

Estimation of COVID-19 risk-stratified epidemiological parameters and policy implications for Los Angeles County through an integrated risk and stochastic epidemiological model

Abigail L. Horn^{a,*}, Lai Jiang^a, Faith Washburn^b, Emil Hvitfeldt^a, Kayla de la Haye^a, William Nicholas^b, Paul Simon^b, Maryann Pentz^a, Wendy Cozen^a, Neeraj Sood^c, David V. Conti^a

^a*Department of Preventive Medicine
Keck School of Medicine
University of Southern California*

^b*Los Angeles County Department of Public Health
313 N Figueroa St, Los Angeles, CA 90012*

^c*Sol Price School of Public Policy, University of Southern California*

Abstract

Summary

Background. Health disparities have emerged with the COVID-19 epidemic because the risk of exposure to infection and the prevalence of risk factors for severe outcomes given infection vary within and between populations. However, estimated epidemic quantities such as rates of severe illness and death, the case fatality rate (CFR), and infection fatality rate (IFR), are often expressed in terms of aggregated population-level estimates due to the lack of epidemiological data at the refined subpopulation level. For public health policy makers to better address the pandemic, stratified estimates are necessary to investigate the potential outcomes of policy scenarios targeting specific subpopulations.

Methods. We develop a framework for using available data on the prevalence of COVID-19 risk factors (age, comorbidities, BMI, smoking status) in subpopulations, and epidemic dynamics at the population level and stratified by age, to estimate subpopulation-stratified probabilities of severe illness and the CFR (as deaths over observed infections) and IFR (as deaths over estimated total infections) across risk profiles representing all combinations of risk factors including age, comorbidities, obesity class, and smoking status. A dynamic epidemic model is integrated with a relative risk model to produce time-varying

*Corresponding author

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

subpopulation-stratified estimates. The integrated model is used to analyze dynamic outcomes and parameters by population and subpopulation, and to simulate alternate policy scenarios that protect specific at-risk subpopulations or modify the population-wide transmission rate. The model is calibrated to data from the Los Angeles County population during the period March 1 - October 15 2020.

Findings. We estimate a rate of 0.23 (95% CI: 0.13,0.33) of infections observed before April 15, which increased over the epidemic course to 0.41 (0.11,0.69). Overall population-average $IFR(t)$ estimates for LAC peaked at 0.77% (0.38%,1.15%) on May 15 and decreased to 0.55% (0.24%,0.90%) by October 15. The population-average $IFR(t)$ stratified by age group varied extensively across subprofiles representing each combination of the additional risk factors considered (comorbidities, BMI, smoking). We found median IFRs ranging from 0.009%-0.04% in the youngest age group (0-19), from 0.1%-1.8% for those aged 20-44, 0.36%-4.3% for those aged 45-64, and 1.02%-5.42% for those aged 65+. In the group aged 65+ for which the rate of unobserved infections is likely much lower, we find median CFRs in the range 4.4%-23.45%. The initial societal lockdown period avoided overwhelming healthcare capacity and greatly reduced the observed death count. In comparative scenario analysis, alternative policies in which the population-wide transmission rate is reduced to a moderate and sustainable level of non-pharmaceutical interventions (NPIs) would not have been sufficient to avoid overwhelming healthcare capacity, and additionally would have exceeded the observed death count. Combining the moderate NPI policy with stringent protection of the at-risk subpopulation of individuals 65+ would have resulted in a death count similar to observed levels, but hospital counts would have approached capacity limits.

Interpretation. The risk of severe illness and death of COVID-19 varies tremendously across subpopulations and over time, suggesting that it is inappropriate to summarize epidemiological parameters for the entire population and epidemic time period. This includes variation not only across age groups, but also within age categories combined with other risk factors analyzed in this study (comorbidities, obesity status, smoking). In the policy analysis accounting for differences in IFR across risk groups in comparing the control of infections and protection of higher risk groups, we find that the strict initial lockdown period in LAC was effective because it both reduced overall transmission and protected individuals at greater risk, resulting in preventing both healthcare overload and deaths. While similar numbers of deaths as observed in LAC could have been achieved with a more moderate NPI policy combined with greater protection of individuals 65+, this would have come at the expense of overwhelming the healthcare system. In anticipation of a continued rise in cases in LAC this winter, policy makers need to consider the trade offs of various policy options on the numbers of the overall population that may become infected, severely ill, and that die when considering policies targeted at subpopulations at greatest risk of transmitting infection and at greatest risk for developing severe outcomes.

Research in Context

Evidence Before This Study

For public health policy makers to address the COVID-19 pandemic, stratified estimates are necessary to investigate the potential outcomes of policy scenarios targeting specific subpopulations. However, estimated epidemic quantities such as rates of severe illness and death, the case fatality rate (CFR) and the infection fatality rate (IFR) have been expressed in terms of aggregated population-level estimates or estimates by age groups alone, due to the lack of epidemiological data at the refined subpopulation level. We searched PubMed for articles published in English from inception to November 25, 2020 with the keywords “covid-19” AND “ifr” OR “infection fatality risk” OR “infection fatality rate” AND “age” AND “obesity” OR “comorbidity” OR “smoking”, which identified 31 results. We found a few estimates of odds ratios and hazard ratios for each potential risk factor controlling for other risk factors, but found no estimates of case fatality rate or infection fatality rate by age and another moderator.

Added Value of This Study

Conventionally, estimates of risk effects and outcomes given combinations of conditions for refined subpopulations are obtained through access to individual-level data and the application of multiple regression techniques. At the time of this study, individual-level COVID-19 data were not widely available nor sampled in an appropriate manner to avoid substantial bias. In the absence of access to appropriate individual-level data, we develop a framework for using available data on the prevalence of COVID-19 risk factors (age, comorbidities, BMI, smoking status) in subpopulations, and epidemic dynamics at the population-level and stratified by age, to estimate subpopulation-stratified probabilities of severe illness and the CFR and IFR (accounting for estimated total infections) across multiple combinations of risk factors. A dynamic epidemic model is integrated with a relative risk modeling approach to produce time-varying subpopulation-stratified estimates. The integrated model allows analyzing dynamic outcomes and parameters by population and subpopulation, and simulating alternate policy scenarios that protect specific at-risk subpopulations and modify the population-wide transmission rate. The model is calibrated to data from Los Angeles County from the period March 1 - October 15, 2020.

Implications of All the Available Evidence

Results highlight the value of strata-specific estimates in anticipating disparities in the impact of the epidemic and the efficacy of targeted subpopulation-level policy interventions. The risk of severe illness given infection and the CFR and IFR vary tremendously across subgroups within a population and have changed over time, suggesting that it may be inappropriate to reduce the variation in risk to static, single summary measures for the entire population. This includes variation not only across age groups, but also within age categories

combined with other risk factors analyzed in this study (comorbidities, obesity status, smoking). The fact that the risk of severe illness and the CFR and IFR vary tremendously across subpopulations raises the question of whether protecting certain subpopulations would have achieved similar results in infections, hospitalizations, and deaths as policies applied population-wide to reduce transmission, including the initial lockdown period from March 19 to May 8. We investigated this question through counter-factual simulation, comparing observed trends with alternate policy scenarios modifying the population-wide transmission rate or protecting specific at-risk subpopulations. We find the strict initial lockdown period in LAC was effective because it both reduced overall transmission and protected individuals at greater risk, preventing both healthcare overload and deaths. While similar numbers of deaths as observed in LAC could have been achieved while avoiding initial societal lockdown through the implementation of a more moderate policy with regard to transmission reduction through non-pharmaceutical interventions (NPIs) combined with greater protection of individuals 65+, this would have come at the potential expense of overwhelming the healthcare system. These results illustrate the value in policy analysis of a reproducible model-based method for producing subpopulation-level disease severity and IFR estimates without access to individual data, which can be applied in other locations and policy settings such as the evaluation of vaccination policy focused on targeting at-risk sub-groups.

1. Introduction

Health disparities have emerged with the COVID-19 epidemic because the risk of exposure to infection and the prevalence of risk factors for severe outcomes given infection vary within and between populations and over time [1, 2, 3, 4, 5, 6, 7]. For public health policy makers to better address the pandemic, models reporting stratified estimates are necessary to investigate the potential outcomes of policy scenarios targeting specific subpopulations. However, estimated epidemic quantities such as rates of severe illness and death, the case fatality rate (*CFR*), and the infection fatality rate (*IFR*) are often expressed in terms of aggregated population-level estimates or by age group alone due to the lack of epidemiological data at the refined subpopulation level [8, 9, 10]. While data may be available for single risk factor strata such as by age [11], data on subpopulations representing individuals with combinations of risk factors are not reported or available. Conventionally, estimates of risk effects and outcomes given combinations of conditions are obtained through access to individual-level data and the application of multiple regression techniques [12, 5]. At the time of this study, individual-level COVID-19 data were not widely available nor sampled in an appropriate manner to avoid substantial bias [13].

In this paper we develop a model that produces stratified estimates of illness severity and death for subpopulations representing individuals with combinations of risk factors from available dynamic epidemiological data at the aggre-

gated population level [11]. In the absence of access to individual-level data, we apply a statistical technique developed for conditional analysis of marginal summary statistics [14] to obtain estimates of the conditional effects of combinations of risk factors for COVID-19 on the probability of severe illness and death using data from published studies reporting the marginal effects of individual risk factors [3, 2]. We consider the risk factors age, existing comorbidities, obesity, and smoking. Separately, we develop a stochastic epidemic model that estimates time-varying probabilities of hospitalization, ICU admission, and death given infection at the population level. We integrate the conditional risk effects and the population-level probabilities, together with available dynamic data on the prevalence of infections stratified by age, in a framework that estimates the probability of advancing to each stage of disease, CFRs, and IFRs, stratified across all plausible combinations of the modeled risk factors.

The integrated model allows the analysis of dynamic outcomes and parameters by population and subpopulation. Focusing on Los Angeles County (LAC), the most populous and one of the most diverse counties in the United States, we analyze the estimated overall and risk-stratified time-varying illness severity probabilities, CFRs, and IFRs in relation to the epidemic timecourse and implemented policy decisions. We focus on the time period March 1 through October 15, marking the end of what we characterize as a second wave of the epidemic and before a third wave of rising infection began. Through simulation, we compare observed trends with alternate policy scenarios protecting specific at-risk subpopulations or modifying the level of non-pharmaceutical intervention (NPIs) implementation. Results highlight the value of strata-specific estimates in understanding disparities in the impact of the epidemic, and the efficacy of targeted subpopulation-level policy interventions in LAC.

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 hospitalization, ICU admission, and death. Using Bayesian methods for parameter estimation and uncertainty quantification, we estimated the population-average time-varying probabilities of transitions between the infected, hospitalized, ICU, mechanical ventilation, death, and recovery compartments, and the resulting population-average time-varying case fatality rate (*CFR*, defined as deaths over observed infections) and infection fatality rate (*IFR*, defined as deaths over all infections) (Section 2.1). In parallel, we used available data from published studies on the marginal effects of individual risk factors (age, existing comorbidities, obesity, smoking) to calculate conditional risk effects estimates for three models: (1) hospitalization given infection, (2) ICU admission given hospitalization, and (3) death given hospitalization. The conditional risk estimates were integrated with the corresponding probability estimates from the dynamic epidemic model to create a *risk model* (Section 2.2). The *risk model* enables us to estimate, stratified across 39 combinations of modeled risk factors (i.e. *risk profiles*), the probability of

each stage of disease given infection within LAC. Furthermore, we integrate the time-varying stratified probability of each stage of disease with the time-series of observed infections, estimated total infections including observed and unobserved, deaths, together with available data on the prevalence of infections stratified by age, to estimate the risk profile-stratified CFR and IFR across time.

2.1. Epidemic model

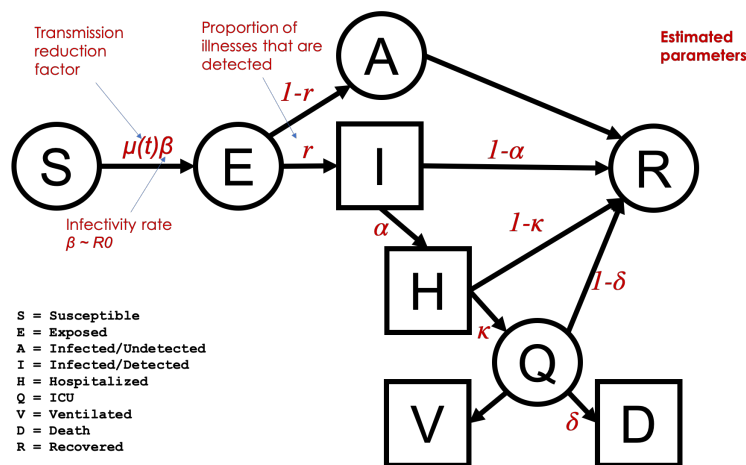


Figure 1: Epidemic model structure and key estimated parameters. Model compartments with available data are represented as square compartments

We develop a model of COVID-19 transmission in a single, homogeneously-mixed population divided into nine compartments representing different disease states (Figure 1). 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 mechanical ventilation support (V), and that die (D). Each individual can only be in one state at each point in time with the exception of the mechanical ventilation support class, which is viewed as a proxy with error for 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.

To model this dynamical state system we employ a discrete-time approximation to the corresponding stochastic continuous-time Markov process in which transitions of individuals between disease stages are seen as stochastic movements between the corresponding population compartments with random transition rates [15, 16]. This model keeps track of the number of individuals in

each compartment and the flows of individuals transitioning between compartments through a set of coupled discrete-time multinomial counting processes with transmission rates defined by Poisson processes. To simulate from this system we employ a Euler numerical scheme for Markov process models [15].
95 For more details see appendix pp 1-6.

The basic reproductive number, R_0 , defined as the mean number of secondary cases generated by a typical infectious individual on each day in a full susceptible population [17], is a function of model parameters including the transmission rate $\beta(t)$ [18] (appendix pp 7-8), defined as the average number of
100 individuals that an infected individual will infect per day. We introduce an additional time-varying parameter, $\mu(t)$, representing modification to the original β_0 and equivalently R_0 , 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
105 different policy 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 (appendix p 6).

2.1.1. Parameter Estimation

110 All transition rate parameters (e.g., the time between exposure and infectiousness) are modeled as fixed values taken directly from published literature (appendix p 3). The model has six unknown parameters, $\theta = \{\beta(t), r(t), \alpha_t, \kappa_t, \delta_t, p_V\}$, which we estimate using Approximate Bayesian Computation (ABC) techniques using multiple data sources to specify informative prior parameter distributions.
115 The prior distribution for R_0 is informed by values estimated from other published studies on COVID-19 [19, 20]. 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 provided by Unacast [21] on reductions in distances travelled and
120 encounter rates [22]. The prior distribution for the fraction of observed cases out of all infections, $r(t)$, was informed by results of a CDC study reporting seroprevalence surveys across 10 communities in March - early May for dates within that time period, and was allowed to vary more widely for dates beyond that time period [23] (appendix pp 8-11).

125 Changes in the probabilities of each disease stage over time are determined by the changing risk profile of the infected population. The only available infection data for LAC by risk factor was for age [11]. We therefore used the distribution of infections by age to inform the timescale and shape of the linear function modeling α_t , κ_t , and δ_t over time. Prior distributions for each probability on
130 key dates were modeled as normal distributions with means informed by the ratios of the observed numbers of infections, hospitalizations, and deaths in LAC (appendix p 11).

The model was fit to the daily and cumulative count of infections (observed), hospitalizations, individuals undergoing ventilation support, and deaths in LAC
135 from a data set from the LACDPH, updated daily by the COVID-19 Outbreak

Data Coordination Team [11](appendix p 12). Using ABC on multiple parameters simultaneously produces joint posterior estimates over all parameters. The $CFR(t)$ and $IFR(t)$ were calculated as estimated deaths over estimated cumulative observed infections or estimated cumulative total infections, respectively. To provide uncertainty estimates, we simulate trajectories with parameter values coming from the joint posterior distribution and aggregate simulations from 1000 jointly estimated parameter sets. For each parameter set, we aggregate simulations for 100 stochastic epidemic model realizations. We pool together all simulations and report their median and 95% credible intervals (CI) as the 2.5th/97.5th quantiles of realizations. 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 parameter values.

2.2. Risk model

Using studies reporting the risk of severe COVID-19 outcomes given individual risk factors [3, 2], we construct a logistic regression model to estimate the probability of infection 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) infection $P_t(H|I)$, (2) the probability the individual is admitted to the ICU given admittance to hospital $P_t(Q|H)$, and (3) the probability that the individual dies given being admitted to the ICU $P_t(D|Q)$. These probabilities correspond to the epidemic model estimated parameters α_t , κ_t , and δ_t , 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 were modeled as binary variables.

We estimate the conditional risk effects corresponding to each factor for each risk model using marginal effects estimates available from reported studies and a method called the joint analysis of marginal summary statistics (JAM) [14]. JAM uses two pieces of information: (i) the marginal effect estimates between risk factors and the outcome and (ii) a reference correlation structure between the risk factors. For information informing (i) we obtain the marginal log RR between individual risk factors and COVID-19 illness severity from peer-reviewed published COVID-19 studies [3, 2] (left column of Table 2). For (ii), we use data from The National Health and Nutrition Examination Survey (NHANES) from 2017-2018 [24] (details in appendix pp 13-14).

180 To construct a model to estimate $P_t(H|I)$, $P_t(Q|H)$, and $P_t(D|Q)$, we combine in a logistic model the 39 risk profiles representing all plausible linear combinations of the risk factors specified in a mean centered design matrix, \mathbf{X} ; and their corresponding conditional log RR obtained from JAM, $\hat{\psi}$; with specified intercepts set to the estimated probabilities from the epidemic model
185 (Section 2.1) for $\hat{\alpha}_t$, $\hat{\kappa}_t$, $\hat{\delta}_t$, respectively (appendix pp 16-17). For example, to estimate the vector of probabilities of hospitalization given infection for all risk profiles we use $\widehat{P}_t(H|I) = \text{expit}(\hat{\alpha}_t + \mathbf{X}\hat{\psi})$. The reference profile are individuals age 0 – 19 with no comorbidity, $BMI < 30 \frac{kg}{m^2}$, and non-smoking. The specification of the frequency of each risk profile at each stage of disease for the mean
190 centered design matrix is obtained from available LAC data on the prevalence of each age group in infections [11] and an estimate of the frequency of each risk profile within the overall population based on prevalence data for individual risk factors [25, 26], updated at each stage of disease (appendix pp 14-16). We report estimates at three time points: May 15, August 1, and October 15, 2020.

195 2.3. Calculating Risk Profile Stratified CFR(t) and IFR(t)

To calculate the time-varying CFR(t) and IFR(t) for each risk profile, the estimated frequency of each profile in the infected population and in the deceased population (obtained from the risk model) are multiplied by each value of the estimated cumulative number of observed infections (I) or total infections (A),
200 and deaths (D), respectively. We find the CFR(t) and IFR(t) for each model realization as the number of deaths over observed infections, and number of deaths over total infections, respectively. Repeating across the 1000 model realizations achieves the 95% CI (appendix pp 18-19).

2.4. Scenario Analysis

205 We implement scenarios modeling the protection of at-risk populations and changes to the population-wide transmission rate at different times. We increase or decrease the value of the reproductive number $R(t)$ to reflect different levels of non-pharmaceutical interventions (NPIs), which could include measures such as physical distancing and/or mask adherence. Specifically, we model three levels of population-wide transmission rates: (1) *NPIs=Observed* implements the observed (epidemic model-estimated) $R(t)$ throughout the epidemic in LAC. (2) *NPIs=Moderate* implements an $R(t)$ equal to the estimated maximum value reached in LAC after the initial lockdown restrictions were eased (Section 3.1). Although this maximum value for $R(t)$ was reached on May 15, in this scenario we decrease $R(t)$ from the initial R_0 to this maximum $R(t)$ value between
210 March 12 and March 27 to reflect no community lockdown implemented. (3) *NPIs=None* implements $R(t) = R_0$ throughout the time interval March - October 2020, representing a baseline scenario in which no actions or behaviors reduce the native R_0 . The trend of $R(t)$ modeled for the three levels is shown
215 in Figure 4a.

220 To simulate scenarios protecting at-risk populations we focus on the shielding of individuals aged 65+ and calculate values of the probabilities of each

stage of disease given infection ($\alpha_t^{Protect}$, $\kappa_t^{Protect}$, and $\delta_t^{Protect}$) over time with the subpopulation of individuals 65+ removed. We isolate a fraction of individuals 65+ above the observed prevalence of this subpopulation in the infected population. We model three levels: (1) *Protect=Observed* models the trend in the probabilities of severe illness estimated by the epidemic model. We call this the “*Protect=Observed*” rather than “*Protect=None*” level because it follows the natural behaviorally-adapted reduction of individuals 65+ in the infected population (see Section ??). (2) *Protect=High* isolates all individuals 65+, approximately 11% of the LAC population, from infection and subsequent stages of disease. We implement this scenario by calculating the population-average $\alpha_t^{Protect.100}$, $\kappa_t^{Protect.100}$, and $\delta_t^{Protect.100}$ with individuals 65+ removed along the same timeline as the observed trend in α_t , κ_t , and δ_t . The frequency of each risk profile in infections is made to reflect the age distribution over infections as observed in LAC data (following the approach described in Section 2.2), renormalized after profiles including individuals 65+ are removed. With the renormalized frequency distribution of each profile over infections, we use the risk model (Section 2.2) to find the population-average adjusted values. The timeline follows that for the observed trend in α_t , κ_t , and δ_t . (3) *Protect=Moderate* isolates 50% of individuals 65+ from community transmission and subsequent stages of disease from the observed distribution of each age group over infections on each date and calculates the population-average adjusted $\alpha_t^{Protect.50}$, $\kappa_t^{Protect.50}$, and $\delta_t^{Protect.50}$, following the approach for the *Protect=High* scenario. The trend of the probabilities of each stage of disease given infection modeled for the three levels is shown in Figure 4b (full details in appendix pp 19-21).

We implement nine scenarios representing all combinations of the three *NPI* and three *Protect* settings. For each scenario, we simulate the model with the estimated parameter values for $R0$, $r(t)$, p_V , and starting time t_0 with the remaining parameters implemented as described above.

2.5. Role of the funding source

Funders had no role in study design, data collection, data analysis, data interpretation, writing of the report, or the decision to submit for publication. The corresponding author had full access to all of the data and the final responsibility to submit for publication.

3. Results

3.1. Model estimates and epidemic trends in LAC

Model fits. Figure 2 summarizes the epidemic model fit with COVID-19 data for LAC from March 1 through October 15 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

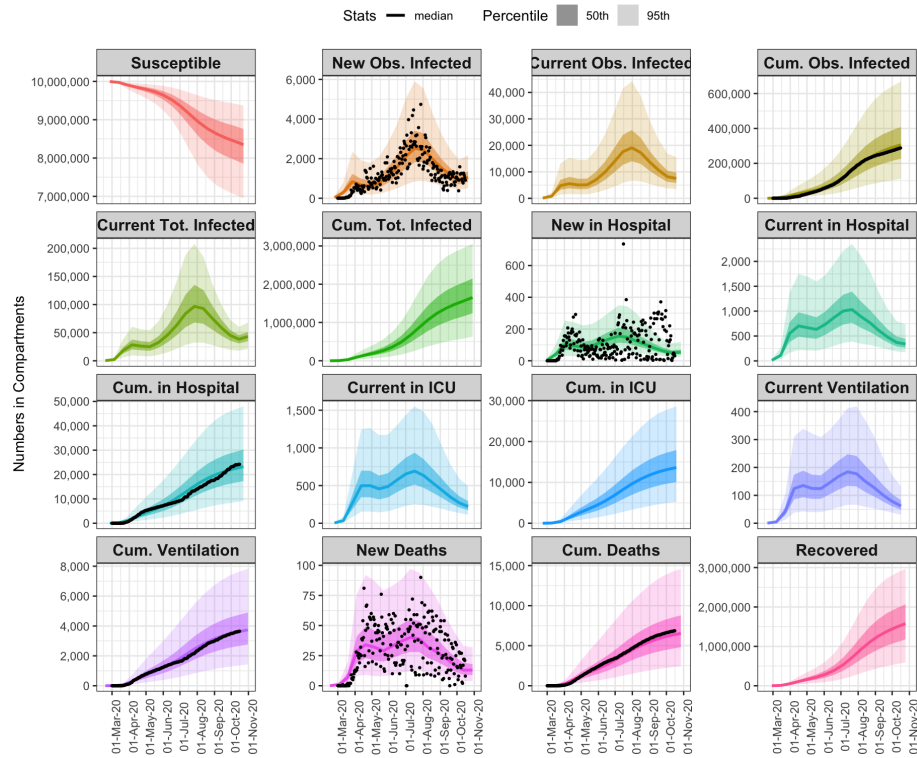


Figure 2: Summary of the epidemic model fit with COVID-19 data for Los Angeles, for all state variables, across multiple views: New cases, representing new daily incidence, current number in a compartment at a specific date, and cumulative counts. Available observed data (for new and cumulative counts) are plotted as black dots. Estimates are shown as the median number in compartments over time, with the 50% (darker) and 95% (lighter) CI.

counts until a specific date. Observed data for available compartments are plotted as black dots. The figure demonstrates that good model fits are achieved in all compartments across time.

Epidemic timecourse in LAC. The LA City Mayor’s Office distinguishes between three stages of the COVID-19 epidemic in LA City and County relating to policy response measures implemented following the orders of the County Health Officer: Stage I, March 19 - May 7: the initial shutdown; Stage II, May 8 - June 11: the first steps towards reopening; Stage III, June 12 - October 15 (and beyond): greater reopening followed by “modifications” closing higher risk settings (including bars and indoor seating in restaurants) [27, 28]. The start of the school year on August 18, although virtual, marked a change in activity level and is also depicted. We characterize three waves of the epidemic occurring across these stages: a first wave, March 1 - May 6, occurring between Stage I and the beginning of Stage II and peaking on April 1; a second and much

It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/).

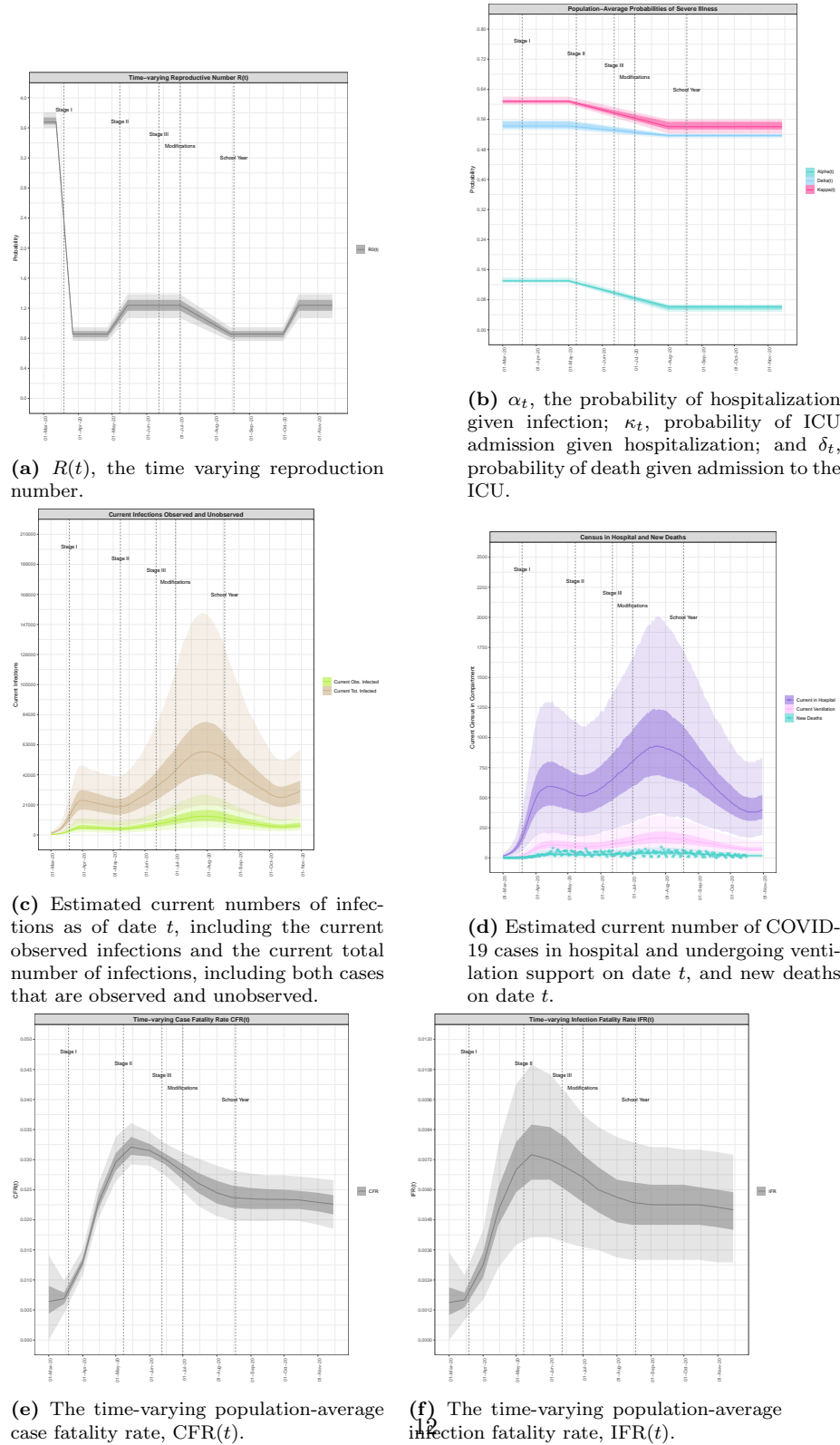


Figure 3: Timeseries of model-estimated parameters and compartmental variables relative to COVID-19 policy decisions in LAC. Model-estimated median curves are plotted along with the 50th% (dark shading) and 95% CI (light shading). Data is not available for current numbers but is plotted for the variable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

larger wave, May 7 - October 14, beginning with Stage II and peaking during Stage III on July 30; and a third wave that began on October 15 and which is
280 outside the scope of this paper. Figures 3a-3f characterize the epidemic course and estimated model parameters relative to these policy stages and epidemic waves. A full time course of the epidemic and policy decisions in LAC can be found at [29].

We estimate that for most of Stage I, the overall observation rate was $r(t) =$
285 0.23 (95% CI: $(0.13,0.33)$) of all infections observed. Beginning in mid-April, the observation rate began to steadily increase until levelling off at a value of $r(t) = 0.41$ $(0.11,0.69)$ by August 15. 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 $R_0=3.675$ $(3.627,3.729)$. Dur-
290 ing the period from March 12 to March 27, beginning just before Stage I was implemented, we estimate a reduction to an $R(t)$ of 0.859 $(0.758,0.947)$. The corresponding reduction in transmission led to a levelling off at $30\ 127$ $(39\ 039, 58\ 842)$ estimated current total infections on April 1 and subsequent decrease until May 6, followed by hospitalizations, ICU admissions, and deaths. No-
295 tably, $R(t)$ began to rise again following increasing mobility behavior on April 27, almost two weeks before the Stage II reopening policy was implemented, portending the increase in infections to follow.

Throughout Stage II and the beginning of Stage III, we estimate an increasing trend in $R(t)$, reaching 1.264 $(1.052,1.378)$ around the time that the
300 “modifications” (re-closures) were implemented at the beginning of July. The increasing $R(t)$ during this period was followed by a second exponential growth phase in infections, hospitalizations, ICU admissions, and deaths. $R(t)$ began to decrease following the implementation of the “modifications,” and by the start of the school year in mid-August (virtually) had returned to the Stage I value
305 where it plateaued until a new rise at the beginning of October. Estimated current total infections continued to increase until a peak of $93\ 720$ $(34\ 404, 184\ 084)$ on July 30, reflecting the expected delay between reductions in $R(t)$ and the infection rate. Current total infections followed a decreasing trend until October 15, marking the beginning the third wave of the epidemic. Hospitaliza-
310 tions and ICU admissions followed the infection rate and reached much larger second peaks in late July before decreasing again until reaching new minimums in mid-October. Importantly, even at upper 95%CI of the peak, the current census of cases in hospital, ICU, and undergoing ventilation support remain well below the capacity limits of approximately 4 000 hospital beds and 2 245
315 ICU beds in LAC [11].

Although current and new hospitalizations and ICU admissions reached levels during the second wave of more than double that of the first wave, new deaths in the second wave only marginally surpassed values in the first, and the rate of all three decreased relative to increasing infections. We identify three
320 phases of the probabilities of severe illness reflecting this behavior: the highest values observed during an initial phase from the beginning of the epidemic until May 1, a decreasing trend between May 1 and August 1, and stabilized reduced values following August 1. Specifically, we estimate that between May 1 and

It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/).

Profile	Group	age	BMI	smoking	comorbidity	Pop.Prev	P(H I),May.15	P(H I),Aug.1	P(H I),Oct.15	P(Q H),May.15	P(Q H),Aug.1	P(Q H),Oct.15	P(D Q),May.15	P(D Q),Aug.1	P(D Q),Oct.15
27	High	65+	30<-BMI<40	Non Smoker	Comorbidity	0.0110	0.2626	0.2036	0.2080	0.6732	0.7066	0.7082	0.7224	0.7026	0.7051
24	High	65+	BMI<30	Non Smoker	Comorbidity	0.0699	0.1635	0.1230	0.1260	0.6624	0.6963	0.6980	0.6998	0.6791	0.6817
32	High	45-64	BMI<30	Smoker	Comorbidity	0.0167	0.1690	0.1274	0.1304	0.6735	0.7068	0.7084	0.6517	0.6294	0.6322
6	High	65+	BMI<30	Non Smoker	No Comorbidity	0.0254	0.1148	0.0852	0.0873	0.6579	0.6921	0.6937	0.6914	0.6703	0.6729
26	High	45-64	20<-BMI<40	Non Smoker	Comorbidity	0.0362	0.1733	0.1308	0.1339	0.5727	0.6104	0.6122	0.5166	0.4924	0.4954
14	High	45-64	BMI<30	Smoker	No Comorbidity	0.0130	0.1189	0.0884	0.0905	0.6691	0.7026	0.7043	0.6426	0.6200	0.6228
8	High	45-64	30<-BMI<40	Non Smoker	No Comorbidity	0.0219	0.1221	0.0908	0.0930	0.5678	0.6056	0.6075	0.5066	0.4824	0.4854
23	High	45-64	BMI<30	Non Smoker	Comorbidity	0.1510	0.1031	0.0763	0.0782	0.5607	0.5987	0.6005	0.4891	0.4650	0.4679
31	High	20-44	BMI<30	Smoker	Comorbidity	0.0206	0.1069	0.0792	0.0811	0.5730	0.6106	0.6125	0.4370	0.4133	0.4162
5	High	45-64	BMI<30	Non Smoker	No Comorbidity	0.1045	0.0709	0.0520	0.0533	0.5557	0.5938	0.5957	0.4791	0.4550	0.4580
13	High	20-44	BMI<30	Smoker	No Comorbidity	0.0300	0.0736	0.0540	0.0553	0.5681	0.6058	0.6077	0.4272	0.4036	0.4065
25	High	20-44	30<-BMI<40	Non Smoker	Comorbidity	0.0233	0.1098	0.0814	0.0834	0.4658	0.5047	0.5067	0.3072	0.2869	0.2894
7	High	20-44	30<-BMI<40	Non Smoker	No Comorbidity	0.0240	0.0757	0.0555	0.0569	0.4608	0.4997	0.5017	0.2887	0.2788	0.2812
22	High	20-44	BMI<30	Non Smoker	Comorbidity	0.1055	0.0634	0.0464	0.0476	0.4536	0.4925	0.4944	0.2843	0.2650	0.2673
4	High	20-44	BMI<30	Non Smoker	No Comorbidity	0.1401	0.0430	0.0313	0.0321	0.4487	0.4875	0.4894	0.2762	0.2572	0.2595
1	High	0-19	BMI<30	Non Smoker	No Comorbidity	0.1463	0.0163	0.0117	0.0120	0.2304	0.2592	0.2607	0.1231	0.1130	0.1142

Table 1: Profile-stratified $P_t(H|I)$, $P_t(Q|H)$, and $P_t(D|Q)$: Risk profiles (characterized by unique combination of age group, BMI range, smoking status, and any comorbidity), risk group (1-5), model-estimated population prevalence, the frequency of each profile in the infected population, and the median probability of hospitalization given infection, ICU admission given hospitalization, and death given admission to the ICU on May 15, August 1, and October 15 2020. Only profiles with a population prevalence $> 1\%$ are shown.

August 1, the population-average probability of hospitalization given infection α_t decreased from 0.138 (0.125,0.15) to 0.048 (0.04,0.055); the probability of ICU admission given hospitalization κ_t decreased from 0.597 (0.593,0.603) to 0.543 (0.537,0.549); and the probability of death given ICU admission δ_t decreased from 0.561 (0.557,0.567) to 0.503 (0.49,0.52). The population-average CFR(t) and IFR(t) followed the same trend; we estimate the CFR(t) was 0.0243 (0.021,0.028) on April 15 (corresponding to the first phase values for α_t , κ_t , and δ_t), reached a peak on May 15 of 0.0342 (0.030,0.039), and had dropped to an almost-stable value of 0.0237 (0.02,0.0292) by August 1 (representing the second phase of values for α_t , κ_t , and δ_t), although a slight downward trend still continued until a value of 0.0211 (0.0176,0.0251) on October 15. The IFR(t) was 0.0055 (0.0028,0.0082) on April 15, peaked at 0.0077 (0.0038,0.0115) on May 15, decreased to 0.0055 (0.0024,0.0090) by August 1 and 0.0051 (0.0023,0.0078) by October 15.

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

Table 2 displays the marginal relative risks (RR) extracted from the literature (left column) and conditional RR estimated by the risk model for each risk factor considered by our model on the rates of hospitalization given infection, ($H|I$), ICU admission given hospitalization, ($Q|H$), and death given ICU admission, ($D|Q$). We find that 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.

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 [†]	1.12 (0.73, 1.71)
Smoker	1 [†]	1.96 (1.33, 2.89)
Any comorbidity	1.64 (0.81, 3.32)	1.05 (0.78, 1.43)

Table 2: The marginal relative risk collected from published studies on COVID-19 and conditional relative risk estimated by the risk model for each risk factor on rates of hospitalization given infection, $(H|I)$, ICU admission given hospitalization, $(Q|H)$ and death given ICU admission, $(D|Q)$ (95% credible interval).

[†]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.

Profile	Group	age	comorbidity	BMI	smoking	Pop.Prev	Inf.May.15	Inf.Aug.1	Inf.Oct.15	CFR.May.15	CFR.Aug.1	CFR.Oct.15	IFR.May.15	IFR.Aug.1	IFR.Oct.15
27	Smoker	65+	Comorbidity	30-BMI<40	Non Smoker	0.0110	0.0205	0.0114	0.0106	0.1175	0.1164	0.1123	0.0269	0.0269	0.0260
24	Smoker	65+	Comorbidity	BMI<30	Non Smoker	0.0889	0.1300	0.0722	0.0656	0.0897	0.0670	0.0648	0.0160	0.0155	0.0150
32	Smoker	45-64	Comorbidity	BMI<30	Smoker	0.0167	0.0170	0.0180	0.0178	0.0882	0.0653	0.0632	0.0156	0.0151	0.0146
6	Smoker	65+	No Comorbidity	BMI<30	Non Smoker	0.0258	0.0472	0.0262	0.0249	0.0480	0.0455	0.0441	0.0110	0.0105	0.0102
26	Smoker	45-64	Comorbidity	30-BMI<40	Non Smoker	0.0362	0.0388	0.0411	0.0408	0.0472	0.0453	0.0439	0.0108	0.0105	0.0102
14	Smoker	45-64	No Comorbidity	BMI<30	Smoker	0.0130	0.0133	0.0140	0.0139	0.0470	0.0443	0.0429	0.0108	0.0102	0.0099
8	Smoker	45-64	No Comorbidity	30-BMI<40	Non Smoker	0.0218	0.0223	0.0236	0.0234	0.0323	0.0306	0.0297	0.0074	0.0071	0.0069
23	Smoker	45-64	Comorbidity	BMI<30	Non Smoker	0.1510	0.1537	0.1623	0.1609	0.0260	0.0245	0.0238	0.0050	0.0057	0.0055
31	Smoker	20-44	Comorbidity	BMI<30	Smoker	0.0206	0.0206	0.0221	0.0216	0.0246	0.0230	0.0224	0.0036	0.0033	0.0032
5	Smoker	45-64	No Comorbidity	BMI<30	Non Smoker	0.1045	0.1064	0.1123	0.1114	0.0174	0.0162	0.0157	0.0040	0.0037	0.0036
13	Smoker	20-44	No Comorbidity	BMI<30	Smoker	0.0307	0.0313	0.0330	0.0327	0.0145	0.0132	0.0148	0.0035	0.0035	0.0034
25	Smoker	20-44	Comorbidity	30-BMI<40	Non Smoker	0.0238	0.0242	0.0255	0.0253	0.0145	0.0138	0.0132	0.0033	0.0031	0.0031
7	Smoker	20-44	No Comorbidity	30-BMI<40	Non Smoker	0.0244	0.0244	0.0258	0.0255	0.0096	0.0088	0.0087	0.0022	0.0021	0.0020
22	Smoker	20-44	Comorbidity	BMI<30	Non Smoker	0.1055	0.1074	0.1134	0.1124	0.0074	0.0070	0.0068	0.0014	0.0016	0.0016
4	Smoker	20-44	No Comorbidity	BMI<30	Non Smoker	0.1401	0.1427	0.1507	0.1493	0.0044	0.0045	0.0044	0.0011	0.0010	0.0010
1	Smoker	0-19	No Comorbidity	BMI<30	Non Smoker	0.1463	0.0348	0.0880	0.1008	0.0004	0.0004	0.0004	0.0001	0.0001	0.0001

Table 3: Profile-stratified CFR(t) and IFR(t): Risk profiles (characterized by unique combination of age group, BMI range, smoking status, and any comorbidity), risk group (1-5), model-estimated population prevalence, the frequency of each profile in the infected population, and the median of the CFR(t) and IFR(t) on May 15, August 1, and October 15 2020. Only profiles with a population prevalence $> 1\%$ are shown.

Table 1 shows the model-estimated population prevalence, the frequency in the infected population, and the median $P_t(H|I)$, $P_t(Q|H)$, and $P_t(D|Q)$ for each profile with an estimated population prevalence greater than 1% (see appendix p 18 for all profiles). Probabilities are provided for the dates May 15, representing the peak value for the overall CFR(t) and IFR(t) (Figures 3e, 3f), August 1, representing the value of the probabilities after the overall $P_t(H|I)$, $P_t(Q|H)$, and $P_t(D|Q)$ had completed their descent but while the CFR(t) and IFR(t) were still decreasing, and October 15, the most recent date.

The probability of hospitalization given infection, ICU admission given hospitalization, and death given ICU admission vary extensively across the risk profiles. Notably, the risks within specific marginal factor groups also vary extensively. For example, within the age group 65+, the probability of hospitalization given infection is approximately five times greater for individuals with at least one comorbidity, a smoking history, and severe obesity then for individuals that have no comorbidities, do not smoke, and have a healthy BMI.

3.3. Risk-stratified CFR and IFR for LAC

Table 3 shows the frequency in the infected population and the median of the CFR(t) and IFR(t) on May 15, August 1, and October 15 2020 for each risk profile with a model-estimated population prevalence $> 1\%$ (see appendix p 20 for all 39 profiles). The probability of death given infection varies tremendously across populations. While the overall median CFR(t) ranged from a high of 0.0342 (0.030,0.039) on May 15 to a low of 0.0211 (0.0176,0.0251) on October 15, the profile-stratified median CFR(t) ranged from 0.0004 to 0.23 across the profiles. The overall median IFR(t) ranged from a high of 0.0077 (0.0038,0.0115) on May 15 to a low of 0.0051 (0.0023,0.0078) on October 15, but ranged from 0.0001 to 0.052 across the profiles. To facilitate interpretation of the probabilities and variability across risk profiles, we group the profiles into five Risk Groups based on similar within-group CFR(t) on May 15, with Risk

375 1 being composed of individuals with $CFR(t) > 0.16$; $0.08 < CFR(t) < 0.16$ in Risk 2; $0.04 < CFR(t) < 0.08$ in Risk 3; $0.01 < CFR(t) < 0.04$ in Risk 4; and $CFR(t) < 0.01$ in Risk 5 (risk groups are indicated for each profile in Table 3). Notably, profiles including individuals 65+ are found in Groups 1, 2, and 3, and profiles for individuals aged 45-64 are found in Groups 1-4 (appendix p 20).

380 3.4. Scenario analysis

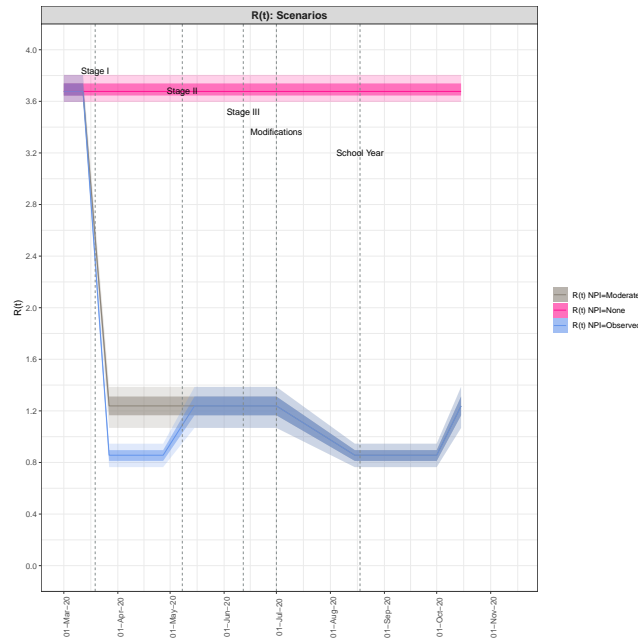
Figure 4a shows the three levels of NPI policies implemented ($NPIs=None, Moderate, Observed$) and Figure 4b shows the three levels of isolating the 65+ population implemented ($Protect=Observed, 50%, 100%$) to create the nine scenarios. Figure 5 shows the results of scenarios that include the three *Protect* levels, $NPI=Observed$ and $NPI=Moderate$; the results of all scenarios including $NPI=Moderate$ are shown in appendix p 22. Results for each scenario by $CFR(t)$, $IFR(t)$, cumulative deaths, current in hospital, current observed infections, and current total infections are pictured. The “Do Nothing” scenario (Scenario 1) results in a median of approximately 75 000 deaths by October 15, exceeding the observed death count in LAC by approximately 11 times the observed median value of 6 855, as well as excessively surpassing hospital capacity for two months with a peak of 40 000 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 3) averts approximately 50% of deaths estimated in the “Do Nothing” Scenario, however surpasses observed deaths in LAC by a factor of 5.5 and exceeds hospital capacity for one month at peak levels 6.5 times as large as maximum capacity. Implementing a moderate NPI policy with no additional measures to protect at-risk populations (Scenario 4) greatly reduces deaths from the scenarios in which no community transmission reduction takes place although still leads to more than 16 000 deaths by October 15, a factor of 2.5 greater than observed.

When the moderate NPI policy is combined with additional measures to protect 100% of individuals 65+ (Scenario 6), the number of deaths accrued is much closer to the observed death toll in LAC by October 15, while hospital demand would have still approached levels exceeding capacity for two months. If only 50% of individuals 65+ had been protected (Scenario 5), the death count would have doubled the observed count and hospital demand would have come closer to exceeding capacity.

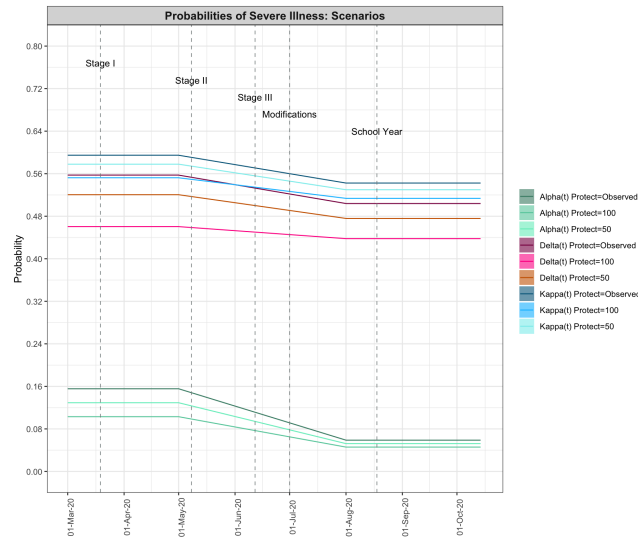
Comparing Scenarios 7-9 in which the observed trend in community transmission is implemented (strict initial lockdown followed by a gradual reopening and stabilization at an $R(t)$ of approximately 1.26), deaths could have been halved if 100% of individuals 65+ had been protected (Scenario 9) or 1 500 deaths if 50% of individuals 65+ had been protected above the observed self-protection that occurred (Scenario 8). Hospitalization demand would not have decreased appreciably.

415 The overall population average $CFR(t)$ and $IFR(t)$ are more strongly determined by the implemented *protection* level; by October 15 these approximate 0.03 and 0.003, respectively, across three scenarios in which 100% of individuals 65+ are protected. Still, much larger absolute numbers of infected individuals,

It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/).



(a) Three levels of non-pharmaceutical-intervention (NPI) policies implemented: In $NPIs=None$ no adaptations are made to R_0 ; 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 α_t, κ_t , and δ_t are implemented; in $Protect=100$ and $Protect=50$, respectively 100% and 50% of individuals 65+ are removed from the observed infection distribution and the population-average $\alpha_t^{Protect}, \kappa_t^{Protect}$, and $\delta_t^{Protect}$ are recalculated.

Figure 4: Implemented policies in scenario analysis.

It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/).

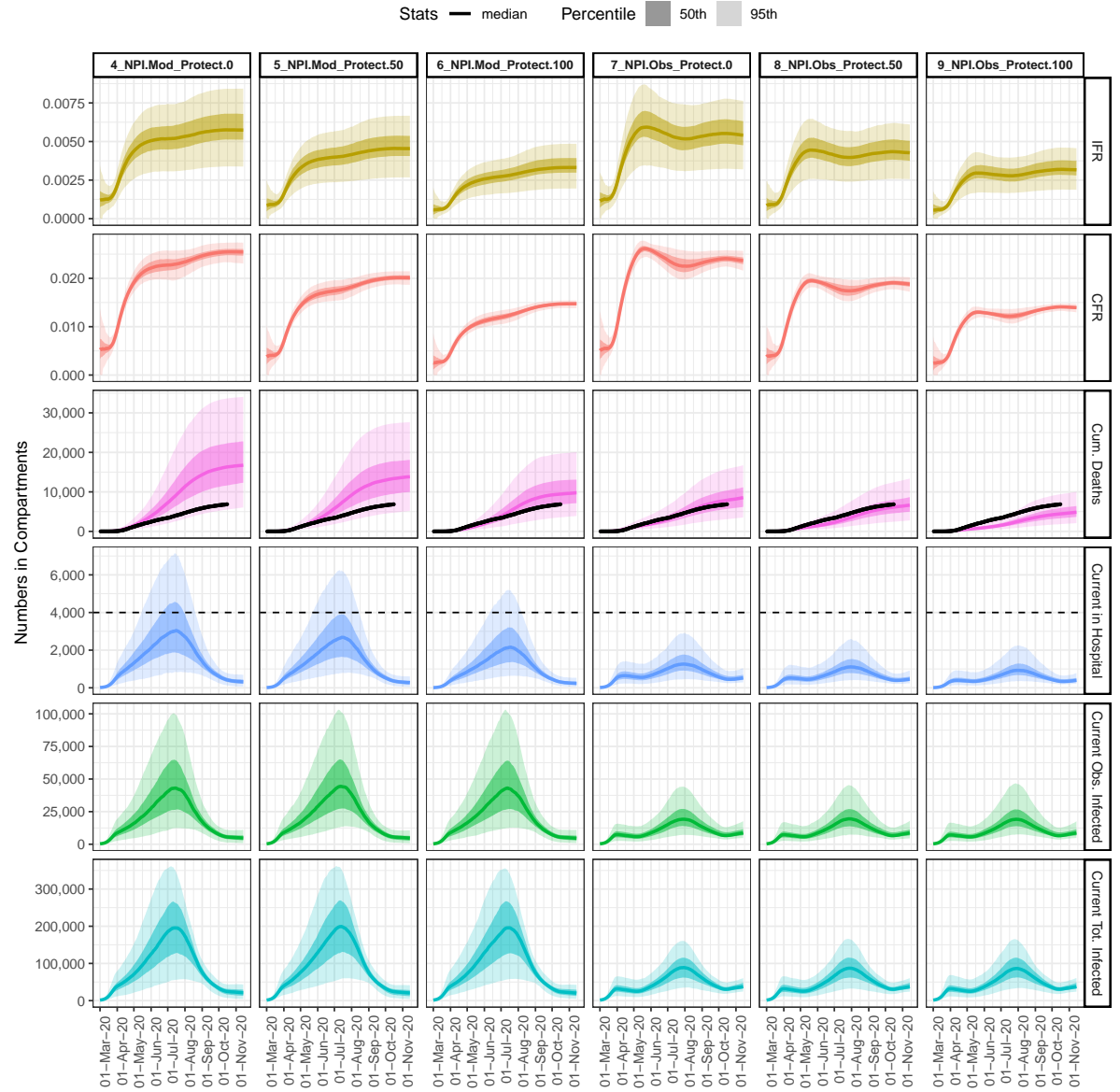


Figure 5: Scenarios implemented and results by CFR(t), IFR(t), cumulative deaths, current in hospital, current observed infections, and current total infections.

420 as in the scenario with $NPI=Nothing$, result in much larger absolute numbers of deaths.

4. Discussion

This work has developed a framework for using available data on COVID-19 epidemic dynamics and prevalences of COVID-19 risk factors at the population level to estimate time-varying subpopulation-stratified probabilities of severe illness and the case fatality rate ($CFR(t)$) and infection fatality rate ($IFR(t)$) across multiple combinations of risk factors. In the absence of individual-level data, the technical contribution of this work was to integrate a dynamic epidemic model with a risk modeling approach to estimate conditional effects from available marginal data and to subsequently produce time-varying subpopulation-stratified estimates for LAC. To reflect the uncertain knowledge of many parameters and the understanding that in non-linear systems small variations to specific parameters can result in large impacts in outputs [30], we account for uncertainty in all results through the use of a stochastic epidemic model and a Bayesian approach to parameter estimation. On its own, the risk model estimates the conditional effects of each risk factor and therefore the overall effect of risk factors in combination. These adjusted effects have not been typically reported in observational studies on COVID-19, yet help to understand which individuals are at highest risk of advancing to each stage of disease. Integrated together with the epidemic model, the modeling framework allows the analysis of dynamic outcomes and parameters relating to each stage of disease by population and subpopulation.

Analyses demonstrate that the risk of severe illness and death has decreased over time and moreover varies tremendously across subpopulations, suggesting that it is inappropriate to summarize epidemiological parameters for the entire population and epidemic time period. This includes variation not only across age groups, but also within age categories combined with other risk factors analyzed in this study (comorbidities, obesity status, smoking). Across the risk profiles we found median IFRs ranging from 0.0095%-0.04% in the youngest age group (0-19), IFRs ranging from 0.1%-1.8% for those aged 20-44, 0.36%-4.3% for those aged 45-64, and 1.02%-5.42% for those aged 65+ (noting that the 95%CI varied beyond these ranges of median values). The highest IFR for each age strata come from profiles also including comorbidities, obesity Class 2 or 3, and except for the youngest age group, current smoking status.

Our age-stratified IFR are comparable to those found in a recent meta-analysis using results from seroprevalence studies in North America and Europe [8], a model-based analysis for estimating IFR for China [9], and a model-based analysis for estimating IFR in New York City during the time period March 1 through June 6th [10], with the exception that we do not account for further substrata above 65 years of age. This can be seen as a limitation of our analysis of IFR for profiles including age strata above age 65+, since the risk of severe outcomes is known to strongly increase with age. A second limitation of our model-based analysis for older risk strata is that we assume

unobserved infections are equally distributed across all risk profiles, whereas
465 there are likely to be far fewer unobserved or asymptomatic infections for those
at higher risk of severe outcomes. For risk profiles including individuals age
65+, our IFRs are therefore underestimated and our CFR estimates are likely
to be a better approximation of the true IFRs; we find median CFRs in the
range 4.4%–23.45% across profiles including individuals 65+, comparable to the
470 range of 2%–22% for IFRs in age groups 65+ found in the meta-analysis [8]. At
the same time, the NYC study found IFRs for individuals age 65–74 of 4.87%
(3.37%–6.89%), comparable to our median IFR estimate for the 65+ age group,
and 14.2% (10.2%–18.1%) for those aged 75 years and older. More generally, in
475 interpreting our results for policy implications, emphasis should be placed on
the relative differences in IFR across risk profiles and the understanding that
the IFR for a specific age strata represents an average across a wide variation
given the presence or absence of other risk factors.

Our overall population-average $IFR(t)$ estimates for Los Angeles County,
which ranged from a peak of 0.77% (0.38%,1.15%) on May 15 down to 0.55%
480 (0.24%,0.90%) by October 15, are similar to the overall IFR estimated by a
model-based analysis for China of 0.66% (95%CI: 0.39%,1.33%), but lower than
the overall IFR estimated for NYC as of June 6 of 1.39% (1.04%–1.77%), or
1.10% if only confirmed COVID-19-related deaths were included [10]. This
could be explained by the larger outbreak experienced by NYC in the spring,
485 which led to healthcare capacity overload in some parts of the city and larger
probabilities of fatality. Our IFR estimates could also be lower since we account
only for underascertainment of infections and not of deaths [31, 32]. Even at
the lowest overall IFR estimated for LAC as of October 15, a key finding is
that COVID-19 is substantially more deadly than seasonal Influenza with a
490 population-average IFR of approximately 0.05% [8, 33].

The time-varying population-average risk probabilities over time provide in-
sights into the changing dynamics of the epidemic in LAC. The decrease in the
probability of hospitalization given infection (α_t) by almost 200% between the
value before May 1 and after August 1 suggests substantive changes in risk
495 of severe illness in the infected population. This follows from the decrease in
the prevalence of the age group 65+ in (observed) infections between April 15
and July 15 from approximately 23% to 12% [11], but may also reflect other
changes in the demographic composition of infected individuals including other
at-risk subpopulations for which LAC data is not available (e.g., individuals
500 with comorbidities), or increases in the prevalence of individuals without ac-
cess to healthcare [34, 35]. The decrease in the probability of ICU admission
given hospitalization (κ_t) and death given ICU admission (δ_t) between May 1
and August 1 both by approximately 10% suggests a potential effect of better
treatment in the healthcare setting [36, 37], as well as seasonal effects [38, 39],
505 but may also result from changes in the infected population making it through
to critical illness.

The fact that the risk of severe illness and the $CFR(t)$ and $IFR(t)$ vary
tremendously across populations raises the question of whether protecting cer-
tain subpopulations would have achieved similar results in infections, hospital-

510 izations, and deaths as policies applied population-wide to reduce transmission, including the initial lockdown period from March 19 to May 8. We investigate this question through simulation, comparing observed trends with alternate policy scenarios modifying the population-wide transmission rate or protecting specific at-risk subpopulations. We implemented scenarios in which no reductions are made to the transmission rate. However, it is likely that individual behavior would have modified to reduce the transmission rate, as observed in locations where formal shelter-at-home policies were not implemented, such as in Sweden [40]. We therefore implemented a more moderate NPI policy that reduced community transmission at the beginning of the epidemic to levels exhibited during Stage III of the epidemic in LAC, which in practice reflected a panel of public health outreach and restriction policies including mask wearing and closing indoor dining and entertainment, but were not as stringent as the more complete community-wide lockdown of Stage I. We focused on isolating individuals aged 65+ because this was a suggested policy response for the state of California [41], and is the suggested policy plan for initial vaccination distribution [42].

The main finding from the scenario analysis is that the strict initial lockdown period in LAC was effective because it both reduced overall transmission and protected individuals at greater risk, resulting in preventing both healthcare overload and deaths (especially among individuals 65+). Simulation results suggest that an initial implementation of a moderate NPI policy on its own would not have been sufficient to avoid overwhelming healthcare capacity, and additionally would have exceeded the observed death count by close to 2.5 times. Combining the moderate NPI policy with a stringent protection of all individuals 65+ would have resulted in a death count at a similar level as observed, however hospital capacity limits would have approached levels exceeding capacity, likely resulting in more severe illnesses and deaths than accounted for by this model. A more moderate protection scenario in which 50% of individuals 65+ are isolated would have resulted in slightly surpassing the observed death toll and similarly threatening to exceed healthcare capacity. Thus, while similar numbers of deaths as observed in LAC could have been achieved while avoiding initial lockdown through the implementation of a more moderate NPI policy combined with greater protection of individuals 65+, this would have come at the expense of overwhelming the healthcare system. We can therefore conclude that the implemented lockdown was well designed, as in hindsight it appears to have been the minimally required course of action to protect the entire LAC population, and particularly those at high risk of severe disease, while also preventing overcapacity within the healthcare system.

The scenario analysis also has implications for future policy. In anticipation of a continued rise in cases in LAC this winter, policy makers need to consider the trade offs of various policy options on the numbers of the overall population that may become infected, severely ill, and deaths when considering policies targeted at subpopulations at greatest risk of transmitting infection and at greatest risk for developing severe outcomes. Furthermore, this scenario analysis mimics potential vaccination policy related to a population-wide rollout (similar

to lockdowns and NPI policies applied to the population) vs. rollout first to specific at-risk populations (similar to scenarios protecting individuals 65+). The results of this scenario analysis suggests policies that first vaccinate only those at greatest risk of severe illness may not be as effective as a more general policy aimed at decreasing the size of the epidemic through vaccinating the overall population; this latter policy, by acting on those at greatest risk of transmitting infection, results in decreasing the total numbers of the population infected as well as subsequent healthcare utilization and deaths [43].

This study is prone to typical limitations occurring when modeling epidemiological dynamics in the context of rapidly evolving infectious disease outbreaks. Data informing the conditional effect estimates within the risk model are aggregated across early, retrospective studies from China (age, comorbidities, smoking) [3] and NYC (BMI) [2] on the fractions of hospitalization, ICU admission, and death by individual risk factors. While we attempt to reframe these results for the demographic composition of the LAC population through regional data on the prevalence of risk factors and the correlation structure between risk factors (appendix p 14), there may be differences in the underlying study population or treatment setting between China, NYC, and LAC that would lead to heterogeneity in effect estimates. However, we believe that the estimates from the Chinese studies do represent population-based estimates as these samples avoid some of the biases present from other potentially available studies, but with highly selected samples.

While this work has focused on demonstrating the substantial heterogeneity in risk probabilities and IFR across subpopulations, we developed a single-population epidemic model. We accounted for differences in the infected population through the observed age distribution in LAC. However, heterogeneity in exposure to COVID-19 infection has been shown to vary extensively across a number of factors including not only age but also race/ethnicity, neighborhood of residence, employment, economic status, and access to PPE, among others [7, 34, 35, 1, 4, 5]. At the time of beginning this study we did not have 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. The approach we developed is a way to use commonly available population-level epidemic timeseries data to model multiple groups in a single population, and combine these population-level estimates with prevalence rates of risk factors to produce stratified estimates for different subpopulations, specific to a given region. Future work will need to develop multi-population models that estimate subpopulation-stratified probabilities and infection rates accounting for key risk factors of both exposure to infection and severe illness given infection. In the meantime, this framework for using limited available data to produce subpopulation-stratified estimates of the severity of illness and CFR/IFR by subpopulations can be generalized to other regional and policy contexts, provided generally-accessible data on epidemic time series and prevalences of marginal risk factors in the overall population.

600 *Contributors.* ALH and DVC designed the study. ALH wrote the paper with support from DVC. ALH developed the epidemic model and implemented the integrated model. DVC and LJ developed and implemented the conditional risk model. ALH implemented the integrated epidemic and risk model. NS helped to design the policy analysis. ALH led the analysis and visualization of results with
605 assistance from EH. FW prepared the LAC epidemiological data with support from WN and PS. NS, PS, KDH, MP, and WC contributed to writing the paper. DVC and WC coordinated management of the team, including data collection and processing. All authors read and approved the final paper.

NS reported receiving grants from Jedel Foundation, Abbott Diagnostics, the
610 Agency for Healthcare Research and Quality, the National Institutes of Health, the National Institute for Health Care Management Foundation, Health Care Services Corporation, and the Patient-Centered Outcomes Research Institute outside the submitted work; he reported serving as an expert witness for Goldman, Ismail, Tomaselli, Brennan, and Baum; serving as a scientific expert for
615 the American Medical Association, the China Development Research Foundation, and the Pharmaceutical Research and Manufacturers of America; serving as strategic advisor for Paysurance. All authors declare no competing interests.

Data sharing. For data and code required to reproduce the analysis see <https://github.com/AbigailHorn/COV2-LA>. The COVID-19 infection, hospitalization,
620 ventilation support, death, and age-stratified infection data analyzed in this paper was provided by the Outbreak Data Coordination Team of the Los Angeles County Department of Public Health, who consented to its sharing.

Acknowledgements. ALH was supported from an NIH Ruth L Kirschstein National Research Service Award (NRSA) Institutional Training Grant T32 5T32CA009492-
625 35. DVC was supported by NIH P01CA196569 and NCI P30CA014089. EH was supported by NIH P01CA196569. DVC and ALH are also funded by COVID-19 Keck Research Fund from the University of Southern California. We would like to acknowledge data processing and visualization support from Claire Jacquillat.

References

- 630 [1] Stokes EK, Zambrano LD, Anderson KN, Marder EP, Raz KM, Felix SEB, et al. Coronavirus disease 2019 case surveillance—united states, january 22–may 30, 2020. *Morbidity and Mortality Weekly Report* 2020;69(24):759.
- [2] Petrilli CM, Jones SA, Yang J, Rajagopalan H, O'Donnell LF, Chernyak Y, et al. Factors associated with hospitalization and critical illness among
635 4,103 patients with covid-19 disease in new york city. *MedRxiv* 2020;.
- [3] 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.

- 640 [4] Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, 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;.
- [5] Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Factors associated with covid-19-related death using opensafely. *Nature* 2020;584(7821):430–6.
- 645 [6] U.S. Department of Health and Human Services . Social determinants of health [online]. 2020. URL: <https://www.healthypeople.gov/2020/topics-objectives/topic/social-determinants-of-healthexternal>.
- [7] Egede LE, Walker RJ. Structural racism, social risk factors, and covid-19—a dangerous convergence for black americans. *New England Journal of Medicine* 2020;383(12):e77.
- 650 [8] Levin AT, Cochran KB, Walsh SP. Assessing the age specificity of infection fatality rates for covid-19: Meta-analysis & public policy implications. Tech. Rep.; National Bureau of Economic Research; 2020.
- 655 [9] Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet infectious diseases* 2020;.
- [10] Yang W, Kandula S, Huynh M, Greene SK, Van Wye G, Li W, et al. Estimating the infection-fatality risk of sars-cov-2 in new york city during the spring 2020 pandemic wave: a model-based analysis. *The Lancet Infectious Diseases* 2020;.
- 660 [11] 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>.
- 665 [12] Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*. Lippincott Williams & Wilkins; 2008.
- [13] Griffith G, Morris TT, Tudball M, Herbert A, Mancano G, Pike L, et al. Collider bias undermines our understanding of covid-19 disease risk and severity. *medRxiv* 2020;.
- 670 [14] 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.
- [15] 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.
- 675 [16] Mode CJ, Sleeman CK. *Stochastic processes in epidemiology: HIV/AIDS, other infectious diseases, and computers*. World Scientific; 2000.

- [17] Keeling MJ, Rohani P. Modeling infectious diseases in humans and animals. Princeton University Press; 2011.
- [18] 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. 680
- [19] 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;.
- [20] Bertozzi AL, Franco E, Mohler G, Short MB, Sledge D. The challenges of modeling and forecasting the spread of covid-19. *Proceedings of the National Academy of Sciences* 2020;117(29):16732–8. doi:10.1073/pnas.2006520117. arXiv:<https://www.pnas.org/content/117/29/16732.full.pdf>. 685
- [21] Unacast . The unacast social distancing scoreboard. 2020. URL: <https://unacast.com/post/the-unacast-social-distancing-scoreboard>. 690
- [22] 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;.
- [23] 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;. 695
- [24] Centers for Disease Control and Prevention (CDC) . National health and nutrition examination survey data (NHANES). 2017-2018. URL: <https://www.cdc.gov/nchs/nhanes/index.htm>. 700
- [25] of Public Health LACD. Los angeles county health survey (LACHS). 2018. URL: <http://publichealth.lacounty.gov/ha/hasurveyintro.htm>.
- [26] UCLA Center for Health Policy Research Los Angeles CA. California health information survey. 2018. URL: <http://ask.chis.ucla.edu>.
- [27] Los Angeles County Department of Public Health . Order of the health officer, reopening safer at work and in the community for control of covid-19. October 6, 2020. URL: http://www.ph.lacounty.gov/media/Coronavirus/docs/H00/2020_10_06_H00_Safer_at_Home.pdf. 705
- [28] Los Angeles Mayor’s Office . Safer l.a. covid-19. November 15, 2020. URL: <https://corona-virus.la/SaferLA>. 710
- [29] Los Angeles Mayor’s Office . Safer l.a. covid-19 timeline. November 24, 2020. URL: <https://lh3.googleusercontent.com/-awbM5oar9Bk/X71VEqOxBmI/AAAAAAAAAJ1w/fWXu03IaEYolw8fMxqC3VLAgE1m6vTeNQCK8BGAsYHg/s0/2020-11-24.png>.

- 715 [30] Adam D. Simulating the pandemic: What covid forecasters can learn from climate models. *Nature* 2020;.
- [31] Woolf SH, Chapman DA, Sabo RT, Weinberger DM, Hill L, Taylor DD. Excess deaths from covid-19 and other causes, march-july 2020. *Jama* 2020;324(15):1562–4.
- 720 [32] Bilinski A, Emanuel EJ. Covid-19 and excess all-cause mortality in the us and 18 comparison countries. *JAMA* 2020;.
- [33] Centers for Disease Control and Prevention (CDC) . Medical visits, hospitalizations, and deaths in the united states—2017–2018 influenza season [online] 2019;URL: <https://www.cdc.gov/flu/about/burden/2018-2019.html>.
- 725 [34] Centers for Disease Control and Prevention (CDC) . Health equity considerations and racial and ethnic minority groups [online] 2020;URL: <https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/race-ethnicity.html#fn1>.
- 730 [35] Tai DBG, Shah A, Doubeni CA, Sia IG, Wieland ML. The disproportionate impact of covid-19 on racial and ethnic minorities in the united states. *Clinical Infectious Diseases* 2020;.
- [36] Wetsman N. Doctors are better at treating covid-19 patients now than they were in march. 2020. URL: <https://www.theverge.com/2020/7/8/21317128/improved-covid-treatment-hospitals-remdesivir-dexamethasone>.
- 735 [37] Centers for Disease Control and Prevention (CDC) . interim clinical guidance for management of patients with confirmed 2019 novel coronavirus (2019-ncov) infection 2020;12. URL: <https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-guidance-management-patients.html>.
- 740 [38] Kifer D, Bugada D, Villar-Garcia J, Gudelj I, Menni C, Sudre CH, et al. Effects of environmental factors on severity and mortality of covid-19. *medRxiv* 2020;.
- 745 [39] Moriyama M, Hugentobler WJ, Iwasaki A. Seasonality of respiratory viral infections. *Annual review of virology* 2020;7.
- [40] Dahlberg M, Edin PA, Grönqvist E, Lyhagen J, Östh J, Siretskiy A, et al. Effects of the covid-19 pandemic on population mobility under mild policies: Causal evidence from sweden. *arXiv preprint arXiv:200409087* 2020;.
- 750 [41] Fuller T. California calls for residents 65 and older to stay at home 2020;URL: <https://www.nytimes.com/2020/03/15/us/coronavirus-newsom-california-seniors-restaurants-bars.html>.

- 755 [42] N M, M C, K K, et al. . The advisory committee on immunization practices' ethical principles for allocating initial supplies of covid-19 vaccine — united states, 2020. MMWR Morb Mortal Wkly November 27 2020;69(47);1782-1786. URL: <http://dx.doi.org/10.15585/mmwr.mm6947e3>.
- 760 [43] Goldman D, Conti D, Kahn M. Why covid-19 vaccines need to prioritize 'superspreaders'. The Conversation September 3 2020;URL: <https://theconversation.com/why-covid-19-vaccines-need-to-prioritize-superspreaders-145232>.