

# Incorporation of risk factors in a stochastic epidemiological COVID-19 model for Los Angeles County

Abigail L. Horn<sup>a,1,\*</sup>, Lai Jiang<sup>a</sup>, Wendy Cozen<sup>a</sup>, ... Other authors here ....,  
David V. Conti<sup>a</sup>

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

---

## Abstract

### Summary

*Background.*

*Methods.*

*Findings.*

*Interpretation.*

### Research in Context

*Evidence Before This Study*

*Added Value of This Study*

*Implications of All the Available Evidence*

---

---

\*Corresponding author

Email address: abigail.horn@usc.edu (Abigail L. Horn )

## 1. Introduction

Health disparities exist for most prominent health issues, ranging from risky health behaviors, to chronic and infectious disease, and healthcare access. Disparities are unjust differences in health outcomes that arise for certain groups and populations because of social, economic, and/or environmental disadvantage [1]. Recognizing how health risks and diseases affect particular groups and populations differently, and addressing these health disparities to achieve health equity, is a central goal of public health policy. As expected, health disparities have also emerged with the COVID-19 epidemic because the prevalence of risk factors for likelihood of infection and severity of outcomes once infected varies within populations.

Throughout the evolving COVID-19 pandemic there has been a significant amount of work to develop epidemiological models to provide country, region, and county-level estimates of key parameters important for policy planning [REFs]. These parameters include the time-varying basic reproduction number,  $R(t)$ , defined as the mean number of secondary cases generated by a typical infectious individual on each day in a full susceptible population [2], and the rate of illness observation, important given the large rates of asymptomatic cases and low testing rates, especially early in the epidemic. However, estimated epidemic parameters such as rates of severe illness and death, the case fatality rate (CFR, defined as deaths over observed illnesses) and infection fatality rate (IFR, defined as deaths over all illnesses) are often expressed in terms of a single population-level estimates [3]. In reality, these rates are vastly heterogeneous across subpopulations, primarily for individuals of advanced age and preexisting health conditions, and across race/ethnicity groups [4]. For public health policy makers to better address the pandemic, models reporting strata-specific estimates are necessary to investigate the potential outcomes of subpopulation-level targeted policy scenarios.

A primary reason that epidemic models have been limited to single-population estimates is the lack of epidemiological data for subpopulation-level groups. Most public health authorities collect and share data on counts of illness, hospitalizations, and deaths from COVID-19 at the population level, as well as data for select subpopulations comprised by single risk factors such as age and race/ethnicity [5]. Single risk factor subpopulations represent an average across all combinations of conditions, whereas it is important to understand the individual contribution of multiple risk factors and the effects of combinations of risk factors. Estimates of the joint or conditional effect of combinations of risk factors are conventionally obtained through analysis on individual-level data. However, individual-level data is unavailable in the U.S. except for in small studies and rarely at a regional population or subpopulation level.

In this paper we develop a model that estimates heterogeneity in illness severity and death by subpopulation from available epidemic dynamic data at the population level. Focusing on Los Angeles County (LAC), a population of approximately 10 million, we develop a framework to estimate the CFR, IFR, and probabilities of hospitalization, ICU admission, and death given infection

across 39 combinations of risk factors, defined here as *risk profiles*, representing all plausible combinations of modeled risk factors for COVID-19. This framework incorporates a stochastic single-population epidemic dynamic model (*epidemic model*) and a model to estimate conditional effects of risk factors (*risk model*). The epidemic model, which uses Bayesian methods for parameter estimation and uncertainty quantification, is used to estimate the time-varying population-average probabilities of severe illness and death for LAC. In the absence of direct access to individual-level data, the risk model applies a statistical technique commonly used in genetics to obtain estimates of the conditional effects of risk factors for COVID-19 on the probability of severe illness and death using data from published studies [] on the marginal risk of individual factors. We consider the risk factors age, existing comorbidities, obesity, and smoking. The comorbidities included are diabetes, hypertension, chronic obstructive pulmonary disease (COPD), hepatitis B, coronary heart disease, stroke, cancer and chronic kidney disease. We integrate the epidemic-model-estimated population-average probabilities of severe illness and the risk-model-estimated conditional effect of each risk factor, together with observed LAC data on the time-varying prevalence of each marginal age subpopulation in illnesses, to estimate the time-varying stratified CFR, IFR and probabilities of each stage of disease across all risk profiles.

The integrated model allows the analysis of dynamic outcomes and parameters by population and subpopulation. Focusing on the period March 1 - July 20, 2020, the time from the first cases until the highest peak of illnesses in LAC [3], we analyze the time course of the epidemic in LAC in relation to interventions and policy decisions in the context of the estimated epidemic parameters  $R(t)$ , the rate of illness observation, the time-varying probabilities of severe illness, and the CFR and IFR. Through simulation, we compare observed trends with the effects of alternate policies relating to adjusting the community illness transmission rate and protecting specific at-risk subpopulations. Results illustrate the extensive variation in hospitalization and CFR and IFR by sub-population, highlighting the importance of strata-specific estimates in understanding the impact of the epidemic and the potential outcomes of future subpopulation-level targeted policy interventions.

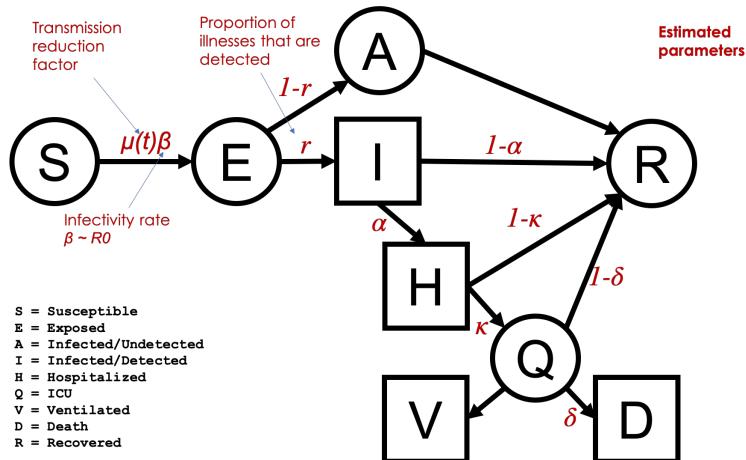
## 2. Methods

We first developed a single-population stochastic dynamic epidemic model that accounts for observed and unobserved transmission of COVID-19 and trajectories through the healthcare system with hospitalizations, ICU admission, and death. In parallel, we used available data from published studies on the marginal risk of individual risk factors (age, existing comorbidities, obesity, smoking) to calculate conditional risk effects estimates for risk factors for three models: (1) hospitalization given illness, (2) ICU admission given hospitalization, and (3) death given hospitalization. These conditional risk estimates were then integrated with corresponding probability parameter estimates from the dynamic epidemic model to create a *risk model*. The estimated probabilities

90 from the epidemic model are population-average parameters corresponding to  
 transitions between the infected, hospitalized, ICU, ventilated, death, and recovery  
 compartments. The *risk model* enables us to estimate, for each risk profile  
 within LAC, the probability of each stage of disease given infection. The distri-  
 bution of each risk profile in illnesses, the initial stage of disease, was informed  
 95 by observed LAC data on the prevalence of each age strata over illnesses, strat-  
 ified across the risk profiles according to an estimate of the frequency of each  
 risk profile within each age group. The estimated frequency of each risk profile  
 in the illness distribution, the probability of each risk profile for each stage of  
 disease, and epidemic model estimates for the cumulative number of infections  
 100 and deaths, all dynamic in time, are used to estimate the risk-stratified CFR  
 and IFR. The resulting stratified risk probabilities and CFR/IFR are therefore  
 mean-centered on the epidemic model estimated probabilities of each stage of  
 disease, and linked to LAC data on illnesses by age group.

105 The epidemic compartmental model uses stochastic differential equations  
 and approximate Bayes calculation techniques for parameter estimation, en-  
 abling presentation of the uncertainty in all estimations and predictions. We  
 influence parameter estimates through prior specification, using mobility data  
 to inform the prior on changes in disease transmission rates, and seroprevalence  
 110 data to inform the prior on the rate of illness detection. Time series data on the  
 numbers of individuals infected (observed), hospitalized, on ventilation, dead,  
 and recovered, are used in the model fitting procedure. Incorporating data be-  
 yond observed infected cases allows for more reliable overall model estimation,  
 since these counts are likely more accurately reported than illnesses.

## 2.1. Epidemic model



**Figure 1:** Epidemic model structure and key estimated parameters

115 We develop a stochastic model of COVID-19 transmission in a single, homogeneously-  
 mixed population divided into 9 compartments representing different disease

states (Figure 1), with available data represented as square compartments. Compartments relating to the transmission of infection are the widely-used susceptible, exposed (latent but not yet infectious), infectious and observed ( $I$ ), and recovered classes. We also include a compartment representing unobserved and/or unconfirmed infections ( $A$ ). We model healthcare utilization and outcome at a more granular level by including compartments representing individuals that are in hospital ( $H$ ), in ICU care ( $Q$ ), undergoing ventilation ( $V$ ), and that die ( $D$ ). Each individual can only be in one state at each point in time with the exception of the ventilated class, which is a subset of the in-ICU class. We assume that new infections are created only by individuals in the infected classes ( $I$  and  $A$ ), and that individuals in all other compartments, including in hospital, do not contribute to transmission.

We modeled a stochastic discrete-valued approximation of these dynamics based on an Eulerian numerical solution approach [2, 4]. In this model, individuals move between compartments according to counting processes with transitions mathematically defined by binomial distributions and multinomial distributions with exponential transition probabilities. Binomial distributions are used to determine the numbers of individuals leaving a given compartment when there is only one direction to go in, and multinomial distributions are used when individuals leaving a compartment are distributed across multiple compartments. The model is fully stochastic in that for a fixed set of parameter values and initial conditions, it will generate an ensemble of epidemic realizations.

The basic reproductive number,  $R_0$ , is not a parameter of the system of state variables, but is function of model parameters including the transmission rate  $\beta$ . An expression for the relationship between  $R_0$ ,  $\beta$ , and other model parameters was derived. Because our model involves multiple infectious classes with different rates of infection duration, we use the *Next Generation Matrix* (NGM) [5] approach, computing  $R_0$  by linearizing the dynamic system of equations around the infection-free steady state and identifying conditions that guarantee increases in infected states [5].

We introduce a time-varying parameter,  $\mu(t)$ , representing modification to the transmission rate and equivalently  $R_0$  (since  $R_0$  is directly proportional to  $\beta$ ), which we call the *transmission reduction factor*. The transmission reduction factor allows us to estimate changes in the transmission rate over time and to explicitly model changes in the transmission rate to simulate different intervention scenarios. We define the time-varying effective reproductive number as  $R(t) = \mu(t)R_0$ , i.e. the initial value of the basic reproductive number at the beginning of the pandemic period before interventions multiplied by the time-varying transmission reduction factor.

### 2.1.1. Parameter Estimation

All transition rate parameters, e.g. the time between exposure and infectiousness, or between hospitalization and recovery, are modeled as fixed values taken directly from published literature (Table ??). The model has ten unknown parameters that we estimate with data. This includes three parameters

relating to virus transmission:  $R_0$ , which we use with the analytical expression to get the transmission rate  $\beta$ ; the transmission rate reduction factor  $\mu(t)$ ; and the illness detection rate  $r$ , or the fraction of all infections that will be observed (vs. unobserved). Four time-varying parameters relate to conditional probabilities of each stage of disease:  $\alpha(t)$ , the probability an infected individual will require hospitalization;  $\kappa(t)$ , the probability a hospitalized patient will require ICU care;  $p_V(t)$ , the probability a patient in ICU care will require ventilation; and  $\delta(t)$ , the probably an ICU patient will die. We also estimate the start time of the epidemic.

We use Approximate Bayesian Computation (ABC) techniques to estimate parameters. We apply the R package EasyABC and the algorithm *Marjoram*, which implements the algorithm of Marjoram et al. (2003) [1] with improvement steps by Wegmann et al. (2009) [2]. We set 100,000 burn-in simulations, 1000 sets of jointly estimated parameter values to save. 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.

We use multiple data sources to specify informative prior parameter distributions where sufficient information allows. We base the prior distribution for  $R_0$  on values estimated from other published studies on COVID-19. We use geolocation trace data from smartphones, i.e. mobility data, to inform both the magnitude and the timing of inflection points in the transmission reduction factor  $\mu(t)$ . We incorporate data for LAC from a dashboard provided by Unacast, a company that builds and maintains anonymized mobility datasets based on aggregated geolocation data aggregated from millions of users in LAC [3]. Specifically, we include measures of reductions in distances travelled, time spent outside of home, visits to specific types of venues, and encounter rates in the community; all of which have been shown to be proxies for reduction in opportunities for transmission [4]. The mobility data informs the timescale and shape of the linear function modeling the transmission reduction factor  $\mu(t)$ , which is estimated at the observed inflection points on dates March 27 and July 5, 2020 (see Supplemental Figure ??). We use the results of a seroprevalence study conducted on a representative sample of the LAC population on April 10-11, 2020 by researchers at the University of Southern California (USC) and the Los Angeles County Department of Public Health (LACDPH) [5] to inform the prior distribution for  $r$ , the fraction of observed illnesses over all illnesses. The study found the prevalence of COVID-19 antibodies in 4.65% (95%CI: 2.8%, 5.6%) of the sampled population. We use the lowest bound estimate from this study to remain more consistent with other published studies on the percentage of unobserved illnesses [6]. We use these observations together with the number of observed recovered illnesses at the same date (which we estimate as a time shift from observed illnesses) to inform the prior distribution for  $r$ . We use a uniform distribution centered around this value, with a wider confidence then identified in the study.

We estimate the probabilities of each disease stage  $\alpha(t)$ ,  $\kappa(t)$ , and  $\delta(t)$  in two phases reflecting differences in the composition of the infected population

by age, the only risk factor modeled in this study for which observed (marginal) data is available. From the beginning of the epidemic in LAC in early March through approximately April 20, the fraction of illnesses made up by individuals 65 years and above ( $65+$ ) increased to approximately 23% and the fraction of illnesses made up by individuals less than 18 years ( $< 18$ ) remained stable at around 1-2%. From April 20 until July 20, the trend in the illness distribution by age reversed, with individuals  $65+$  decreasing to 12% and individuals  $< 18$  increasing to 9%. From July 20 through mid-September this relative age distribution has remained stable [3]. There is a delay in the effect of the changing profile of infections by age on the population-average probabilities of severe illness and death, as the infected population advances across the stages of disease. Qualitative model best-fit analysis was used to identify the two inflection points representing the approximate beginning of change in the probability of each stage of disease and death, and the approximate end to the decrease and beginning of the stabilization of the probabilities. The probability parameters were estimated on April 20, a time point representing initial values before the decrease began, and July 20, a time point representing second set of values after the decrease. Prior distributions for each probability on each date were modeled as normal distributions with means informed by the ratios of observed numbers of illnesses, hospitalizations, and deaths in LAC.

### 2.1.2. Data

The stochastic epidemiological model was fit with prior parameter distributions to multiple data sources on counts in LAC, with a few modifications. We use both the daily and cumulative count of numbers of infected (observed), hospitalized, ventilated, and dead individuals in LAC, using a data set from the Los Angeles County Department of Public Health, updated daily by the COVID-19 Outbreak Data Coordination Team. To improve reliability of the data used for estimation we do not consider unstable early counts and use observations beginning March 15, 2020 for illness, hospitalization, and ventilation counts, and March 30, 2020 for death counts.

We also use counts of recovered individuals in model fitting. These are estimated using the seroprevalence study conducted by USC and LACDPH [8] as a single data point for the number of counts in the recovered compartment as of a week before the seroprevalence study date (approximately the time it takes for an antibody response to develop []). Counts of recovered individuals equal to zero at the start time of the epidemic (a model parameter) up until this date are interpolated with an exponential curve and used in the model fitting procedure.

We remove an estimated number of counts from each compartment that are attributable to residents of Skilled Nursing Facilities (SNFs), which include nursing homes. This processing step was taken because a single population model is not capable of addressing the disparate risk of exposure and severe illness experienced by this group. Data on counts of illnesses and deaths for individuals in SNF facilities are available via the LACDPH COVID-19 dashboard.

We estimate numbers of hospitalized and ventilated individuals from the death counts.

### 2.1.3. Uncertainty quantification

Using ABC on multiple parameters simultaneously produces joint posterior estimates over all parameters rather than the marginal posterior of a single parameter. In forward simulations of our model, we simulate trajectories with parameter values coming from this joint posterior distribution, thereby simulating trajectories across distributions of likely parameter values.

To provide uncertainty estimates, we aggregate simulations from 1000 estimated parameter sets. For each parameter set, we aggregate simulations for 500 stochastic realizations. We pool together predictions across all parameter sets from their stochastic realizations, and report their mean and 2.5th/97.5th percentiles, and/or the 25th/75th percentiles, which we call credible intervals.

This procedure quantifies 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.

## 2.2. Risk model

Using studies reporting the risk of severe COVID-19 outcomes given individual risk factors, we construct a logistic regression model to estimate the probability of COVID-19 illness trajectories for individuals with combinations of risk factors. Specifically, we estimate three models for: (1) the probability that an individual is admitted to hospital given acquired (observed) illness  $Pr(H|I)$ , (2) the probability the individual is admitted to the ICU given admittance to hospital  $Pr(Q|H)$ , and (3) the probability that the individual dies given being admitted to the ICU  $Pr(D|Q)$ . These probabilities correspond to the epidemic model estimated parameters  $\alpha$ ,  $\kappa$ , and  $\delta$ , respectively. Each of these regression models includes indicator variables for the presence or absence of specific risk factors.

The risk factors included in our analysis are age, body mass index (BMI), smoking status, and any comorbidity. The comorbidities included are diabetes, hypertension, chronic obstructive pulmonary disease (COPD), hepatitis B, coronary heart disease, stroke, cancer and chronic kidney disease. We modeled age and BMI as an ordinal variable and assume an additive effect of both age and BMI on the three outcomes. Age was categorized within four groups: 0 – 19, 20 – 44, 45 – 64, and 65+, and 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 status are modeled as binary variables.

Since we do not have access to individual-level data, we estimate the conditional risk effects corresponding to each risk factor within each logistic regression using marginal effect estimates available from reported studies and a method called the joint analysis of marginal summary statistics (JAM) [10]. This method was originally developed for use in genome-wide analyses to obtain conditional effect estimates for a trait with joint combinations of single

nucleotide polymorphisms (SNPs). The method uses two pieces of information: (i) the marginal effect estimates between single SNPs and the phenotype and (ii) the correlation structure between the SNPs. We apply the JAM model to estimate the conditional log relative risks (RR) of COVID-19 illness severity (hospitalization, ICU, and death) given joint combinations of risk factors [10].  
 300 For information informing (i) we obtain the marginal log RR between individual risk factors and COVID-19 illness severity from published COVID-19 studies, peer-reviewed where available [11, 12] (see left column of Table 1). For (ii), we obtain the correlation structure between the risk factors using data from The  
 305 National Health and Nutrition Examination Survey (NHANES) [13]. NHANES is a survey research program conducted by the National Center for Health Statistics (NCHS) 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 the correlation separately for each race/ethnicity group (Latinx, White, African American, Asian, and Other) and weighted the correlation matrix by the distribution of each race/ethnicity in LAC (Supplement Section ??).

To construct a model to estimate  $Pr(H|I)$ ,  $Pr(Q|H)$ , and  $Pr(D|Q)$ , we  
 315 combine in a logistic regression framework all (plausible) linear combinations of the risk factors specified in a mean centered design matrix,  $\mathbf{X}$ , their corresponding conditional log RR obtained from JAM,  $\hat{\psi}$ , with intercepts that are set to the estimated probabilities from the epidemic model (see Section 2.1) for  $\hat{\alpha}$ ,  $\hat{\kappa}$ ,  $\hat{\delta}$ , respectively. For example, to estimate the vector of probabilities of hospitalization given illness for all risk profiles we use  $\widehat{Pr(H|I)} = \text{expit}(\hat{\alpha} + \mathbf{X}\hat{\psi})$ .  
 320 Patients age < 20, are assumed to be the baseline, assuming the risk factors non-smoker,  $BMI < 30\frac{kg}{m^2}$ , and no comorbidity.

The estimation of the three probability models requires estimates of the frequency of each risk profile at each stage of disease. The distribution over  
 325 infections, the first stage of disease, was informed by observed LAC data on the prevalence of each of the four age groups over illnesses and an estimate of the frequency of each risk profile within the overall population. The latter was produced by simulating a sample population for LAC based on the prevalence of each marginal risk factor in LAC from multiple data sources and the  
 330 weighted correlation structure between the risk factors obtained from NHANES data. The frequency of each age group over illnesses was then stratified across the risk profiles according to their relative frequency within each age group in the overall LAC population. The frequency of each risk profile in subsequent stages of disease was found 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.

### 2.3. Calculating Risk Profile Specific CFR /IFR

To calculate the risk-stratified CFR/IFR, the frequency of each risk profile in the illness population and the estimated frequency of each risk profile in the

340 deceased population (obtained from the risk model) is multiplied by each value  
 341 of the estimated cumulative number of observed infections ( $I$ ), total infections  
 342 ( $A$ ), and deaths ( $D$ ). We find the CFR and IFR for each model realization as  
 343 the number of deaths over observed infections, and number of deaths over total  
 344 infections, respectively. Repeating across the model realizations achieves the  
 345 95% credible intervals.

#### 2.4. Scenario Analysis

We implement scenarios modeling changes in the community transmission  
 346 rate at different times, and isolating specific populations at different times. We  
 347 operate on the community transmission rate by increasing or decreasing the  
 348 value of the reproductive number  $R(t)$  over time, reflecting different levels of  
 349 non-pharmaceutical interventions (NPIs), which could include measures such as  
 350 physical distancing and mask adherence. Specifically, we model three levels of  
 351 community transmission: (1)  $NPIs=Observed$  implements the epidemic model  
 352 estimated observed  $R(t)$  throughout the epidemic in LAC.(2)  $NPIs=Moderate$   
 353 implements an  $R(t)$  equal to the epidemic model-estimated value in Stage III of  
 354 the epidemic in LAC (described in Section 3.1). In this scenario  $R(t)$  decreases  
 355 from  $R_0$  to the Stage III value between March 12 and March 27, following the  
 356 observed timeline of the drop between the initial  $R_0$  and the estimated lowest  
 357 value for  $R(t)$  observed during Stage I lockdown. (3)  $NPIs=None$  implements  
 358  $R(t) = R_0$  throughout the modelled period March - July representing a scenario  
 359 in which no actions are taken to reduce the native  $R_0$  observed at the beginning  
 360 of the epidemic in LAC. The trend of  $R(t)$  modeled for the three levels is shown  
 361 in Figure 7a.

We simulate the effects of protecting the elderly by calculating the values  
 362 of the probabilities of each stage of disease given infection ( $\alpha(t)$ ,  $\kappa(t)$ , and  
 363  $\delta(t)$ ) over time with the subpopulation of individuals 65+ removed. We model  
 364 three levels: (1)  $Protect=Observed$  models the observed trend in the proba-  
 365 bilities of severe illness for LAC, which decreased between April and July fol-  
 366 lowing the estimated distribution described in Section 3.1. We call this the  
 367 “ $Protect=Observed$ ” rather than “ $Protect=None$ ” level because it follows the  
 368 observed reduction of individuals 65+ in the infected population distribution,  
 369 which decreased from 23% to 12% between April 20 and July 20 (as described in  
 370 Section 2.1.1), indicating behaviorally-adapted protection of this age group. (2)  
 371  $Protect=High$  isolates all individuals 65+, approximately 11% of the LAC pop-  
 372 ulation, from general community transmission and subsequent stages of disease.  
 373 We implement this scenario by calculating the population-average  $\alpha(t)$ ,  $\kappa(t)$ ,  
 374 and  $\delta(t)$  on April 20 and July 20 with individuals 65+ removed. The scenario  
 375 is informed by observed LAC data on the distribution of each age group over  
 376 illnesses such that the frequency of each risk profile in illnesses reflects the ob-  
 377 served age distribution (as described in Section 2.2). The frequency distribution  
 378 over the profiles is then renormalized after removing all profiles that include in-  
 379 dividuals 65+. With the renormalized frequency distribution in illnesses, we use  
 380 the risk model (Section 2.2) to find the population-average adjusted  $\alpha(t)_{Protect}$ ,  
 381  $\kappa(t)_{Protect}$ , and  $\delta(t)_{Protect}$ . The timeline follows that for the observed trend in

<sup>385</sup>  $\alpha(t)$ ,  $\kappa(t)$ , and  $\delta(t)$ . (3) *Protect=Moderate* isolates 50% of individuals 65+ from community transmission and subsequent stages of disease. Similar as for 100% protection, we remove 50% of individuals 65+ from the observed distribution of each age group over illnesses on each date, and calculate the population-average adjusted  $\alpha(t)_{Protect}$ ,  $\kappa(t)_{Protect}$ , and  $\delta(t)_{Protect}$ . The trend of the probabilities of each stage of disease given infection modeled for the three levels is shown in Figure 7b.

<sup>390</sup> We implement eight scenarios representing most combinations between the three NPI and three Protect settings. For each scenario, we simulate the model with the estimated parameter values for  $R0$ ,  $r$ ,  $T0$ , and  $p_V$  with the remaining parameters implemented as described above. The model is simulated over the 1000 values of the joint posterior distribution of the estimated parameters with 500 stochastic realizations as described in Section 2.1.3.

### 3. Results

#### 3.1. Model estimates and epidemic trends in LAC

<sup>400</sup> *Model fits.* Figure 2 summarizes the epidemic model fit with COVID-19 data for LAC from March 1, 2020, through July 20, 2020 for all disease states across multiple views: New cases, representing new daily incidence; the current number in a compartment at a specific date, relevant for understanding current prevalence rates and comparing with healthcare capacity limitations; and cumulative counts until a specific time. Observed data, with the estimated number of cases attributable to nursing home residents removed, are plotted as black dots. The figures demonstrate that good model fits are achieved in all compartments across time.

<sup>410</sup> *Epidemic timecourse in LAC.* Between March 1 and July 20, 2020, the Los Angeles City Mayor's Office distinguishes between three stages of the COVID-19 epidemic in LAC relating to policy response measures implemented: Stage I, March 19 - May 7: the initial shutdown; Stage II, May 8 - June 11: the first steps towards reopening; Stage III, June 12 - July 20 (and beyond): greater reopening followed by modifications closing higher risk settings (such as bars and indoor seating in restaurants) [14]. Figures 3a-3d characterize the epidemic course in LAC relative to these stages. Figure 3a shows the time-varying reproductive number  $R(t)$  and Figure 3c the current illnesses as of date  $t$ , including both observed and total (observed and unobserved) model-estimated current cases that are ultimately observed (current observed illnesses). Figure 3b shows  $\alpha(t)$ , the probability of hospitalization given illness,  $\kappa(t)$ , probability of ICU admission given hospitalization, and  $\delta(t)$ , probability of death given admission to the ICU; and Figure 3d shows the current number of COVID-19 cases in healthcare, including hospitalizations, ICU, and ventilation on time  $t$ , and new deaths on date  $t$ .

<sup>420</sup> We estimate that the overall observation rate is  $r = 5.10\%$  (95% CI: 2.67%, 9.57%) of all illnesses observed. In the initial period of the outbreak before

public behavior began to change and policy interventions were implemented, we estimate the basic reproduction number was  $R_0 = 3.89$  (95% CI: 3.48, 4.28). During the period from March 12 to March 27, beginning just before Stage I was implemented, we estimate a reduction to an  $R(t)$  of 0.86 (95% CI: 0.72, 1.09).  
 430 The corresponding reduction in transmission led to a levelling off and later decrease in the numbers of illnesses, followed by hospitalizations, ICU admissions, and deaths. Notably, we estimate that  $R(t)$  began to rise again following increasing mobility behavior around April 27, almost two weeks before the Stage  
 435 II reopening policy was implemented, portending the increase in illnesses to follow. Throughout Stage II and the beginning of Stage III, we estimate an increasing trend in  $R(t)$ , which plateaued at 1.79 (95% CI: 1.53, 2.25) around the time the modifications (re-closures) were implemented. With the increasing  $R(t)$  during this period we estimate increases in current illnesses to levels  
 440 exceeding the initial exponential growth phase, similarly followed by hospitalizations, ICU admissions, and deaths. Importantly, even at upper 95% credible interval of the peak, the current census of cases in hospital, ICU, and ventilation remain well below the capacity limits of approximately 4,000 hospital beds and 2,245 ICU beds in LAC [3].

445 Although hospitalizations and ICU admissions continued to rise during period between Stage II and III, new deaths did not increase, and the rate of all three decreased relative to increasing illnesses. We estimate that between June 1 and July 1, the estimated inflection points between the first and second phase of the illness severity probabilities, the population-average probability of hospitalization given illness  $\alpha(t)$  decreased from 0.22 (0.20, 0.23) to 0.08 (0.07, 0.09); the probability of ICU admissions given hospitalization  $\kappa(t)$  decreased from 0.66 (0.64, 0.68) to 0.29 (CI: 0.27, 0.31); and the probability of death given ICU admission  $\delta(t)$  decreased from 0.49 (0.47, 0.50) to 0.41 (0.39, 0.43). The population-average CFR and IFR correspondingly dropped; we estimate  
 450 the CFR was 3.7% (3.2%, 4.2%) on April 20 (representing the first phase values for  $\alpha(t)$ ,  $\kappa(t)$ , and  $\delta(t)$ ) and had dropped to 1.6% (1.0%, 2.4%) by July 20 (representing the second phase). The IFR was 0.2% (0.18%, 0.23%) on April 20 and dropped to 0.09% (0.06%, 0.13%) by July 20.

### 3.2. Risk-stratified probabilities of severe illness and death for LAC

460 Table I displays the marginal relative risks (RR) extracted from the literature (left column) and conditional RR estimated by the risk model of each risk factor considered by our model (age, comorbidities, smoking status, obesity status) on the rates of hospitalization given illness, ( $H|I$ ), ICU admission given hospitalization, ( $Q|H$ ), and death given ICU admission, ( $D|Q$ ). We find that  
 465 the independent effect of age is stronger than from having any comorbidities. The independent effect of comorbidities and obesity attenuate with increasing severity of disease, while that of age and smoking increase.

470 Using the conditional RR together with the population-average probabilities of each stage of disease estimated by the epidemic model and the observed distribution of illnesses by age in LAC, we estimated the probability of each stage of disease given infection for all risk profiles. Probabilities are estimated for the

Risk Factors	Marginal RR (95% CI)	Conditional RR (95% CI)
$(H I)$		
Ordinal Age	2.11 (1.88, 2.37)	1.7 (0.67, 4.28)
Ordinal BMI	2.98 (2.61, 3.39)	1.82 (1.06, 3.15)
Smoker	1.40 (0.90, 2.17)	1.76 (0.21, 14.52)
Any comorbidity	3.18 (2.42, 4.18)	1.50 (0.59, 3.84)
$(Q H)$		
Ordinal Age	1.71 (1.40, 2.08)	1.54 (1.23, 1.92)
Ordinal BMI	1.01 (0.86, 1.18)	1.05 (0.65, 1.69)
Smoker	1.71 (0.87, 3.38)	1.61 (1.45, 1.79)
Any comorbidity	1.34 (0.87, 2.06)	1.02 (0.86, 1.20)
$(D Q)$		
Ordinal Age	4.29 (2.50, 7.34)	2.42 (1.70, 3.44)
Ordinal BMI	1 <sup>†</sup>	1.12 (0.73, 1.71)
Smoker	1 <sup>†</sup>	1.96 (1.33, 2.89)
Any comorbidity	1.64 (0.81, 3.32)	1.05 (0.78, 1.43)

**Table 1:** The marginal relative risk collected from published studies on COVID-19 and conditional relative risk estimated by the risk model of each risk factor on rates of hospitalization given illness,  $(H|I)$ , ICU admission given hospitalization,  $(Q|H)$  and death given ICU admission,  $(D|Q)$  (95% credible interval).

<sup>†</sup>We set the marginal RR for ordinal BMI and smoker to 1 because we did not find the association between obesity class, smoking status, and the likelihood of death given ICU admission  $D|Q$  due to COVID-19 in the published literature.

dates April 20, representing the initial set of values for these probabilities, and July 20, representing the plateaued value of the probabilities after decreasing (see Figure 3b). Supporting Table 4 shows for each profile the model-estimated population prevalence, the frequency in the illness population on April 20 and July 20, and the probability of hospitalization given illness, ICU admission given hospitalization, and death given admission to the ICU on April 20 and July 20.

The probability of hospitalization given illness, ICU admission given hospitalization, and death given admission to the ICU vary extensively across the risk profiles. Notably, the risk within specific marginal factor groups also varies extensively. For example, within the age group of 65+, the probability of hospitalization given infection is 3 and 4 times higher on April 20 and July 20, respectively, for individuals with at least one comorbidity, a smoking history, and morbid obesity than for individuals that have no comorbidities, do not smoke, and have a healthy BMI. The decreases in probabilities between April 20 and July 20 are not equally distributed across all profiles but increase conversely with risk such that the largest ratio decreases are estimated for the lowest risk profiles.

*Risk-stratified CFR and IFR for LAC.* A key result from the combined epidemic and risk factor model are estimates of risk-profile specific CFR and IFR. Table 5 shows the model-estimated population prevalence, the frequency of each profile in the illness population on April 20 and July 20, and the median of the CFR and IFR on April 20 and July 20 for each profile. Profiles with a population prevalence lower than 0.000001 do not yield well-defined CFR or IFR and are not included.

The probability of death given infection varies tremendously across populations; while the population-average CFR was 3.7% (3.2%, 4.2%) on April 20, the profile-stratified CFR ranged from 0 to 17.7% (15.2%, 19.6%). On July 20, the population-average CFR was 1.6% (1.0%, 2.4%), and ranged from 0 to 13.6% (9.4%, 20.8%). Given the model-estimated proportion of  $r = 5.1\%$  of total illnesses being observed, the IFR are substantially lower than the CFR, with the highest (median) IFR being 0.96% (0.83%, 1.1%) on April 20 and 0.74% (0.53%, 1.0%) on July 20, compared with a CFR of 17.7% (15.2%, 19.6%) for the same highest-risk profile.

To facilitate interpretation of the probabilities and variability across risk profiles, we group the profiles into 5 Risk Groups based on similar within-group CFR on April 20, with Risk 1 being composed of individuals with  $CFR > 10\%$ ;  $6\% < CFR < 10\%$  in Risk 2;  $3\% < CFR < 6\%$  in Risk 3;  $0.1\% < CFR < 3\%$  in Risk 4; and  $P(D|I) < 0.1\%$  in Risk 5 (risk groups are indicated for each profile in Table 5). All profiles including individuals 65 years and above are found in Risk 1 and 2, whereas profiles for individuals 45-64 are found in Risk 1-4.

### 3.3. Group-stratified population at each stage of disease

Figure 6 shows the model estimated frequency of risk factor groups in the population of individuals in each stage of disease: infected, hospitalized, ad-

mitted to ICU, and deceased, compared to their frequency in the overall LAC population (left bar), on April 20 and July 20, for: (A) Risk Groups, (B) age group (C) group of individuals with any comorbidity vs. no comorbidity, and (D) group of individuals who smoke vs. do not smoke. The population considered does not include individuals in SNF facilities.

We estimate that the change in the composition of infections by age (discussed in Section 2.1.1) results in a change in composition of the deceased population (not including fatalities coming from SNFs) of 53% individuals over 65 years and 36% individuals 45-64 years on April 20, to approximately 40% individuals over 65 years and 47% individuals 45-64 on July 20. Individuals 65+ comprise a fraction of total deaths 4.8 and 3.6 times greater on April 20 and July 20, respectively, than their prevalence in the overall population, whereas individuals 45-64 comprise a fraction equal to 1 and 1.3 times greater on April 20 and July 20 than their overall population prevalence. The percentage of individuals with comorbidities in the deceased population remains stable at 72 - 73%. 1.5 times as many individuals with comorbidities are in the deceased population then in the overall population on both dates.

Comparing across increasing stages of disease, the figures illustrate the conditional effect each risk factor has on the overall population. For example, there is a larger effect of age 65+ on the probability of death given ICU admission than the probability of hospitalization given illness or ICU admission given hospitalization; consequently, the proportion increase of individuals 65+ between the in-ICU and the deceased population is more than twice as large as the proportion increase between the hospitalized and in-ICU population. Comorbidities have a larger effect on the probability of hospitalization given illness than death given hospitalization and there is a large increase in the proportion of individuals in this group between the infected and hospitalized populations. The proportion increase between the in-ICU and deceased population is much lower than for advanced age.

#### 3.4. Scenario analysis

Figure 7a shows the three levels of NPI policies implemented (none, moderate, observed) and Figure 7b shows the three levels of isolating the elderly implemented (observed, 50%, 100%) in Scenarios 1-8.

Figure 8 shows the results of the eight scenarios. Doing nothing to decrease  $R_0$  or protect at-risk populations, the “Do Nothing” Scenario (Scenario 1), results in a median of approximately 30,000 deaths by July 20, exceeding the observed death count in LAC by 13 times the median value of 2,300, as well as severely exceeding hospital capacity for two months with a peak of 22,000 individuals requiring daily hospital care. Combining a stringent protection policy that isolates 100% of individuals 65+ with the “Do Nothing” to reduce community transmission policy (Scenario 2) averts approximately 50% of deaths estimated in the “Do Nothing” Scenario, however surpasses observed deaths in LAC by a factor of 6 and exceeds hospital capacity for one month at peak levels over 100% of maximum capacity. Implementing a moderate NPI policy that reduces community transmission to levels observed in July while taking no

additional measures to protect at-risk populations (Scenario 3) greatly reduces deaths from the scenarios in which no community transmission reduction takes place, leading to a 3,800 deaths by July 20, a factor of 1.6 greater than observed. However, this scenario approaches or exceeds healthcare capacity for over one month.

565

Whe the moderate NPI policy is combined with additional measures to protect 100% of individuals 65+ (Scenario 4), the number of deaths accrued would have approximated the observed death toll in LAC by July 20, while hospital demand would have still approached levels exceeding capacity. If a more feasible 570 50% of individuals 65+ had been protected (Scenario 5), the death count would have exceeded the observed count by approximately 1000 and hospitalization demand would come closer to exceeding capacity.

575

Comparing Scenarios 6, 7, and 8 in which the observed trend in community transmission is implemented (strict initial lockdown followed by a gradual re-opening and stabilization at an  $R(t)$  of approximately 1.8), deaths could have been reduced by a median of 300 deaths if 100% of individuals 65+ had been protected (Scenario 7) or 100 deaths if 50% of individuals 65+ in the observed count had been protected (Scenario 8). Hospitalization demand would not have decreased appreciably.

580

#### 4. Discussion

585

This work has developed a framework for using available COVID-19 epidemic dynamic data and prevalences of risk factors at the population level to estimate subpopulation-stratified probabilities of severe illness and CFR and IFR across multiple combinations of risk factors for Los Angeles County from the period March 1 - July 20, 2020. In the absence of individual level data, the technical contribution of this work was to integrate a dynamic epidemic model with a modeling approach used in genetics to estimate conditional effects from available marginal data to produce time-varying subpopulation-stratified estimates for LAC. On its own, the risk model provides adjusted risk estimates 590 for age, smoking, BMI, and comorbidities and provides a better understanding of the impact of combinations of factors on severe outcomes from COVID-19, more refinement as to which individuals are at highest risk of advancing to each stage of disease. The integrated model allows us to analyze dynamic outcomes and parameters by population and subpopulation. Analysis demonstrates that the risk of severe illness and death varies tremendously across populations, and that it is insufficient to reduce the variation in risk to single, summary measures 595 for the entire population. For example, within the aggregate group of individuals 65+ years old, the IFR is 12 times greater for individuals with at least one comorbidity, a smoking history, and morbid obesity than for individuals that have no comorbidities, do not smoke, and have a healthy BMI.

600

The fact that the risk of severe illness and the CFR/IFR vary tremendously across populations raises the question of whether protecting certain subpopulations would have achieved similar results in illnesses, hospitalizations, and deaths as policies relating to reductions to community transmission, including

605 the initial lockdown period of from March 19 to May 8. We simulated the effects  
of alternate scenarios relating to policies modifying community transmission and  
protecting specific subpopulations to compare with observed trends in LAC. The  
main finding from the scenario analysis is that the strict initial lockdown period  
saved many lives because it both reduced transmission and protected individ-  
610 uals at greater risk. Doing nothing to decrease transmission or protect at-risk  
populations, the “Do Nothing” Scenario, would have resulted in devastating  
consequences, surpassing the observed death toll by 13 times. Protecting the  
elderly while doing nothing to decrease transmission would still have surpassed  
the observed death toll by 6 times and significantly overwhelmed healthcare  
615 capacity for months. However these scenarios are unrealistic because individual  
behavior would likely have modified the transmission rate, as observed in other  
countries where shelter-at-home policies were not formally implemented. We  
therefore implemented a more moderate NPI policy that reduced community  
transmission at the beginning of the epidemic to levels exhibited during Phase  
620 3 in LAC, which in practice reflected panel of public health outreach and restriction  
policies including mask wearing, schools and indoor dining/entertainment being closed,  
but not as stringent as the more complete community lock-down of Stage I. Simulation  
results suggest that this policy would not have been sufficient to avoid overwhelming  
625 healthcare capacity, and additionally would have exceeded the observed death count by over 100%. If this moderate NPI policy were combined with stringent measures to protect all individuals over 65 years,  
the death count would have looked similar to the observed trends, while hospital  
capacity limits would have been approached, likely resulting in more severe  
illnesses and deaths then accounted for by this model as well as overwhelming  
630 an already severely overtaxed healthcare system. A more realistic scenario in  
which 50% of individuals over 65 years were protected would have surpassed the  
observed death toll and similarly threatened to exceed healthcare capacity.

As we continue to reopen society, this analysis illustrates the effectiveness  
635 of more nuanced policies that can better target at-risk individuals while not  
over-controlling lower risk situations and individuals. In anticipation of a surge  
in cases in the fall across the country [], to prevent critical illnesses and deaths  
it may be more effective to ramp up efforts to protect individuals of advanced  
age and other at-risk populations while continuing to maintain more moderate  
640 community-level NPI restrictions. The modeled scenarios illustrate that these  
protection measures, including more moderate scenarios that have 50% efficacy  
at protecting the at-risk subpopulation of individuals of advanced age, can be  
as effective in preventing critical illnesses and deaths as untargeted society-wide  
NPI policies to restrict community transmission.

We comment on a few estimated model parameters in context. We identified  
645 two stages of the epidemic in the larger LAC population relative to the probabilities  
of severe illness and death: a first phase between March 1 and June 1, a decrease  
in probabilities between June 1 and July 1, and a second stage from July 1 to July 20 (continuing through mid-September). The large decrease in  
the probability of hospitalization given illness ( $\alpha$ ) by almost a factor of three  
650 suggests substantive changes in the infected population. This follows from the

halving in the prevalence of individuals 65+ in illnesses from 23% on April 20 to 12% on July 20, but may also reflect other changes in the demographic distribution of individuals at risk or with access to the healthcare system. The decrease in the probability of ICU admission given hospitalization ( $\kappa$ ) by more than half  
655 suggests better treatment in healthcare. The probability of death given hospitalization ( $\delta$ ) decreased marginally in comparison with  $\alpha$  and  $\kappa$ , suggesting that for the individuals that make it through to critical illness, their probability of death has not changed.

Our estimate for the fraction of observed illnesses of  $r = 5.1\%(2.7, 9.6)$  is  
660 consistent with the findings from the major CDC COVID-19 Response Team seroprevalence study conducted across 10 locations in the U.S., which reported mean detection rates in the range of 4.5% to 16% [9]. We stress that there is substantial uncertainty in our estimate of the observation rate, due to the prior on this parameter being informed by a single data point: the seroprevalence  
665 study conducted on the LAC on April 5, 2020. This translates to an estimated prevalence of 200,000 (95 CI: 50,000, 500,000) illnesses in LAC by July 20, 2020, out of which 191,500 (95% CI: 48,000, 470,000) are unobserved. Despite the substantial uncertainty, conclusions can be drawn from these estimates. Even at the lowest estimates for the total numbers of illnesses, an overall prevalence of  
670 around 50,000 represents a very large number of infected individuals, with major implications for the effect on sustaining a high level of community transmission. The corresponding model-estimated number of recovered cases including unobserved illnesses by July 15 ranges from 400,000 to 2.5 million, or 4% to 25% of the LAC population, which at either extreme is far from herd immunity.

The estimate for the initial basic reproduction number  $R_0$  of 3.89 (3.48, 675 4.28) is consistent with COVID-19 SEIR model projections in other settings, which have reported  $R_0$  over a range 1.40 to 6.49 with a mean of 3.28 [15] due to different settings, data sources, and modeling and statistical methods, noting that values for  $R_0$  estimated for SEIR models are consistently and significantly higher than for SIR models, as has been demonstrated by previous  
680 researchers [16]; for example, Bertozzi et al. [17] demonstrate an estimated  $R_0$  for California of 2.4 using a SIR model and 4.9 using an SEIR model. Within the context of the epidemic trajectory in LAC, a key finding of our work is that  $R(t)$  began to increase even before Stage II policy was implemented to further open the community. A closer integration between epidemic estimates of  $R(t)$  and policy measures could have helped to reduce the surge of cases observed with Stage II and III reopening and could potentially be helpful in the future. At the same time, we estimate that the measures taken in the period March 1 -  
685 July 20 2020 had large success, preventing the capacity of the healthcare system from becoming severely overwhelmed and averting approximately 28,000 total deaths (the difference between the “Do Nothing” to decrease  $R_0$  Scenario 1 and the observed death rate). This estimate is comparable to the 32,218 deaths reached in New York City by July 20, 2020 [].

This study is prone to typical limitations occurring in modeling epidemiological dynamics in the context of rapidly evolving infectious disease outbreaks.  
695 The accuracy of the parameter estimation depends on the accuracy of the prior

specifications and assumptions employed. Such limitations include biases in how cases are observed and reported over time, including in who is getting tested and at what stage of disease. While we include uncertainty, transition time parameters, including the incubation and recovery periods, are modeled as fixed values based on studies from China and Italy. For the model-estimated IFR, it is important to interpret these in context of the multiple embedded assumptions that are included in the estimation of the IFR, which include the assumptions in the case observation rate  $r$ , in the risk model and prior estimates of each risk profile, and the assumption that unobserved illnesses will be equally distributed across the population; as well as the assumptions embedded in the risk model.

While the risk model is specified to the LAC region through the integration of available regional data on (i) the population prevalence of risk factors (ii) and observed epidemic timeseries data on illness, hospitalization, and death counts, and (iii) subpopulation-level timeseries data on illness counts by age strata, it utilizes data unspecific to the region. The data used to produce estimates of the relative prevalence of each stratified risk profiles comes from early published clinical studies coming from China on the fractions of hospitalization, ICU admission, and death by individual risk factors. The studies used are retrospective, whereas the most reliable studies often follow people over time (i.e. prospective). Additionally, estimates of the distribution of each risk profile in the infected population are pinned to the prevalence of each age group in the infected population for LAC, combined with our model-estimated prevalence of each individual risk profile within each age group in the overall LAC population. This was our only way of pinning the infected population to observed data because more detailed data on the prevalence of each individual combination of risk factors was not available, but could lead to errors since the relative distribution of each profile may not be equal to the prevalence of that profile given an age group.

Our risk model provides estimates of COVID-19 illness severity probabilities (probability of infection given illness) due to risk factors for COVID-19 alone (age, comorbidities, obesity, smoking). This can be seen as a “baseline” risk model, whereas a multitude of exogenous factors may impact what we observe about the epidemic, including differences in exposure dynamics, and other demographic characteristics encompassing factors such as employment, economic status, access to PPE, access to care – not only “biological” indirect risk factors from underlying conditions.

While this work has focused on demonstrating how substantially the probability of severe outcomes given infection varies across individuals and subsequent transitions between related model compartments, we developed a system dynamics that assumes homogeneous mixing and a single population. While single population models are widely used and are notably effective during times going into the peak of the disease [17], they apply critical simplifications to an extremely complex problem. Heterogeneities in exposure to COVID-19 infection, which have been shown in multiple studies to vary extensively across subpopulations both spatially and socially, were accounted for only through the observed age distribution in LAC. The dynamics of who is getting infected has become

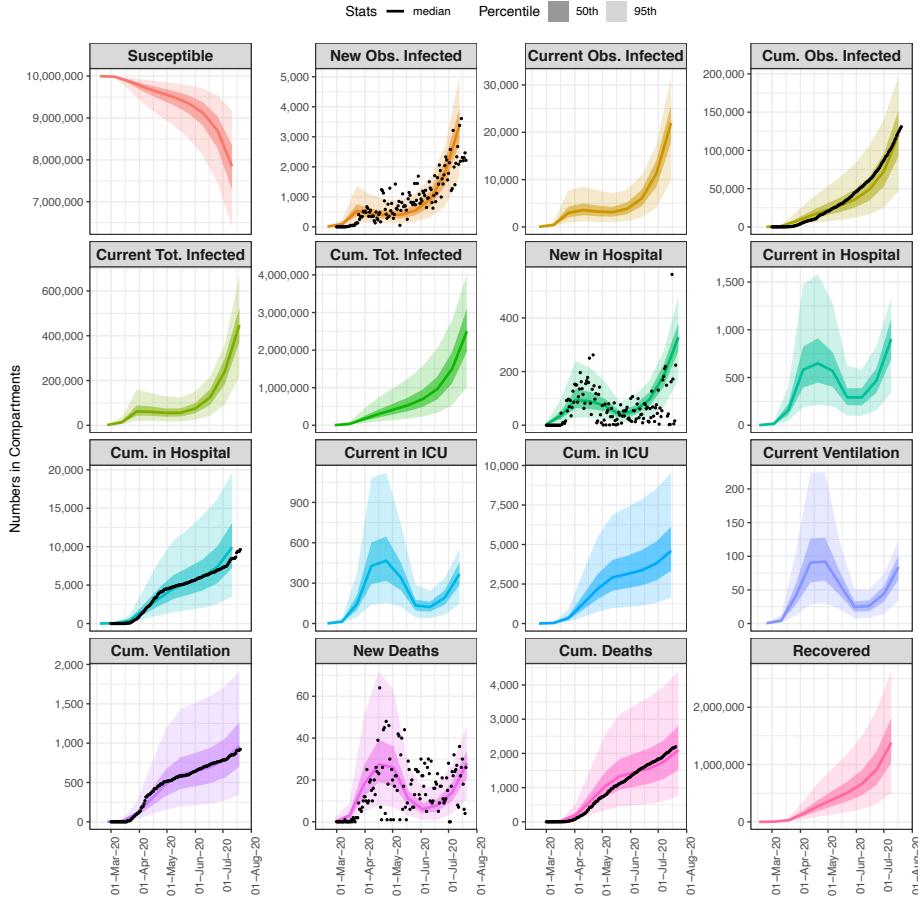
more concentrated in younger populations, as well as other race/ethnicity groups not discussed in this article. However, at the time of this study we did not have  
745 the data to formally model subpopulation-specific probabilities of exposure or the data on hospitalization and death counts for different groups necessary to fit the parameters of a multi-population model to data. The approach we developed is a way to use commonly available community-level data to model multiple groups in a single population, and use population-level estimates together with  
750 prevalence rates of risk factors to estimate stratified rates for different subpopulations. Future work will need to develop multi-population models that account for risk-stratified, time-varying probabilities of exposure, and critical illness given exposure, that are a function of location, age, and other critical risk factors such as race/ethnicity.

755 More generally, our results underscore the value and importance of including risk factors in epidemic models given the great area-level variation in risk factors and the impact that could have on an area's healthcare capacity. The framework developed here to produce subpopulation-stratified estimates of the severity of illness and CFR/IFR by subpopulations, and to understand the impact of  
760 subpopulation-specific policy interventions, can be generalized to other regional contexts, given access to epidemic time series and prevalences of marginal risk factors in the general population. In future studies, such an approach could be used to determine a portfolio of health factor priorities that would most effectively reduce disease incidence and severity.

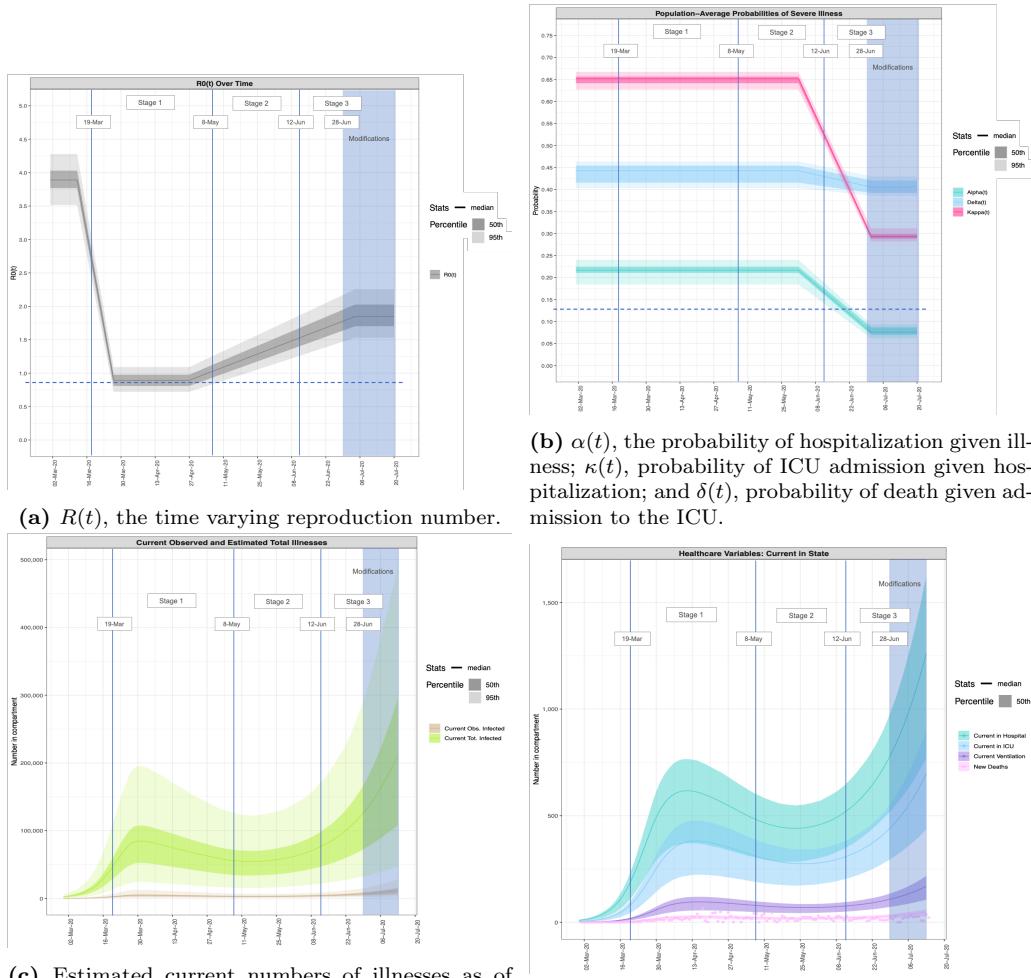
## 765 References

- [1] of Disease Prevention O, Promotion H. Healthy people 2020. 2020. URL: <https://www.healthypeople.gov/2020/>.
- [2] Keeling MJ, Rohani P. Modeling infectious diseases in humans and animals. Princeton University Press; 2011.
- 770 [3] of Public Health LACD. La county daily covid-19 data dashboard. 2020. URL: <http://publichealth.lacounty.gov/media/coronavirus/data/index.htm>
- [4] Bretó C, He D, Ionides EL, King AA, et al. Time series analysis via mechanistic models. *The Annals of Applied Statistics* 2009;3(1):319–48.
- 775 [5] Diekmann O, Heesterbeek J, Roberts MG. The construction of next-generation matrices for compartmental epidemic models. *Journal of the Royal Society Interface* 2010;7(47):873–85.
- [6] Unacast . The unacast social distancing scoreboard. 2020. URL: <https://unacast.com/post/the-unacast-social-distancing-scoreboard>
- 780 [7] Oliver N, Lepri B, Sterly H, Lambiotte R, Deletaille S, De Nadai M, et al. Mobile phone data for informing public health actions across the covid-19 pandemic life cycle. *Science* 2020;.

- 785 [8] Sood N, Simon P, Ebner P, Eichner D, Reynolds J, Bendavid E, et al. Seroprevalence of sars-cov-2-specific antibodies among adults in los angeles county, california, on april 10-11, 2020. *Jama* 2020;;
- [9] Havers FP, Reed C, Lim T, Montgomery JM, Klena JD, Hall AJ, 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;.
- 790 [10] Newcombe PJ, Conti DV, Richardson S. Jam: a scalable bayesian framework for joint analysis of marginal snp effects. *Genetic epidemiology* 2016;40(3):188–201.
- [11] Guan Wj, Ni Zy, Hu Y, Liang Wh, Ou Cq, He Jx, et al. Clinical characteristics of coronavirus disease 2019 in china. *New England journal of medicine* 2020;382(18):1708–20.
- 795 [12] Petrilli CM, Jones SA, Yang J, Rajagopalan H, O'Donnell LF, Chernyak Y, et al. Factors associated with hospitalization and critical illness among 4,103 patients with covid-19 disease in new york city. *MedRxiv* 2020;.
- [13] for Disease Control C, Prevention (CDC) NCfHSN. National health and nutrition examination survey data (nhanes). 2017-2018. URL: <https://www.cdc.gov/nchs/nhanes/index.htm>
- 800 [14] MayorOfLA . Los angeles county mayor's office covid-19 policy website. August 15, 2020. URL: <https://corona-virus.la/>.
- [15] Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of covid-19 is higher compared to sars coronavirus. *Journal of travel medicine* 2020;.
- 805 [16] Ridenhour B, Kowalik JM, Shay DK. Unraveling r 0: Considerations for public health applications. *American journal of public health* 2018;108(S6):S445–54.
- [17] Bertozzi AL, Franco E, Mohler G, Short MB, Sledge D. The challenges of modeling and forecasting the spread of covid-19. *arXiv preprint arXiv:200404741* 2020;.



**Figure 2:** Summary of the epidemic model fit with COVID-19 data for Los Angeles, for all state variables, across multiple views: New cases, representing new daily incidence; current number in a compartment at a specific date, relevant for understanding current prevalence rates and comparing with healthcare capacity limitations; and cumulative counts until a specific time. Observed data, for available compartmental variables, are shown with our model estimated reductions after removing the estimated number of nursing home related cases, and are plotted as black dots. Estimates are shown as the median number in compartments over time, with the 50% (darker) and 95% (lighter) credible intervals.



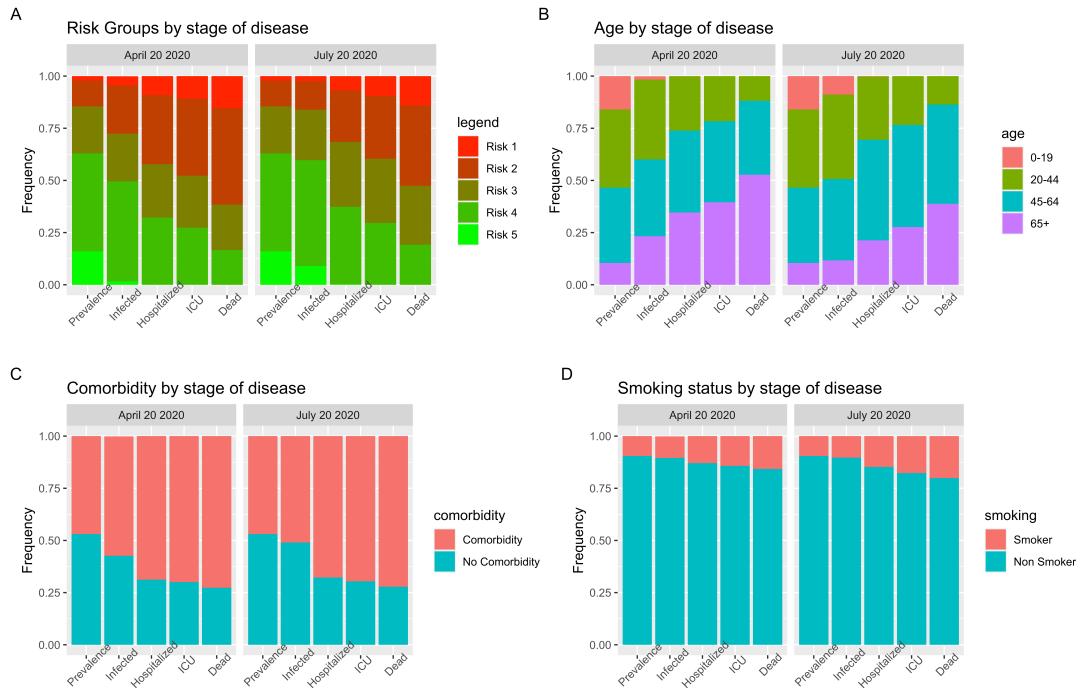
(c) Estimated current numbers of illnesses as of date  $t$ , including the current observed illnesses and the current total number of illnesses, including both cases that are observed and unobserved.

(b)  $\alpha(t)$ , the probability of hospitalization given illness;  $\kappa(t)$ , probability of ICU admission given hospitalization; and  $\delta(t)$ , probability of death given admission to the ICU.

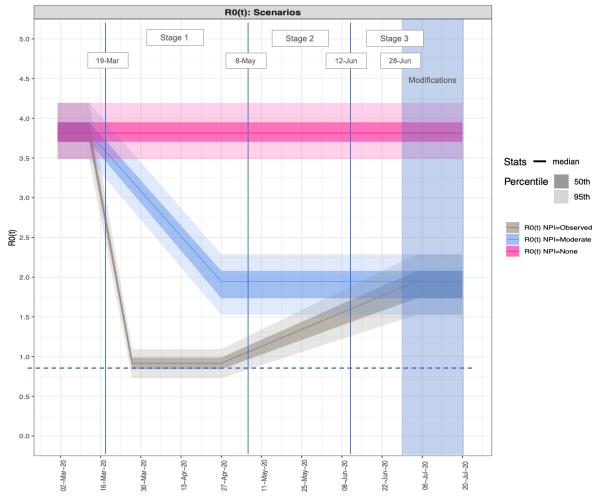
**Figure 3:** Timeseries of model-estimated parameters and compartmental variables relative to COVID-19 policy decisions in LAC. Model-estimated median curves are plotted along with the 50th% (light shading) and 95% (dark shading) credible intervals. Data is not available for current numbers but is plotted for the variable available (new deaths).



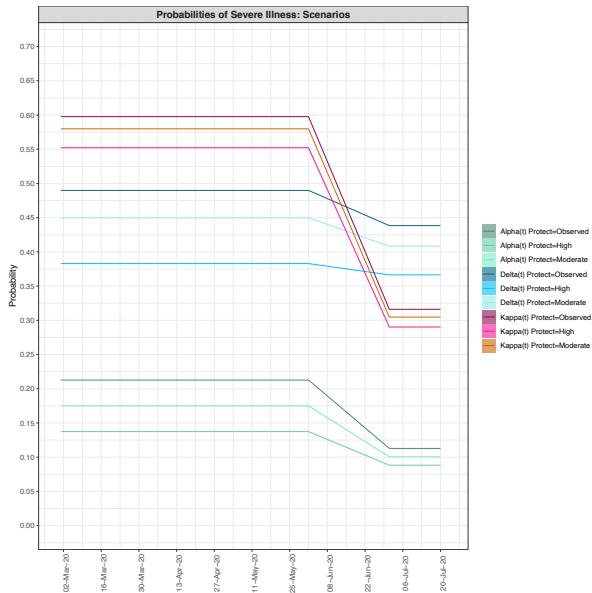




**Figure 6:** Estimated frequency of groups in the overall LAC population, and the population of individuals in each stage of disease from infected, hospitalized, admitted to ICU, to deceased on April 20 and July 20: (A) Risk Groups, (B) age group (C) any comorbidity group, and (D) smoking status group. Individuals in SNF facilities are not included in the prevalence estimates.

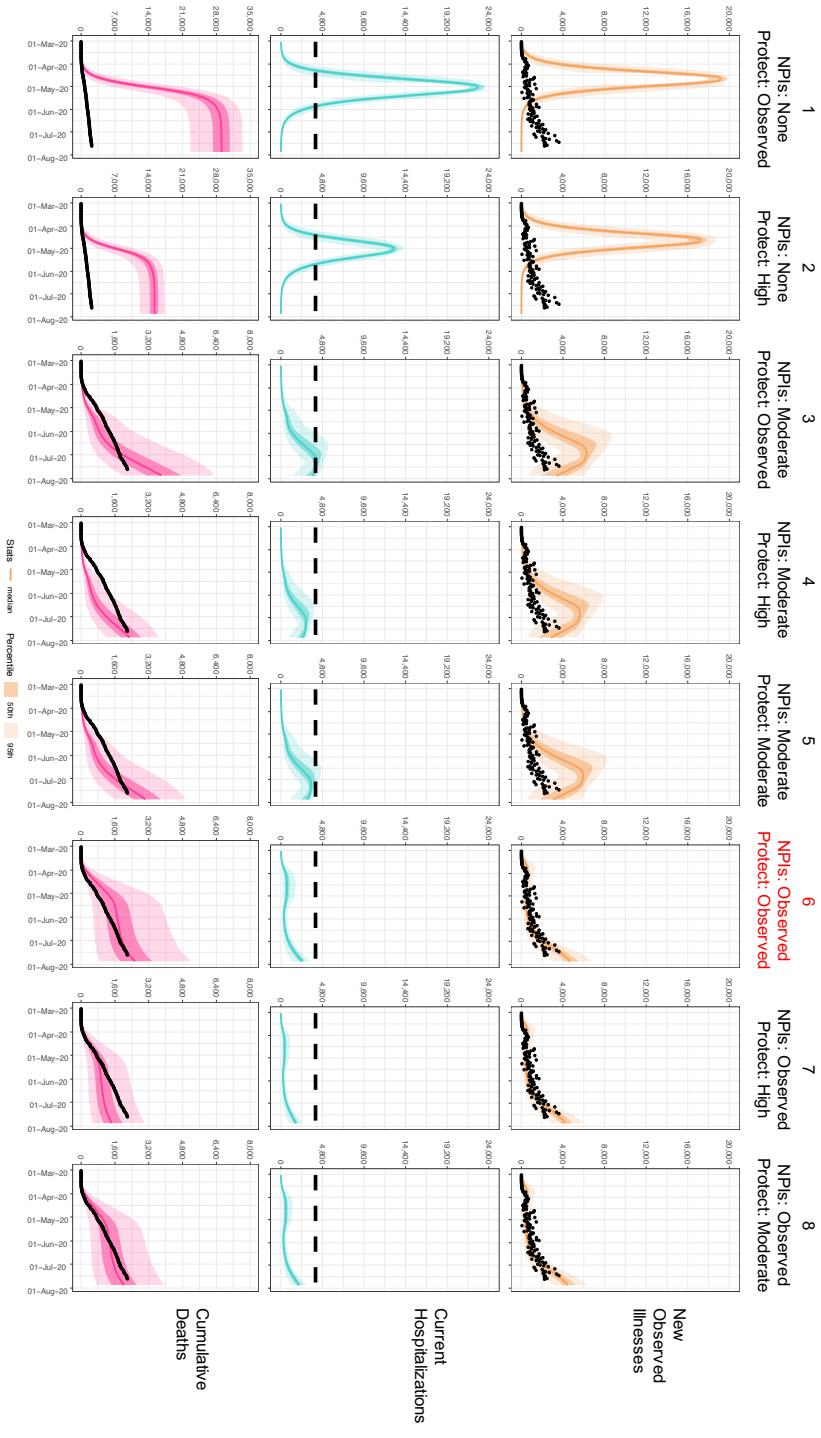


**(a)** Three levels of non-pharmaceutical-intervention (NPI) policies implemented: In NPIs=None no adaptations are made to  $R(t)$ ; in NPIs=Moderate  $R(t)$  quickly adapts from  $R_0$  to the observed  $R(t)$  during Stage III; in NPIs=Observed the observed  $R(t)$  is implemented.



**(b)** Three levels of protection of individuals 65 years and older implemented in scenario analysis: In Protect=Observed the observed  $\alpha(t), \kappa(t)$ , and  $\delta(t)$  are implemented; in Protect=High and Protect=Moderate, respectively 100% and 50% of individuals 65+ are removed from the observed illness distribution and the population-average  $\bar{\alpha}(t), \bar{\kappa}(t)$ , and  $\bar{\delta}(t)$  are recalculated.

**Figure 7:** Implemented policies in scenario analysis.



**Figure 8:** Scenarios implemented and results by current observed illnesses, current hospitalizations, and cumulative deaths.