{R} - \mathbf{f}_{t,p,Q}.
$$

### <span id="page-23-1"></span>**7.3 Calculating risk-stratified probabilities**

We find the probability of each risk profile for model 1,  $\bar{P}_t(H|\bar{I})$ , using the frequency of each risk factor in the infected population, f*p*,*<sup>I</sup>* , as:

<span id="page-23-4"></span>
$$
\widehat{P_t(H|I)} = \text{expit}(\text{logit}(\hat{\alpha}_t) + \mathbf{X}_1 \hat{\underline{\psi}}_{m=1}^{Cond})
$$

We can then find the probability of each risk profile for model 2,  $P_t(Q|H)$ , using the frequency of each risk factor in the hospitalized population,  $f_{p,H}$ , as:

$$
\widehat{P_t(Q|H)} = \text{expit}(\text{logit}(\hat{\kappa}_t) + \mathbf{X}_2 \underline{\hat{\Psi}}_{m=2}^{Cond}).
$$

Finally, we find the probability of each risk profile for model 3,  $P_t(D|Q)$ , using the frequency of each risk factor in the in-ICU population,  $f_{p,Q}$ , as:

<span id="page-23-5"></span>
$$
\widehat{P_t(D|Q)} = \text{expit}(\text{logit}(\hat{\delta}_t) + \mathbf{X}_3 \underline{\hat{\mathbf{\Psi}}}_{m=3}^{Cond}).
$$

### <span id="page-23-2"></span>**8 Step 5: Conditional RR for age**

We estimate the conditional RR of each age group separately from that of the other three risk factors. This is done because while we can estimate the conditional RR using the same methodology as for the other factors as described in Step 2, we have observed data on the distribution of each age group over deaths for LAC that we can use to estimate the conditional RR for age, given the other model inputs and attributes. Given this information, we aim to find the solution set that minimizes the distance between the distribution over deaths produced by the logistic model and the observed distribution.

Specifically, we choose the conditional RR for age,  $RR_{p'}$ , such that the distance between the frequency of each age group over deaths produced in Step 4,  $f_{t,p',D}$  (marginalized from the profile-specific values  $f_{t,p',D}$  using equation [10\)](#page-23-4) and the observed distribution of each age group over deaths in LAC,  $\mathbf{f}_{t,p',D_{obs}}$ , is minimized, i.e.

$$
\arg\min_{RR_{p'}\in[0,10]}\text{Dist}_D = \sum_{t,p'}\mathbf{f}_{t,p',D}(RR_{p'}) - \mathbf{f}_{t,p',D_{obs}},\tag{11}
$$

where the frequency of each age group over deaths is written as  $f_{t,p',D}(RR_{p'})$  to emphasize that it is a function of the conditional RR of the age groups  $RR_{p'}$ .

<span id="page-23-3"></span>This objective is solved by varying  $RR_{p'}$  between [0,10], rerunning Steps 3 and 4, and computing the distance metric Dist<sub>*D*</sub> in equation [11.](#page-23-5)

### **9 Step 6: Risk-profile-stratified** *CFRq*,*<sup>t</sup>* **and** *IFRq*,*<sup>t</sup>*

We estimate the time-varying risk-profile-stratified case fatality rate (*CFRt*) and the infection fatality rate (IFR(*t*)) from cumulative counts of observed infections,  $I_{cum}$ , total infections,  $I_{cum} + A_{cum}$ , and cumulative deaths  $D_{cum}$  coming from the epidemic model; and estimates of the time-varying frequency of each risk profile in the infected population, f*t*,*q*,*<sup>I</sup>* (from Step 3) and of each risk profile in the deceased population (from Step 4).

We simulate model realizations with parameter values coming from this joint posterior distribution [2.3.](#page-10-0) For each realization, we estimate the number of individuals in each risk profile subpopulation in the observed infected population as the estimated frequency of each profile in the infected population, f*t*,*q*,*<sup>I</sup>* , multiplied by the epidemic-model-estimated time series of the cumulative number of observed infections (*Icum*). Similarly, we estimate the number of individuals in each risk profile subpopulation in the total infected population (including observed an unobserved illnesses) as f*t*,*q*,*<sup>I</sup>* multiplied by the estimated time series of cumulative total infections, i.e.  $I_{cum} + A_{cum}$ . We estimate the number of deceased individuals from each risk profile as f*t*,*q*,*<sup>D</sup>* multiplied by the estimated time series of cumulative deaths (*Dcum*).

We find the CFR and IFR for each model realization as the estimated number of deaths for each profile over estimated observed infections for each profile, and number of deaths over total infections, respectively.

To produce uncertainty estimates, we aggregate model simulations from 100 jointly estimated parameter sets and 20 stochastic epidemic model realizations for each parameter set. We pool together all simulations and report their mean and 2.5th/97.5th percentiles.

### <span id="page-24-0"></span>**10 Results for risk-profile-stratified estimates**

### <span id="page-24-1"></span>**10.1 Comparison between observed and estimated frequency of each age group in deceased population (for Step 5)**

Figure [11](#page-25-0) shows the frequency of the age groups in the deceased population estimated by the model,  $f_{t,p',D}(RR_{p'})$ , compared with the observed frequency,  $f_{t,p',D_{obs}}$ . Minimizing the summed differences between age groups  $p'$  in the distributions  $f_{t,p',D}(RR_{p'})$ and  $f_{t,p',D_{obs}}$ , i.e. the metric Dist<sub>D</sub>, is the objective used to estimate the conditional RR for age in Step 5 of the risk model methodology. Figure [11](#page-25-0) demonstrates a very low difference error is achieved.

### <span id="page-24-2"></span>**10.2 Profile-stratified probabilities of severe illness and death for LAC**

The resulting probabilities across each risk profile for each of three models  $P_t(H|I)$ ,  $P_t(Q|H)$ , and  $P_t(D|Q)$ , as well as the estimated frequency of each profile in the overall LAC population across dates every two weeks from May 15 2020 - March 1 2021 are shown in Tables [9,](#page-26-0) [10,](#page-27-0) and [11,](#page-28-0) respectively.

<span id="page-25-0"></span>

**Figure 11.** Frequency of the age groups in the deceased population estimated by the model,  $f_{t,p',D}(RR_{p'})$ , compared with the observed frequency,  $f_{t,p',D_{obs}}$ , used to estimate the conditional RR for age in Step 5 of the risk model.

<span id="page-26-0"></span>

Table 9. Profile-stratified *<sup>P</sup>* Table 9. Profile-stratified  $\widehat{P_t(H|I)}$ : Risk profiles (characterized by unique combination of age group, BMI range, smoking status, and any comorbidity), model-estimated population prevalence in LAC, and the probability

<span id="page-27-0"></span>

50	$80+$	30 < BMI < 40	Smoker	Comorbidity	0.00022	0.6338	0.7344	0.6632	0.6654	0.6695	1.6684	0.6663	0.6662
◡	19-49	BMI >40	Smoker	Comorbidity	0.00112	0.0909	0.1378	0.1022	0.103	0.1048	0.1043	0.1034	0.1034
ے ر	50-64	BMI > 40	Smoker	Comorbidity	0.00026	).2055	0.2924	0.2273	0.229	0.2324	0.2315	0.2298	0.2297
	65-79	BMI > 40	Smoker	Comorbidity	0.00013	0.4008	J.5166	0.432	0.4344	0.4391	0.4378	0.4355	0.4353
	$80+$	BMI > 40	Smoker	Comorbidity	0.00001	0.645	0.7439	0.674	0.6762	0.6803	0.6791	0.677	0.677

**Table 10.** Profile-stratified  $\widehat{P_t(Q|H)}$ : Risk profiles (characterized by unique combination of age group, BMI range, smoking status, and any comorbidity), model-estimated population prevalence in LAC, and the probabil

<span id="page-28-0"></span>



Table 11. Profile-stratified  $\widehat{P_t(D|Q)}$ : Risk profiles (characterized by unique combination of age group, BMI range, smoking status, and any comorbidity), model-estimated population prevalence in  $LAC$ , and the probability of death given ICU admission.

### **10.3 CFR**(*t*) **and IFR**(*t*) **over all risk profiles**

The median and 95% CI of resulting CFR(*t*) and IFR(*t*) across each risk profile, as well as the estimated frequency of each profile in the overall LAC population across dates every two weeks from May 15 2020 - March 1 2021 are shown in Tables [12](#page-29-1) and [13.](#page-30-0)

<span id="page-29-1"></span><span id="page-29-0"></span>



Table 12. Profile-stratified CFR(*t*) (as decimal). Risk profiles (characterized by unique combination of age group, BMI range, smoking status, and any comorbidity), model-estimated population prevalence in LAC, the frequency of each profile in the infection population, and the median and 95%CI of the Case Fatality Rate (CFR $(t)$ ).

<span id="page-30-0"></span>

Table 13. Profile-stratified IFR(*t*) (as decimal). Risk profiles (characterized by unique combination of age group, BMI range, smoking status, and any comorbidity), model-estimated population prevalence in LAC, the frequency of each profile in the infection population, and the median and 95%CI of the Infection Fatality Rate (CFR $(t)$ ).

### **References**

- <span id="page-31-0"></span>1. Li, Q. *et al.* Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia. *New England Journal of Medicine* (2020).
- <span id="page-31-1"></span>2. Wang, C. *et al.* Evolving epidemiology and impact of non-pharmaceutical interventions on the outbreak of coronavirus disease 2019 in wuhan, china. *MedRxiv* (2020).
- <span id="page-31-2"></span>3. Wang, D. *et al.* Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in wuhan, china. *Jama* 323, 1061–1069 (2020).
- <span id="page-31-3"></span>4. Zhou, F. *et al.* Clinical course and risk factors for mortality of adult inpatients with covid-19 in wuhan, china: a retrospective cohort study. *The lancet* (2020).
- <span id="page-31-4"></span>5. Coulson, T., Rohani, P. & Pascual, M. Skeletons, noise and population growth: the end of an old debate? *Trends in Ecology & Evolution* 19, 359–364 (2004).
- <span id="page-31-5"></span>6. Breto, C., He, D., Ionides, E. L., King, A. A. ´ *et al.* Time series analysis via mechanistic models. *The Annals of Applied Statistics* 3, 319–348 (2009).
- <span id="page-31-6"></span>7. Mode, C. J. & Sleeman, C. K. *Stochastic processes in epidemiology: HIV/AIDS, other infectious diseases, and computers* (World Scientific, 2000).
- <span id="page-31-7"></span>8. Lekone, P. E. & Finkenstädt, B. F. Statistical inference in a stochastic epidemic seir model with control intervention: Ebola as a case study. *Biometrics* 62, 1170–1177 (2006).
- <span id="page-31-8"></span>9. Keeling, M. J. & Rohani, P. *Modeling infectious diseases in humans and animals* (Princeton University Press, 2011).
- <span id="page-31-9"></span>10. Jacquez, J. A. & O'Neill, P. Reproduction numbers and thresholds in stochastic epidemic models i. homogeneous populations. *Mathematical Biosciences* 107, 161–186 (1991).
- <span id="page-31-10"></span>11. Diekmann, O., Heesterbeek, J. & Roberts, M. G. The construction of next-generation matrices for compartmental epidemic models. *Journal of the Royal Society Interface* 7, 873–885 (2010).
- <span id="page-31-11"></span>12. Britton, T. Basic stochastic transmission models and their inference. *Handbook of Infectious Disease Data Analysis* 111 (2019).
- <span id="page-31-12"></span>13. Ridenhour, B., Kowalik, J. M. & Shay, D. K. Unraveling r0: Considerations for public health applications. *American journal of public health* 108, S445–S454 (2018).
- <span id="page-31-13"></span>14. Liu, Y., Gayle, A. A., Wilder-Smith, A. & Rocklöv, J. The reproductive number of covid-19 is higher compared to sars coronavirus. *Journal of travel medicine* (2020).
- <span id="page-31-14"></span>15. Ridenhour, B., Kowalik, J. M. & Shay, D. K. Unraveling r 0: Considerations for public health applications. *American journal of public health* 108, S445–S454 (2018).
- <span id="page-31-15"></span>16. Bertozzi, A. L., Franco, E., Mohler, G., Short, M. B. & Sledge, D. The challenges of modeling and forecasting the spread of covid-19. *arXiv preprint arXiv:2004.04741* (2020).
- <span id="page-31-16"></span>17. Oliver, N. *et al.* Mobile phone data for informing public health actions across the covid-19 pandemic life cycle (2020).
- <span id="page-31-17"></span>18. Unacast. The unacast social distancing scoreboard (2020). URL [https://unacast.com/post/](https://unacast.com/post/the-unacast-social-distancing-scoreboard) [the-unacast-social-distancing-scoreboard](https://unacast.com/post/the-unacast-social-distancing-scoreboard).
- <span id="page-31-18"></span>19. Havers, F. P. *et al.* Seroprevalence of antibodies to sars-cov-2 in 10 sites in the united states, march 23-may 12, 2020. *JAMA Internal Medicine* (2020).
- <span id="page-31-19"></span>20. Los Angeles County Department of Public Health. La county daily covid-19 data dashboard (2020). URL [http://](http://publichealth.lacounty.gov/media/coronavirus/data/index.htm) [publichealth.lacounty.gov/media/coronavirus/data/index.htm](http://publichealth.lacounty.gov/media/coronavirus/data/index.htm).
- <span id="page-31-20"></span>21. Los Angeles Times Data and Graphics Department. california-coronavirus-data github (2021). URL [https://github.](https://github.com/datadesk/california-coronavirus-data) [com/datadesk/california-coronavirus-data](https://github.com/datadesk/california-coronavirus-data).
- <span id="page-31-21"></span>22. of Public Health, C. D. California department of public health open data portal (2021). URL [https://data.ca.gov/](https://data.ca.gov/dataset/covid-19-hospital-data1) [dataset/covid-19-hospital-data1](https://data.ca.gov/dataset/covid-19-hospital-data1).
- <span id="page-31-22"></span>23. Jabot, F., Faure, T. & Dumoulin, N. Easy abc: performing efficient approximate b ayesian computation sampling schemes using r. *Methods in Ecology and Evolution* 4, 684–687 (2013).
- <span id="page-31-23"></span>24. Marjoram, P., Molitor, J., Plagnol, V. & Tavaré, S. Markov chain monte carlo without likelihoods. *Proceedings of the National Academy of Sciences* 100, 15324–15328 (2003).
- <span id="page-31-24"></span>25. Wegmann, D., Leuenberger, C. & Excoffier, L. Efficient approximate bayesian computation coupled with markov chain monte carlo without likelihood. *Genetics* 182, 1207–1218 (2009).
- <span id="page-31-25"></span>26. Newcombe, P. J., Conti, D. V. & Richardson, S. Jam: a scalable bayesian framework for joint analysis of marginal snp effects. *Genetic epidemiology* 40, 188–201 (2016).
- <span id="page-31-26"></span>27. Guan, W.-j. *et al.* Clinical characteristics of coronavirus disease 2019 in china. *New England journal of medicine* 382,

<span id="page-32-1"></span><span id="page-32-0"></span>1708–1720 (2020).

- <span id="page-32-2"></span>28. Petrilli, C. M. *et al.* Factors associated with hospitalization and critical illness among 4,103 patients with covid-19 disease in new york city. *MedRxiv* (2020).
- <span id="page-32-3"></span>29. Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data (NHANES) (2017-2018). URL [https://www.cdc.](https://www.cdc.gov/nchs/nhanes/index.htm) [gov/nchs/nhanes/index.htm](https://www.cdc.gov/nchs/nhanes/index.htm).
- <span id="page-32-4"></span>30. Los Angeles County Department of Public Health. Los angeles county health survey (LACHS) (2018). URL [http:](http://publichealth.lacounty.gov/ha/hasurveyintro.htm) [//publichealth.lacounty.gov/ha/hasurveyintro.htm](http://publichealth.lacounty.gov/ha/hasurveyintro.htm).
- <span id="page-32-5"></span>31. UCLA Center for Health Policy Research, C. A. .-., Los Angeles. California health information survey (2018). URL <http://ask.chis.ucla.edu>.
- <span id="page-32-6"></span>32. Los Angeles County Department of Public Health. Los angeles county case counts, crude rates by selected region (2020). URL <https://github.com/AbigailHorn/COV2-LA/tree/master/data>.
- <span id="page-32-7"></span>33. U.S. Census Bureau via R package <tidycensus>. Selected age groups, 2014-2018 american community survey 5-year estimates (2018). URL <https://CRAN.R-project.org/package=tidycensus>.
- <span id="page-32-8"></span>34. Strum, R. & Hattori, A. Morbid obesity rates continue to rise rapidly in the us. *International Journal of Obesity (London)* 37, 889–891 (2013).