

Comparison of COVID-19 vaccine prioritization strategies

Lloyd A. C. Chapman PhD, Poojan Shukla, Isabel Rodríguez-Barraquer MD PhD, Priya B. Shete MD MPH, Tomás M. León MSc PhD, Kirsten Bibbins-Domingo MD PhD, George W. Rutherford MD, Robert Schechter MD and Nathan C. Lo MD PhD

Supplementary Information

Supplementary methods

Simulated California population

We simulated the entire population of California (N=39,148,760) with information for each individual on:

- demographic characteristics: age, sex, race, ethnicity and county of residence;
- unique special population status: healthcare worker (HCW), incarcerated individual (state/federal/local/youth prisoner), skilled nursing facility (SNF) resident, assisted living facility (ALF) resident, education worker, person experiencing homelessness, frontline essential worker, non-frontline essential worker;
- binary status on each of seven comorbidities: asthma, diabetes, heart disease, heart failure, hypertension, obesity, active tobacco use.

Age-, sex- and race-specific population totals for each of the 58 counties in California from the 2018 American Community Survey (ACS) 5-year estimates (1) were used to assign age category (0-4, 5-9, 10-14, 15-17, 18-19, 20-24, 25-29, 30-34, 35-44, 45-54, 55-64, 65-74, 75-84, 85+ years), sex and race for individuals of each county (see Table S1 for variable definitions). Integer ages within each age group were simulated by uniform random sampling. To incorporate binary ethnicity status E (Hispanic/Latino, non-Hispanic/Latino), we estimated the number of individuals of each ethnicity in each county-age-race group (denoted by subscript C, A, R), $N_{C,A,R,E}$, as:

$$N_{C,A,R,E} = N_{C,A,R} p_{E|C,A} p_{E|C,R}$$

where $p_{E|C,A}$ is the estimated proportion of the age group in the county that is of ethnicity E and $p_{E|C,R}$ is the proportion of the race group in the county that is of ethnicity E based on the ACS data, and randomly assigned this number of individuals in the county-age-race group to be of ethnicity E .

Overall population estimates for each of the special populations and their sources are given in Table S2. Special population status was assigned by drawing from a categorical distribution with probabilities for membership of each of the special populations dependent on one or more demographic factors. These probabilities were derived as follows. The probabilities of being a HCW and of being an education worker in each county-age-sex group, $p_{HCW|C,A,S}$ and $p_{educ|C,A,S}$ (subscript S denoting sex), were estimated as:

$$p_{X|C,A,S} = \frac{N_{X,C,S} p_{A|X}}{N_{C,A,S}}$$

where $X \in \{HCW, educ\}$, $N_{X,C,S}$ is the number of HCWs/education workers in each county-sex group according to the ACS data, $p_{A|X}$ is the proportion of HCWs/education workers

who are of age group A according to data from the Bureau of Labour Statistics (2), and $N_{C,A,S}$ is the number of individuals in the county-age-sex group. The probabilities of being a prisoner and of being a SNF resident in each age-sex-race-ethnicity group, $p_{prison|A,S,R,E}$ and $p_{SNF|A,S,R,E}$, were estimated as:

$$p_{Y|A,S,R,E} = \frac{N_{Y,A,S} p_{R|Y} p_{E|Y}}{N_{A,S} p_R p_E}$$

where $Y \in \{prison, SNF\}$, $N_{Y,A,S}$ is the state-level prisoner/SNF population of age A and sex S , $N_{A,S}$ is the state-level population of age group A and sex S , $p_{R|Y}$ and $p_{E|Y}$ are the state-level proportions of prisoners/SNF residents who are of race R and ethnicity E respectively, and p_R and p_E are the corresponding proportions for California residents.

The total ALF population was estimated as approximately 150,000 from the current capacity of ALFs in California (185,000) (3) and data on the occupancy of ALFs from the National Center for Health Statistics' (NCHS) National Study of Long-Term Care Providers (NSLTCP) 2015-16 (4), which suggests 81% occupancy of ALFs in California. The probability of being an ALF resident in age group A , $p_{ALF|A}$, was then estimated as:

$$p_{ALF|A} = \frac{N_{ALF} p_{A|ALF}}{N_A}$$

where N_{ALF} is the total ALF population, N_A is the population of age group A , and $p_{A|ALF}$ is the proportion of ALF residents who are in age group A , taken from the NCHS NSLTCP, assuming that all ALF residents are ≥ 55 years-of-age.

The probability of being a person experiencing homelessness in each county-age-sex-race group was calculated as:

$$p_{homeless|C,A,S,R} = \frac{N_{homeless,C,A} p_{S|homeless} p_{R|homeless}}{N_{C,A} p_S p_R}$$

where $N_{homeless,C,A}$ is the number of people experiencing homelessness in the county-age group from the US Department of Housing and Urban Development's (DHUD) Continuum of Care homelessness data (5), $N_{C,A}$ is the county-age group population, $p_{S|homeless}$ is the proportion of people experiencing homelessness in California who are of sex S from the DHUD data, and $p_{R|homeless}$ is the proportion of the homeless population who are of race R based on Los Angeles Almanac homelessness data (6).

The probabilities of being a frontline essential worker and of being a non-frontline essential worker in each age group were estimated as follows. The national-level percentages of workers in each age group in each of the 5 broad categories of occupations used by the US Bureau of Labor Statistics (Management and professional, Service, Sales and Office, Natural resources, Production and transportation) (2) who are frontline and non-frontline essential workers was calculated using the CDC classification of frontline and non-frontline essential

workers (7). These percentages were then multiplied by the percentages of each age group in California who are employed according to data from the 2018 ACS.

We estimated the probability of an individual in each county-age-sex-race group having each of the comorbidities ($p_{comorb|C,A,S,R}$) using data from the California Health Interview Survey (CHIS) (8) on prevalence of each comorbidity in each county by age ($p_{comorb|C,A}$), and that at the state level by sex, by race, and overall ($p_{comorb|S}$, $p_{comorb|R}$, and p_{comorb}), as:

$$p_{comorb|C,A,S,R} = p_{comorb|C,A} \frac{p_{comorb|S} p_{comorb|R}}{p_{comorb} p_{comorb}}$$

We used this probability to simulate binary status for each comorbidity for each individual in each county-age-sex-race group via a Bernoulli random draw. Obesity was defined as a body mass index (BMI) ≥ 30 , where BMI was simulated using CHIS data on the state-level mean BMI and standard deviation for each age group, adjusted by sex and race as in the above formula. Each individual's BMI was simulated via a random draw from a truncated normal distribution with the age-sex-and-race-adjusted mean and standard deviation, and truncation limits of 5 and 50.

COVID-19 case and death data

De-identified individual-level data on 2,215,972 confirmed COVID-19 cases (individuals who tested PCR positive for SARS-CoV-2) in California up to December 30, 2020, was provided by the California Department of Public Health (CDPH). Among these cases, 28,175 individuals died due to COVID-19. The data included information on county of residence, age category, sex, race/ethnicity, date of positive PCR test result, and, where applicable, date of death (see Table S3 for variable definitions). Complete demographic data was available for 1,578,988 (71%) cases and 27,580 (98%) individuals who died.

Overview of COVID-19 simulation model

We simulated COVID-19 deaths, clinical cases, infections and disability-adjusted life-years (DALYs) using the following steps, each of which is described below in further detail:

1. Simulate COVID-19 deaths using Poisson regression model fitted to COVID-19 death data with death hazard rate adjusted by special population status and comorbidity status, and calculate associated DALYs.
2. Estimate infections from predicted deaths using published age- and sex-dependent COVID-19 infection fatality rate (IFR).
3. Estimate clinical cases from predicted infections using published age-dependent clinical fraction for COVID-19.
4. Calculate DALYs associated with infections and clinical cases.

Table S1. Definitions and sources of variables in simulated California population

| Variable | Values | Source |
|---------------------|---|--|
| County of residence | Alameda, Alpine, Amador, Butte, Calaveras, Colusa, Contra Costa, Del Norte, El Dorado, Fresno, Glenn, Humboldt, Imperial, Inyo, Kern, Kings, Lake, Lassen, Los Angeles, Madera, Marin, Mariposa, Mendocino, Merced, Modoc, Mono, Monterey, Napa, Nevada, Orange, Placer, Plumas, Riverside, Sacramento, San Benito, San Bernardino, San Diego, San Francisco, San Joaquin, San Luis Obispo, San Mateo, Santa Barbara, Santa Clara, Santa Cruz, Shasta, Sierra, Siskiyou, Solano, Sonoma, Stanislaus, Sutter, Tehama, Trinity, Tulare, Tuolumne, Ventura, Yolo, Yuba | (1) |
| Age | Integer between 0 and 100 | (1) |
| Sex | Male, Female | (1) |
| Race | White, African American, American Indian or Alaska Native, Asian Alone, Native Hawaiian or Pacific Islander, Some other race alone, Two or more races | (1) |
| Ethnicity | Hispanic/Latino, non-Hispanic/Latino | (1) |
| Special population | Healthcare worker, Prisoner, Skilled Nursing Facility Resident, Assisted Living Facility Resident, Education Worker, Person Experiencing Homelessness, Frontline Essential Worker, Non-Frontline Essential Worker | See Table S2 |
| Asthma | 0, 1 | “AB40” variable in (8) |
| Diabetes | 0, 1 | “DIABETES” variable in (8) |
| Heart Disease | 0, 1 | “AB34” variable in (8) |
| Heart Failure | 0, 1 | “AB52” variable in (8) |
| Hypertension | 0, 1 | “AB29” variable in (8) |
| Obesity | 0, 1 | BMI \geq 30 where BMI simulated using “BMI” variable in (8) (see text) |
| Active tobacco use | 0, 1 | “SMKCUR” variable in (8) |

Table S2. Population estimates for special populations in California

| Special population | Population estimate* | Source |
|------------------------------------|----------------------|-----------|
| Healthcare workers | 1,559,001 | (1) |
| Prisoners | 227,428 | (1) |
| Skilled nursing facility residents | 136,282 | (1) |
| Assisted living facility residents | 150,340 | (3, 4) |
| Education workers | 962,684 | (1) |
| Persons experiencing homelessness | 132,427 | (5) |
| Frontline essential workers | 2,521,235 | (1, 2, 7) |
| Non-frontline essential workers | 6,994,107 | (1, 2, 7) |

* Note that population totals reflect values simulated as described in the text, which may not exactly match totals given in sources.

Table S3. Definitions of variables in COVID-19 case and death data used in analysis

| Variable | Definition | Values | Modifications |
|-------------------|----------------------------------|---|---|
| county_res | County of residence of case | Alameda, Central Sierra, Contra Costa, Fresno, Greater Sacramento, Imperial, Kern, Los Angeles, Marin, Merced, Monterey, Napa_Sonoma, Northern California, Northern Sacramento Valley, Orange, Placer, Riverside, Sacramento, San Bernardino, San Diego, San Francisco, San Joaquin, San Joaquin Valley, San Luis Obispo, San Mateo, Santa Barbara, Santa Clara, SantaCruz_SanBenito, Solano, Stanislaus, Tulare, Ventura | Counties with population <250,000 combined into single region (except Imperial), San Benito combined with Santa Cruz, Napa combined with Sonoma |
| age_cat | Age category | <10, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80+ | |
| sex | Sex | Male, Female | |
| race_ethnicity | Combined race/ethnicity variable | Hispanic/Latino, non-Hispanic Black, non-Hispanic White, Other | |
| first_report_date | Date of first report | 02/03/2020 - 12/30/2020 | Randomly moved forward or back 1 day |
| date_of_death | Date of death | 02/05/2020 - 12/30/2020 | Randomly moved forward or back 1 day |

To account for variation in the IFR across different settings, we recalibrated the age- and sex-dependent ensemble IFR estimate from (9) for California, using a combination of data on confirmed cases and deaths from CDPH up to late October, 2020; seroprevalence estimates for California up to October 8-13, 2020 (from 6 biweekly surveys of approximately 980 individuals), from the CDC’s database of nationwide commercial laboratory seroprevalence surveys (10); and age-stratified data on numbers of COVID-19 deaths in California among long-term care facility (LTCF) residents (SNF and ALF residents) and non-LTCF residents provided by CDPH. The IFR estimate from (9) is for non-LTCF residents and LTCF residents account for approximately 30% of COVID-19 deaths in California, so we used the age-and-LTCF-stratified death data to estimate the number of deaths in each demographic risk group, of 19,501 cumulative deaths up to October 31, 2020, that were among non-LTCF residents. We then divided this number by the IFR to estimate the number of infections among non-LTCF residents in each demographic risk group, and multiplied by the estimated age-dependent clinical case fraction from (11) (which fitted an age-structured transmission model to age-stratified clinical case data from 6 countries to the estimate the clinical fraction) to estimate the number of clinical cases among non-LTCF residents. We calculated an overall multiplier for the IFR of 2.64 to match the cumulative number of clinical cases among non-LTCF residents from our model with the cumulative number of non-LTCF confirmed cases in the CDPH case line list up to October 22, 2020 (827,867) (accounting for a mean reporting-to-death time of 9 days calculated from the data). We

note that this does not account for the fact that some confirmed cases up to October 22, 2020, were asymptomatic and that symptomatic cases were under-reported early in the pandemic. However, these two biases will have acted in opposite directions and information on whether confirmed cases were asymptomatic was unavailable, so we assume that the cumulative number of confirmed cases up to October 22, 2020, provides a reasonable approximation of the cumulative number of clinical cases up to that date. We verified that this approach gave a reasonable match between the age distribution of clinical cases from the model and that of confirmed cases (Figure 2A in main text), and between the age distribution of cumulative infections estimated from the model (2,433,000) and that estimated from the mean seroprevalence estimates for California up to October 8-13, 2020 (1,927,000) (using an approximately 12-day infection-to-reporting delay between the seroprevalence data centred on October 10 and case data calculated from the median onset-to-reporting delay in the data and incubation period (12)) (Figure 2B in main text). We note that the estimate of the number of individuals infected based on the seroprevalence data is likely to be an underestimate of the true number as it does not account for waning of antibodies (13).

As a further check on this approach, we performed a Poisson regression of cumulative case numbers against demographic risk factors (age, sex, race/ethnicity and county of residence) for both the estimated clinical case counts and confirmed case counts and compared the resulting parameter estimates. Overall, the two sets of parameter estimates agreed closely with each other (Figure S1). The hazard ratio estimates from the estimated clinical case counts were higher than those from the confirmed case counts for certain counties, possibly suggesting greater under-reporting of cases in these counties and/or higher IFRs, e.g., due to overburdened hospitals.

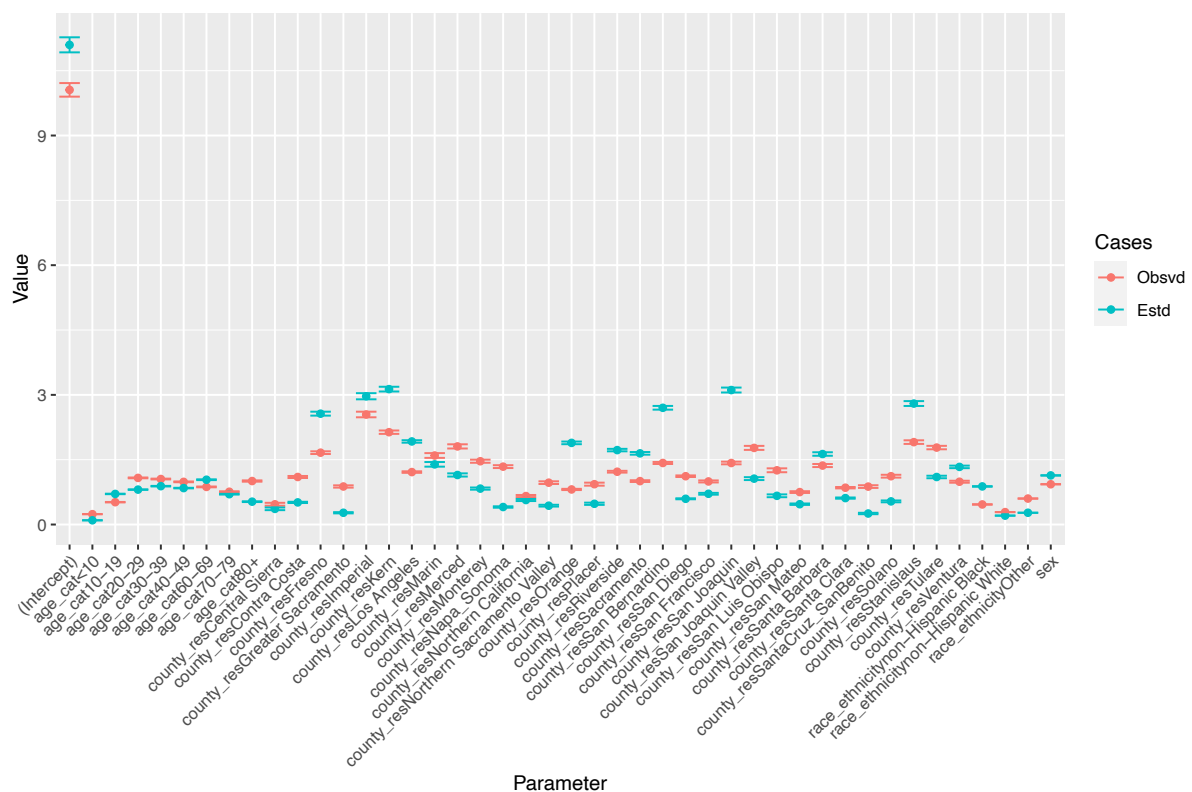


Figure S1. Model calibration check. Parameter estimates for Poisson regressions of COVID-19 case risk associated with different demographic risk factors for confirmed cases up to October 22, 2020, from the California Department of Public Health (red) and clinical cases estimated from our model (turquoise). Logarithm of total exposure time used as an offset variable, such that intercept represents COVID-19 case incidence rate (cases/100,000 person-days) among 50-59-year-old Hispanic/Latino females in Alameda County. Dots show maximum likelihood parameter estimates and error bars show 95% CIs.

Poisson regression model for COVID-19 death risk

We applied a Poisson regression model to the CDPH COVID-19 death data to estimate the hazard ratios for COVID-19 death associated with different demographic risk factors, namely county of residence, age, sex and race/ethnicity. The outcome variable was the cumulative number of deaths for each combination of the risk factors and the logarithm of the cumulative survival time for individuals with each combination of the risk factors was used as an offset variable to account for differences in time to death between individuals. Representing the different risk factors (county, age group, sex, race/ethnicity) by a vector of p covariates $\mathbf{X} = (X_1, \dots, X_p)$ whose value is indexed by the index set (i_1, \dots, i_p) , and the total number of COVID-19 deaths and total survival time of individuals with covariate levels (i_1, \dots, i_p) by $Y_{i_1 \dots i_p}$ and $T_{i_1 \dots i_p}$ respectively (such that $Y_{i_1 \dots i_p} / T_{i_1 \dots i_p}$ is the COVID-19 death rate among individuals with covariate levels (i_1, \dots, i_p)), the Poisson regression model is:

$$Y_{i_1 \dots i_p} \sim \text{Po}(\mu_{i_1 \dots i_p})$$

with

$$\log(\mu_{i_1 \dots i_p}) = \log(T_{i_1 \dots i_p}) + \log(\lambda_0) + \mathbf{X}^T \boldsymbol{\beta},$$

where $\mu_{i_1 \dots i_p}$ is the conditional mean of the Poisson distribution for the cumulative number of deaths, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ is a column vector of the covariate coefficients (such that e^{β_i} is the hazard ratio for covariate X_i), and λ_0 is the baseline COVID-19 death rate in California (assumed constant). This model is equivalent to a proportional hazards model for time to COVID-19 death (14, 15), with hazard rate:

$$\lambda_{i_1 \dots i_p} = \frac{\mu_{i_1 \dots i_p}}{T_{i_1 \dots i_p}} = \lambda_0 e^{\mathbf{X}^T \boldsymbol{\beta}}$$

Hazard ratios and their standard errors were estimated using the `glm` function in R (version 4.0.3) (16). We compared parameter estimates for death data from different periods of time (the last 3 months and 6 months of 2020, and time from the first death to the end of 2020) to assess their stability over time.

Adjustment of COVID-19 death risk by special population status and comorbidity status

We adjusted risk of COVID-19 death by special population status and comorbidity status using published literature estimates for the relative risk of SARS-CoV-2 infection and death from COVID-19 for the special populations and hazard ratios for death given SARS-CoV-2 infection associated with the comorbidities (Table S4). For LTCF residents, we used estimates of age-specific relative risk of COVID-19 death calculated from the CDPH age-and-

LTCF-stratified data. For essential workers, we calculated estimates of the relative risk of COVID-19 death for frontline and non-frontline essential workers using estimates of relative risk of death by occupation from a recent analysis of excess mortality in California during the pandemic (17), categorising employment sectors according to the CDC classification of essential workers (7). For each special population j for which data on relative risk of death was not available (healthcare workers, prisoners, education workers, and people experiencing homelessness), we derived an estimate of the relative risk of death, RR_j , by multiplying an estimate of relative risk of infection in the special population, $RR_{inf,j}$ (from published literature or calculated from publicly available data on cumulative incidence of confirmed cases in the special population), by an estimate of the relative risk of death given infection, $RR_{death|inf,j}$, derived from the age- and sex-specific IFR, to account for differences in age distributions between special populations:

$$RR_j = RR_{inf,j} RR_{death|inf,j} = RR_{inf,j} \frac{\overline{IFR}_j}{\overline{IFR}}$$

where \overline{IFR}_j is the population-weighted average IFR for special population j and \overline{IFR} is the population-weighted average IFR for the general population. We assumed that comorbidities only affected risk of death once infected, and that they had a multiplicative effect on risk of death. The death rate for each demographic risk factor group $\lambda_{i_1 \dots i_p}$ was multiplied by the following factor to ensure the overall death risk remained the same (i.e. to account for the higher death risk for individuals in special populations and with comorbidities):

$$\frac{N_{i_1 \dots i_p}}{\sum_{j, k_1, \dots, k_7} RR_{jk_1 \dots k_7} N_{i_1 \dots i_p j k_1 \dots k_7}}$$

where $N_{i_1 \dots i_p}$ is the number of individuals with demographic risk factor levels (i_1, \dots, i_p) , $RR_{jk_1 \dots k_7}$ is the relative risk of death associated with special population j (where $j = 0$ represents individuals that are not part of one of the 8 special populations) and comorbidity status (k_1, \dots, k_7) (the particular combination of binary statuses for the 7 comorbidities), and $N_{i_1 \dots i_p j k_1 \dots k_7}$ is the number of individuals in each demographic-special-population-comorbidity-status subgroup.

Estimation of number of individuals already infected

We estimated the number of individuals already infected in each demographic risk group by dividing the observed cumulative number of deaths up to December 30, 2020, in each group by the age- and sex-specific recalibrated IFR for that group (see *Calibration of infection fatality rate for California*). For LTCF residents, we first multiplied the IFR by a frailty index representing increased risk of death given infection among LTCF residents compared to the general population, which we assumed to be 3 based on previous estimates (9, 18). The probability of past infection in each age-sex-special-population group was calculated by dividing the estimated number of infected individuals by the group population size, and used to simulate whether or not individuals in the simulated population had been previously infected. Individuals simulated as previously infected were treated as being immune to reinfection with SARS-CoV-2, based on apparently limited confirmed instances of

reinfection and data suggesting protective antibody responses to infection last for at least 6 months (19–21).

Table S4. Estimates of relative risks of SARS-CoV-2 infection and COVID-19 death for special populations and hazard ratios for death given infection for different comorbidities

| Parameter | Distribution/Value* | Source |
|--|-----------------------------|--|
| Relative risk of SARS-CoV-2 infection for special population (Ref. = general population)** | | |
| Healthcare worker | $TN(3.4, 0.04, 3, 4)$ | (22) |
| Prisoner | $TN(5.5, 0.07, 5, 6)$ | (23) |
| Person experiencing homelessness | $TN(1.65, 0.006, 1.5, 1.8)$ | (24, 25) |
| Education worker | $TN(1.8, 0.09, 1.2, 2.8)$ | (26) |
| Relative risk of COVID-19 death for special population (Ref. = general population) | | |
| Skilled nursing facility resident | Age-dependent | California Department of Public Health |
| Assisted living facility resident | Age-dependent | California Department of Public Health |
| Frontline essential worker*** | $TN(1.2, 0.01, 1, 1.4)$ | (17) |
| Non-frontline essential worker*** | $TN(1.1, 0.003, 1, 1.2)$ | (17) |
| Hazard ratio for COVID-19 death given infection for individual with comorbidity | | |
| Asthma | 1 | (27, 28) |
| Diabetes | 1.51 | (29) |
| Heart disease | 1.16 | (27–29) |
| Heart failure | 1.77 | (28) |
| Hypertension | 1 | (28, 29) |
| Obesity | 1.33 | (27) |
| Smoker | 1 | (27–29) |

* $TN(\mu, \sigma^2, a, b)$ = truncated normal distribution with mean μ , non-truncated variance σ^2 and left and right truncation limits a and b . Uncertainty bounds for relative risks of infection reflect a combination of statistical uncertainty (95% CIs for published estimates) and uncertainty due to variation in values from different sources. Death hazard ratios with 95% CI overlapping 1 in original source taken to be 1.

** Further details of the sources and calculations used to derive these values are given in File S1.

*** Estimated as mean of relative risks of death for workers in different employment sectors from (17) categorised into frontline vs non-frontline essential work using the CDC classification (7).

Simulation of COVID-19 deaths

To simulate COVID-19 deaths, we first calculated the cumulative probability of death for individuals in the simulated population over 6 months according to their demographic characteristics, special population status, comorbidities, simulated history of infection, and whether or not they were vaccinated. The cumulative probability of death over 6 months ($T = 182$ days) for a susceptible unvaccinated individual with covariate levels (i_1, \dots, i_p) in special population j with comorbidity status (k_1, \dots, k_7) was:

$$p_{i_1 \dots i_p j k_1 \dots k_7} = 1 - e^{-\lambda_{i_1 \dots i_p} RR_{j k_1 \dots k_7} T}.$$

We assumed that the vaccine is ‘leaky’ and has efficacy $v_e \in [0,1]$, i.e. reduces the relative risk of death by a factor $(1 - v_e)$. The cumulative probability of death over 6 months for a vaccinated individual not previously infected was therefore:

$$p_{i_1 \dots i_p j k_1 \dots k_7} = 1 - e^{-(1-v_e)\lambda_{i_1 \dots i_p} RR_{j k_1 \dots k_7} T}.$$

The number of deaths in each demographic-special-population-comorbidity risk group was simulated as $X_{i_1 \dots i_p j k_1 \dots k_7} \sim \text{Bin}(N_{i_1 \dots i_p j k_1 \dots k_7} - I_{i_1 \dots i_p j k_1 \dots k_7}, p_{i_1 \dots i_p j k_1 \dots k_7})$, where $I_{i_1 \dots i_p j k_1 \dots k_7}$ is the number of previously infected individuals in the risk group. We ran 1000 simulations of past infections and deaths for the full California population ($N = 39,148,760$) to account for stochastic uncertainty. We also accounted for uncertainty in the estimated death rate and relative risks of infection for the special populations, by drawing values for $\log(\lambda_{i_1 \dots i_p})$ from truncated normal distributions with bounds and standard deviations derived from the 95% confidence intervals (CIs) for the regression coefficients (Table S5), and values for the relative risks from truncated normal distributions with bounds and standard deviations derived from uncertainty in literature estimates (Table S4).

Estimation of infections and clinical cases

Simulated deaths from COVID-19 over the first 6 months of vaccine rollout were summed for each demographic-special-population-comorbidity risk group. The cumulative number of SARS-CoV-2 infections in each demographic-special-population-comorbidity risk group was then calculated from the simulated number of deaths by dividing by the recalibrated age- and sex-dependent IFR. Numbers of clinical cases were then calculated from estimated numbers of infections by multiplying by the estimated age-dependent clinical case fraction from (11).

Calculation of DALYs

DALYs associated with illness and/or death due to a disease are given by the sum of years of life lost (YLL) and years lived with disability (YLD):

$$DALYs = YLL + YLD$$

and thus provide a composite measure of mortality and morbidity. Years lived with disability depend on the duration of illness d and the disability weight $w \in [0,1]$ associated with the illness (where 0 represents perfect health and 1 represents death):

$$YLD = wd.$$

Total DALYs associated with simulated COVID-19 deaths, $DALYs_d$, were calculated as:

$$DALYs_d = \sum_A n_{d,A} (\overline{YLL}_A + w_s d_s),$$

i.e. by adding the average remaining life expectancy \overline{YLL}_A of each individual in age group A who died, calculated from 2018 US life tables (30) and the age, sex, and race/ethnicity distribution within the age group, to YLD associated with severe illness from COVID-19,

$w_s d_s$, multiplying by the number of deaths in the age group, $n_{d,A}$, and summing over all age groups. Total DALYs from clinical cases, $DALY_{S_c}$, and subclinical infections, $DALY_{S_i}$, were calculated as:

$$\begin{aligned} DALY_{S_c} &= n_c w_m d_m \\ DALY_{S_i} &= (n - n_c) w_{mi} d_{mi} \end{aligned}$$

where n_c and n are the estimated numbers of clinical cases and overall infections respectively, and $w_m d_m$ and $w_{mi} d_{mi}$ are the YLD associated with clinical and subclinical infection respectively. We used estimates of disability weights for acute episodes of mild, moderate and severe illness from (31) for subclinical infection, clinical infection and death from COVID-19 respectively: $w_{mi} = 0.005$, $w_m = 0.053$, $w_s = 0.210$. For the corresponding durations of illness d_{mi} , d_m and d_s , we used data on symptom duration stratified by symptom severity from a study of 273 COVID-19 outpatients in Atlanta, Georgia (32) and data from other studies and systematic reviews (33–35) that suggests median durations of symptoms for subclinical infection, clinical infection, and severe illness prior to death are 7 days, 10 days and 18 days respectively.

Approximate 95% uncertainty intervals (UIs) for predicted outcomes were calculated as the 2.5%-97.5% quantiles of the distributions of the simulated outcomes.

Vaccine prioritization

For all vaccine prioritization strategies, we assumed that all HCWs and LTCF residents were vaccinated first as per the CDC guidelines for the first phase of the vaccine rollout (36). For the remainder of the population, we calculated the expected average risk of DALYs per person per day for all subgroups under the prioritization strategy, and then ranked subgroups in descending order of DALY risk. Vaccines were then allocated to these subgroups in this order and randomly within each subgroup until the total number of vaccines available, n_v , was reached. So, e.g., for age targeting, after vaccination of HCWs and LTCF residents, vaccinations were allocated to the age group with the highest average DALY risk and then to age groups with progressively decreasing DALY risk until the vaccine quota was reached. The random allocation of vaccinations within subgroups was varied in each of the 1000 simulations to account for uncertainty from variation in the individuals vaccinated before the vaccine quota was met. Three different initial vaccine quotas were considered ($n_v = 2$ million, 5 million, 10 million), reflecting variation in initial vaccine availability.

Under special population targeting, essential workers (frontline and non-frontline) were grouped with the section of the population not belonging to any special population. Under essential worker targeting, frontline and non-frontline essential workers were grouped together and the rest of the non-HCW, non-LTCF-resident population was treated as one group, and allocation was assumed to be random within these two groups. For comorbidity targeting, we grouped individuals into those with any comorbidities and those with no comorbidities, and assumed allocation was random within these groups, on the basis that targeting by number of comorbidities would be practically and economically infeasible. Under age-and-special-population targeting, essential workers were included among the special populations targeted for vaccination. In all strategies, excess vaccines remaining

after complete target group coverage were randomly allocated amongst the remaining population.

Data and code availability

All analysis code was developed in R version 4.0.3 (16) and is available online at <https://github.com/LloydChapman/COVIDVaccineModelling>. The CDPH case data required for fitting the Poisson regression model contain personally identifiable information and therefore cannot be made freely available. Individuals interested in accessing the data should contact CDPH. All the data required to run the vaccine prioritization simulations is available at <http://doi.org/10.5281/zenodo.4516526>.

Supplementary results

Regression model parameter estimates

Parameter estimates for the Poisson regression model (Table S5 and Figure S2) show significant variation in COVID-19 death risk across counties, with hazard ratios (HRs) relative to Alameda County varying from 0.38 for Northern California to 3.7 for Imperial County (when fitting to all data since the first recorded death on February 5, 2020). The hazard ratio estimates for age, sex and race/ethnicity reflect the increase in death risk with increasing age, higher death risk for males, and higher death risk for Hispanic/Latino individuals described in the main text. There was some variation in parameter estimates when fitting to different periods of past data (the last 3, 6, and 11 months of 2020), e.g., a few counties went from being lower risk than Alameda County to higher and vice versa, and there was a slight downward shift in the age distribution of deaths over time, but overall parameter estimates were highly consistent (Table S5).

Vaccine availability

The order of impact of the different prioritization strategies on DALYs, deaths and cases was robust to variation in the vaccine availability. However, with 2 million vaccinations instead of 5 million, i.e. only sufficient doses to vaccinate 5% of California's population rather than 13%, none of the prioritization strategies targeting by a single risk factor averted significantly more DALYs than random allocation over 6 months since the vast majority of the 2 million vaccinations were used up in vaccinating HCWs and LTCF residents (~1.85 million individuals in California) in the first phase of the rollout, and prioritization of the remaining 150,000 vaccinations made only a small difference to the numbers of DALYs averted. Prioritizing older individuals averted 19% (95% UI 18-20%) of the overall DALY burden under no vaccination, while prioritizing special populations, individuals with comorbidities, essential workers and random allocation all averted 18% (95% UI 17-19%) (Table S6 and Figure S3E). However, targeting across all risk factors simultaneously averted a higher proportion of DALYs (24%, 95% UI 23-25%). There was little difference in impact on cases between the different strategies – the percentage of cases averted was between 7% and 8% for all strategies (Table S6 and Figure S3A) – but allocating the vaccinations remaining after the first phase to older individuals averted more deaths (32%) than allocating them to special populations, individuals with comorbidities, or essential workers (all 29%).

With double the vaccine availability – 10 million vaccinations instead of 5 million, i.e., enough vaccine to vaccinate just over a quarter of California's population – age targeting

averted by far and away the most DALYs of the strategies targeting by a single risk factor and 66% (95% UI 64-67%) of the simulated overall burden over 6 months (Table S7 and Figure S3F). This was principally due to averting a much higher proportion of deaths than any of the other strategies – 83% (95% UI 82-83%) compared to 57% (95% UI 55-58%) for comorbidity targeting as the next best performing strategy (Table S7 and Figure S3D). Special population targeting and essential worker targeting averted only 40% (95% UI 39-42%) and 37% (95% UI 36-38%) of the DALY burden respectively, due to only averting 47% (95% UI 46-49%) and 39% (95% UI 38-40%) of deaths. As for 5 million individuals vaccinated, age-and-county targeting and targeting across all risk factors simultaneously averted a higher proportion of DALYs (71% and 76% respectively) than any of the strategies targeting by a single risk factor. Essential worker targeting averted the highest proportion of clinical cases (31%, 95% UI 28-35%), and age targeting averted the lowest proportion (27%, 95% UI 24-29%).

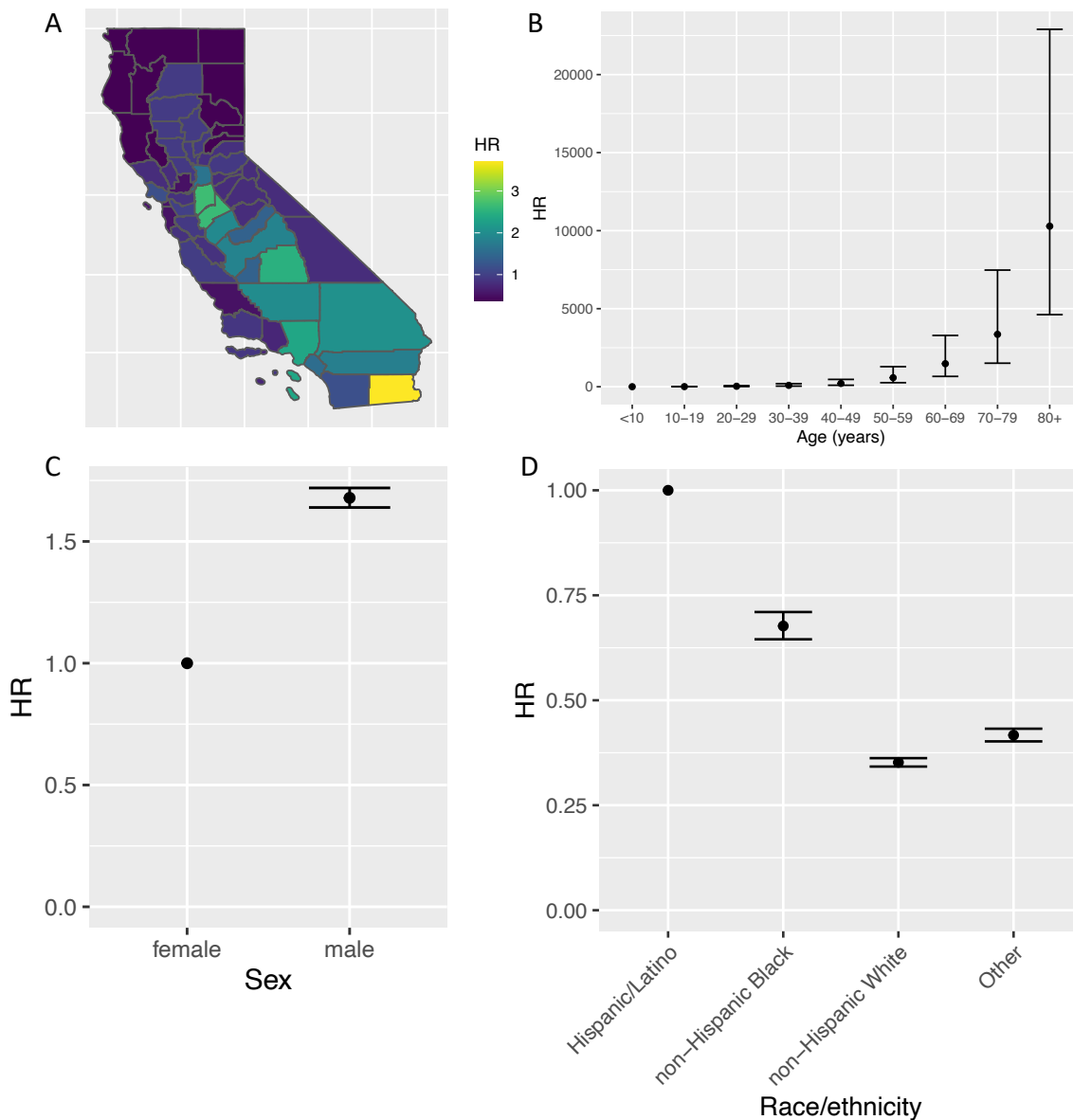


Figure S2. Estimates of hazard ratios for risk of death from COVID-19 for demographic risk factors. Hazard ratios (HR) for (A) county of residence, (B) age, (C) sex, and (D) race/ethnicity. Estimates from Poisson regression model fitted to data on 28,175 COVID-19 deaths in California from February 5, 2020 to December 30, 2020, from the California Department of Public Health. Counties with population <250,000 (except Imperial) were combined into a single region by their economic region, San Benito County was combined with Santa Cruz County, and Napa County was combined with Sonoma County. Plotted hazard ratios for these counties represent the hazard ratio of the combined region.

Table S5. Sensitivity of COVID-19 death risk model parameter estimates to time period of data used.

| Parameter | Maximum likelihood estimate (95% CI) | | |
|--|--------------------------------------|----------------------------|-----------------------------|
| | Feb 5, 2020 - Dec 30,2020 | Jun 30, 2020 - Dec 30,2020 | Sep 30, 2020 - Dec 30, 2020 |
| Baseline COVID-19 death hazard rate (deaths/million person-days) | 0.0023 (0.001-0.0051) | 0.0042 (0.0019-0.0094) | 0.0036 (0.0012-0.011) |
| HRs for counties | | | |
| Alameda | 1 (Ref.) | 1 (Ref.) | 1 (Ref.) |
| Central Sierra | 0.8 (0.63-1) | 1 (0.82-1.3) | 0.61 (0.39-0.95) |
| Contra Costa | 0.85 (0.75-0.96) | 0.9 (0.78-1) | 0.85 (0.69-1.1) |
| Fresno | 1.9 (1.7-2.1) | 2.2 (1.9-2.4) | 2.2 (1.8-2.6) |
| Greater Sacramento | 0.93 (0.79-1.1) | 1.1 (0.89-1.2) | 1.7 (1.4-2.2) |
| Imperial | 3.7 (3.3-4.2) | 2.8 (2.4-3.3) | 2.2 (1.7-2.8) |
| Kern | 2 (1.8-2.2) | 2.2 (2-2.5) | 1.9 (1.6-2.3) |
| Los Angeles | 2.4 (2.2-2.5) | 2.1 (1.9-2.3) | 2.7 (2.3-3.1) |
| Marin | 1.2 (0.98-1.4) | 1.3 (1.1-1.6) | 0.51 (0.33-0.8) |
| Merced | 1.9 (1.7-2.3) | 2.4 (2-2.8) | 2.3 (1.8-3) |
| Monterey | 0.97 (0.82-1.1) | 1.2 (0.98-1.4) | 1.8 (1.4-2.3) |
| Napa-Sonoma | 0.79 (0.68-0.92) | 0.98 (0.84-1.2) | 0.83 (0.64-1.1) |
| Northern California | 0.38 (0.3-0.47) | 0.47 (0.38-0.6) | 0.56 (0.41-0.78) |
| Northern Sacramento Valley | 0.95 (0.81-1.1) | 1.2 (1-1.4) | 1.7 (1.3-2.1) |
| Orange | 1.6 (1.4-1.7) | 1.6 (1.4-1.7) | 1.4 (1.2-1.6) |
| Placer | 0.83 (0.69-1) | 1 (0.83-1.2) | 1.4 (1.1-1.9) |
| Riverside | 1.8 (1.6-2) | 1.7 (1.6-1.9) | 1.9 (1.7-2.3) |
| Sacramento | 1.7 (1.5-1.8) | 2 (1.8-2.2) | 2.3 (2-2.8) |
| San Bernardino | 2.1 (1.9-2.2) | 2.1 (1.9-2.4) | 1.8 (1.6-2.2) |
| San Diego | 1.2 (1.1-1.3) | 1.2 (1.1-1.3) | 1.8 (1.5-2.1) |
| San Francisco | 0.55 (0.47-0.65) | 0.54 (0.45-0.65) | 0.57 (0.44-0.75) |
| San Joaquin | 2.7 (2.4-3) | 3.2 (2.8-3.6) | 2.1 (1.7-2.6) |
| San Joaquin Valley | 1.4 (1.2-1.7) | 1.4 (1.2-1.7) | 1 (0.72-1.4) |
| San Luis Obispo | 0.53 (0.41-0.68) | 0.68 (0.52-0.88) | 1 (0.72-1.4) |
| San Mateo | 0.53 (0.45-0.63) | 0.25 (0.19-0.32) | 0.24 (0.16-0.36) |
| Santa Barbara | 0.89 (0.75-1.1) | 0.95 (0.78-1.1) | 0.77 (0.57-1.1) |
| Santa Clara | 0.93 (0.83-1) | 0.92 (0.81-1) | 1.4 (1.1-1.6) |
| SantaCruz-SanBenito | 0.84 (0.69-1) | 1 (0.84-1.3) | 2 (1.6-2.6) |
| Solano | 0.52 (0.42-0.64) | 0.51 (0.4-0.65) | 0.3 (0.19-0.49) |
| Stanislaus | 2.6 (2.4-2.9) | 3.1 (2.7-3.5) | 2.4 (2-3) |
| Tulare | 2.5 (2.2-2.8) | 2.3 (2-2.7) | 2.8 (2.3-3.5) |
| Ventura | 0.73 (0.63-0.84) | 0.77 (0.66-0.9) | 0.84 (0.66-1.1) |
| HRs for age | | | |
| <10 | 1 (Ref.) | 1 (Ref.) | 1 (Ref.) |
| 10-19 | 6.5 (2.8-15) | 5.9 (2.5-14) | 5.5 (1.6-19) |

| | | | |
|------------------------|--------------------|-------------------|-------------------|
| 20-29 | 30 (13-67) | 25 (11-57) | 17 (5.5-56) |
| 30-39 | 88 (39-200) | 65 (29-140) | 55 (17-170) |
| 40-49 | 210 (94-470) | 160 (74-370) | 140 (45-430) |
| 50-59 | 580 (260-1300) | 450 (200-1000) | 390 (120-1200) |
| 60-69 | 1500 (660-3300) | 1200 (520-2600) | 1000 (320-3100) |
| 70-79 | 3400 (1500-7500) | 2600 (1200-5700) | 2500 (790-7600) |
| 80+ | 10000 (4600-23000) | 7700 (3400-17000) | 7300 (2400-23000) |
| HR for male sex | 1.7 (1.6-1.7) | 1.7 (1.6-1.7) | 1.7 (1.6-1.8) |
| HRs for race/ethnicity | | | |
| Hispanic/Latino | 1 (Ref.) | 1 (Ref.) | 1 (Ref.) |
| non-Hispanic Black | 0.68 (0.65-0.71) | 0.61 (0.57-0.64) | 0.58 (0.53-0.63) |
| non-Hispanic White | 0.35 (0.34-0.36) | 0.34 (0.32-0.35) | 0.38 (0.36-0.4) |
| Other | 0.42 (0.4-0.43) | 0.4 (0.39-0.42) | 0.47 (0.45-0.5) |

Parameter estimates for COVID-19 death risk regression model for different past periods of data on deaths (~11 months from date of first recorded COVID-19 death to end of data, 6 months, and 3 months) from the California Department of Public Health. The estimates derived from ~11 months of data were used in the simulation of the different vaccine prioritization strategies.

Table S6. Simulated impact of different COVID-19 vaccine prioritization strategies in California on numbers of averted COVID-19 clinical cases, deaths and DALYs for 2 million individuals vaccinated.

| Strategy | Cases averted, mean (95% UI) | Percentage of cases averted, mean (95% UI) | Deaths averted, mean (95% UI) | Percentage of deaths averted, mean (95% UI) | DALYs averted, mean (95% UI) | Percentage of DALYs averted, mean (95% UI) |
|---|------------------------------|--|-------------------------------|---|------------------------------|--|
| i) Random allocation | 46,000 (37,000-54,000) | 7% (6-9%) | 4,100 (4,000-4,200) | 29% (28-30%) | 38,000 (37,000-40,000) | 18% (18-19%) |
| ii) Special population targeting | 48,000 (39,000-56,000) | 8% (6-9%) | 4,100 (3,900-4,200) | 29% (28-30%) | 39,000 (37,000-40,000) | 18% (18-19%) |
| iii) Age targeting | 44,000 (36,000-53,000) | 7% (6-9%) | 4,500 (4,300-4,600) | 32% (31-33%) | 40,000 (38,000-42,000) | 19% (18-20%) |
| iv) Essential worker targeting | 46,000 (37,000-55,000) | 8% (6-9%) | 4,000 (3,900-4,200) | 29% (28-30%) | 38,000 (36,000-40,000) | 18% (17-19%) |
| v) Comorbidity targeting | 46,000 (37,000-55,000) | 8% (6-9%) | 4,100 (4,000-4,200) | 29% (28-30%) | 39,000 (37,000-40,000) | 18% (18-19%) |
| vi) Age-and-county targeting | 46,000 (38,000-54,000) | 7% (6-9%) | 4,800 (4,600-5,000) | 34% (33-35%) | 43,000 (41,000-44,000) | 20% (20-21%) |
| vii) Age-and-special-population targeting | 45,000 (37,000-53,000) | 7% (6-9%) | 4,500 (4,300-4,600) | 32% (31-33%) | 40,000 (39,000-42,000) | 19% (19-20%) |
| viii) Optimal allocation | 51,000 (43,000-60,000) | 8% (7-10%) | 5,200 (5,100-5,400) | 37% (36-38%) | 50,000 (48,000-51,000) | 24% (23-25%) |

Vaccine impact simulated over 6 months. All strategies assumed that all healthcare workers and long-term care facility residents were vaccinated first as per the CDC recommendation (36).

Table S7. Simulated impact of different COVID-19 vaccination prioritization strategies in California on numbers of averted COVID-19 clinical cases, deaths and DALYs for 10 million individuals vaccinated

| Strategy | Cases averted, mean (95% UI) | Percentage of cases averted, mean (95% UI) | Deaths averted, mean (95% UI) | Percentage of deaths averted, mean (95% UI) | DALYs averted, mean (95% UI) | Percentage of DALYs averted, mean (95% UI) |
|---|------------------------------|--|-------------------------------|---|------------------------------|--|
| i) Random allocation | 175,000 (157,000-193,000) | 29% (26-32%) | 6,800 (6,500-7,100) | 48% (47-50%) | 85,000 (81,000-89,000) | 41% (39-43%) |
| ii) Special population targeting | 175,000 (159,000-194,000) | 29% (26-32%) | 6,700 (6,400-7,000) | 47% (46-49%) | 84,000 (80,000-88,000) | 40% (39-42%) |
| iii) Age targeting | 162,000 (153,000-172,000) | 27% (24-29%) | 11,700 (11,400-11,900) | 83% (82-83%) | 137,000 (134,000-140,000) | 66% (64-67%) |
| iv) Essential worker targeting | 192,000 (173,000-212,000) | 31% (28-35%) | 5,500 (5,400-5,600) | 39% (38-40%) | 77,000 (74,000-80,000) | 37% (36-38%) |
| v) Comorbidity targeting | 187,000 (172,000-203,000) | 31% (28-34%) | 8,000 (7,700-8,200) | 57% (55-58%) | 100,000 (97,000-104,000) | 48% (47-49%) |
| vi) Age-and-county targeting | 201,000 (191,000-212,000) | 33% (30-36%) | 11,800 (11,700-12,100) | 84% (84-85%) | 149,000 (145,000-152,000) | 71% (70-72%) |
| vii) Age-and-special-population targeting | 164,000 (155,000-174,000) | 27% (25-30%) | 11,700 (11,500-11,900) | 83% (82-84%) | 138,000 (134,000-141,000) | 66% (65-67%) |
| viii) Optimal allocation | 244,000 (231,000-256,000) | 40% (37-44%) | 11,900 (11,700-12,200) | 85% (84-85%) | 158,000 (154,000-161,000) | 76% (75-77%) |

Vaccine impact simulated over 6 months. All strategies assumed that all healthcare workers and long-term care facility residents were vaccinated first as per the CDC recommendation (36).

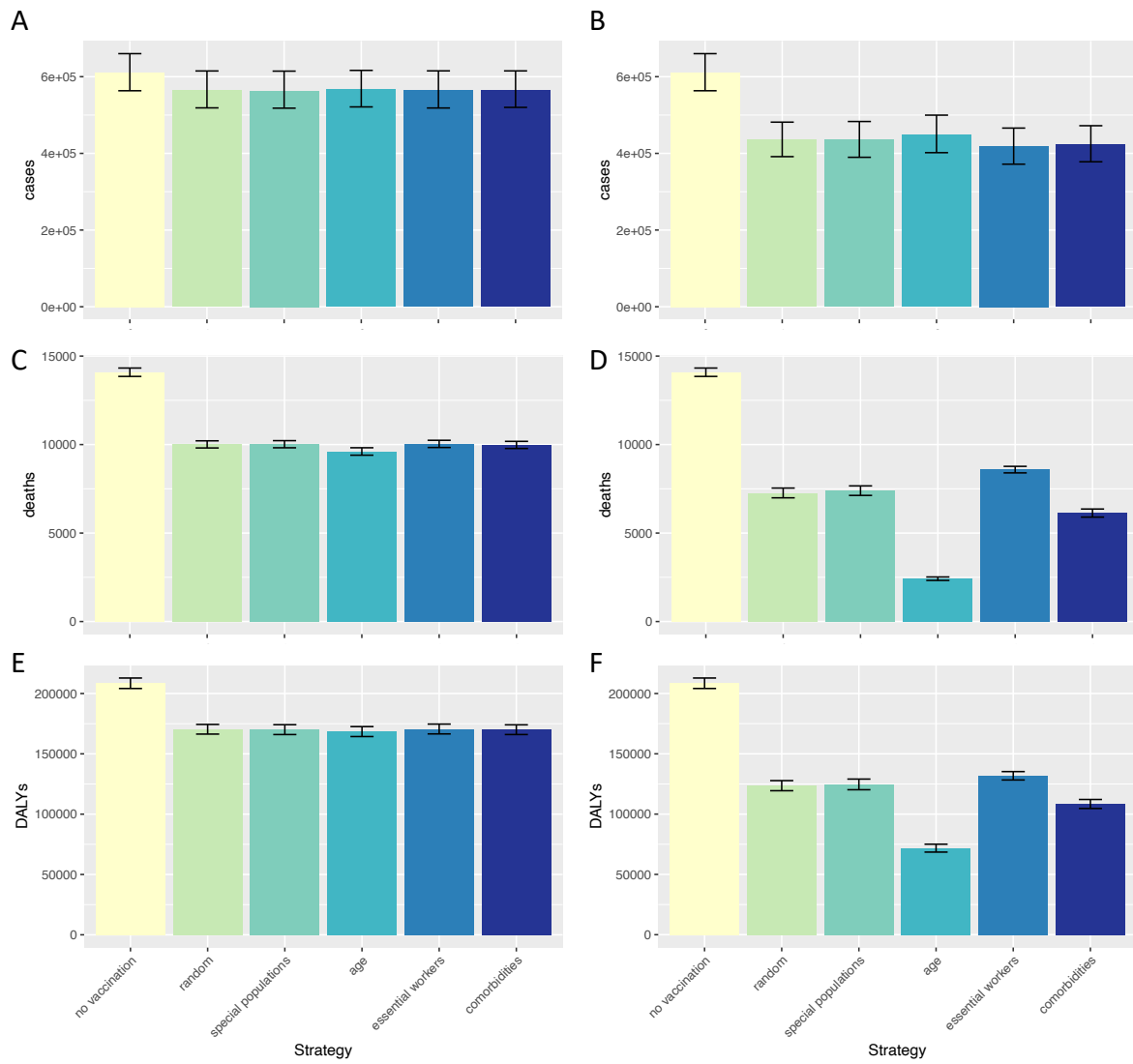


Figure S3. Simulated impact of different COVID-19 vaccine prioritization strategies in California on numbers of COVID-19 clinical cases, deaths and DALYs for 2 million (left column) and 10 million (right column) individuals vaccinated. Simulated (A-B) clinical cases, (C-D) deaths and (E-F) DALYs in California in the first 6 months of 2021. The vaccine was assumed to have 95% age-independent efficacy for preventing clinical disease and infection. Black lines show 95% uncertainty intervals from stochastic uncertainty and parameter uncertainty.

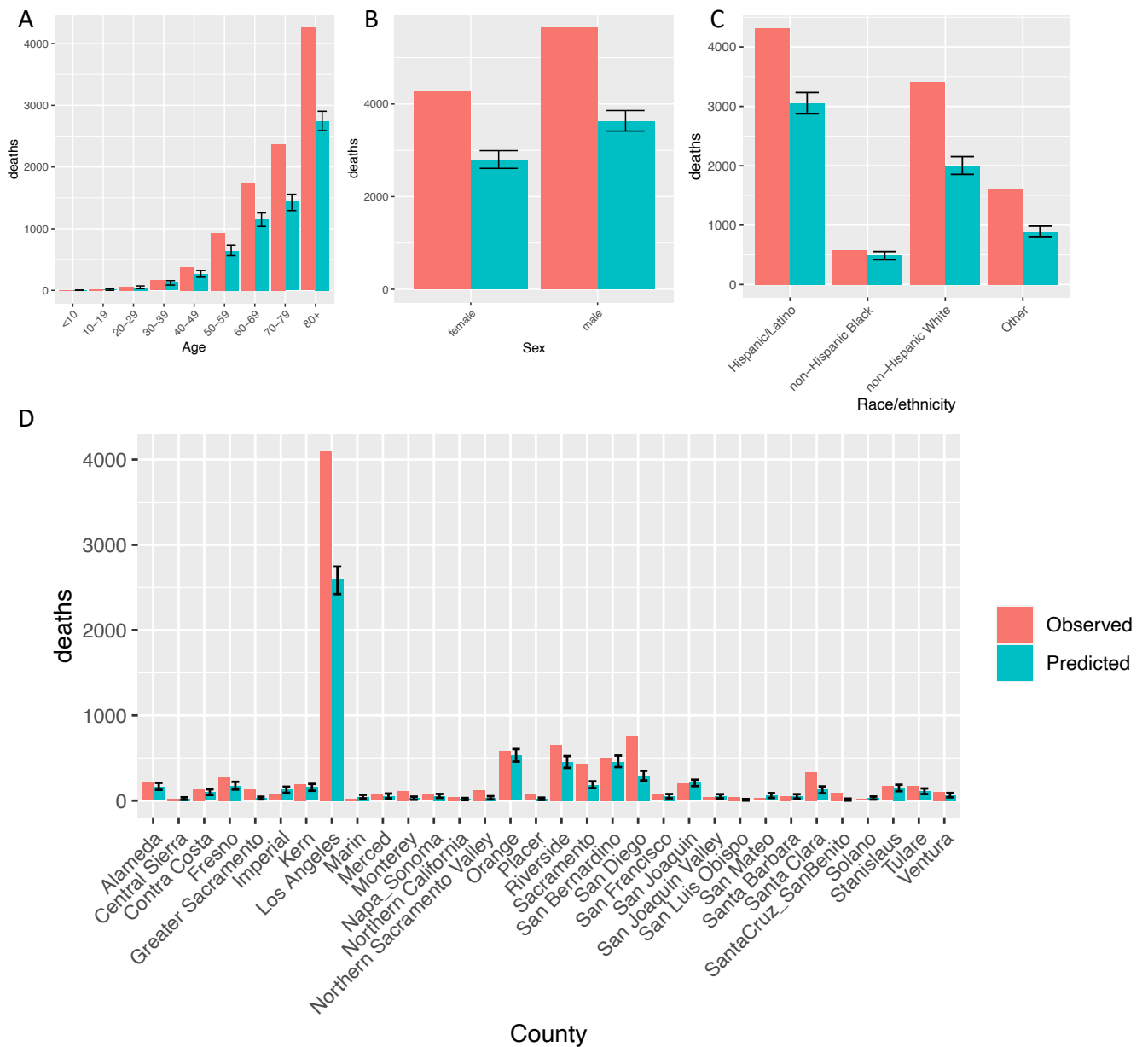


Figure S4. Model checking. Observed COVID-19 deaths in California Department of Public Health data and predicted COVID-19 deaths from the model for September 30 - December 30, 2020, by (A) age, (B) sex, (C) race/ethnicity and (D) county of residence. Model was fitted to data on 17,671 COVID-19 deaths from February 5 - September 30, 2020, and then simulated 1000 times to predict deaths from October 1 - December 30, 2020. Bars for predicted deaths represent mean of 1000 simulations, and error bars show 95% uncertainty intervals of the model predictions.

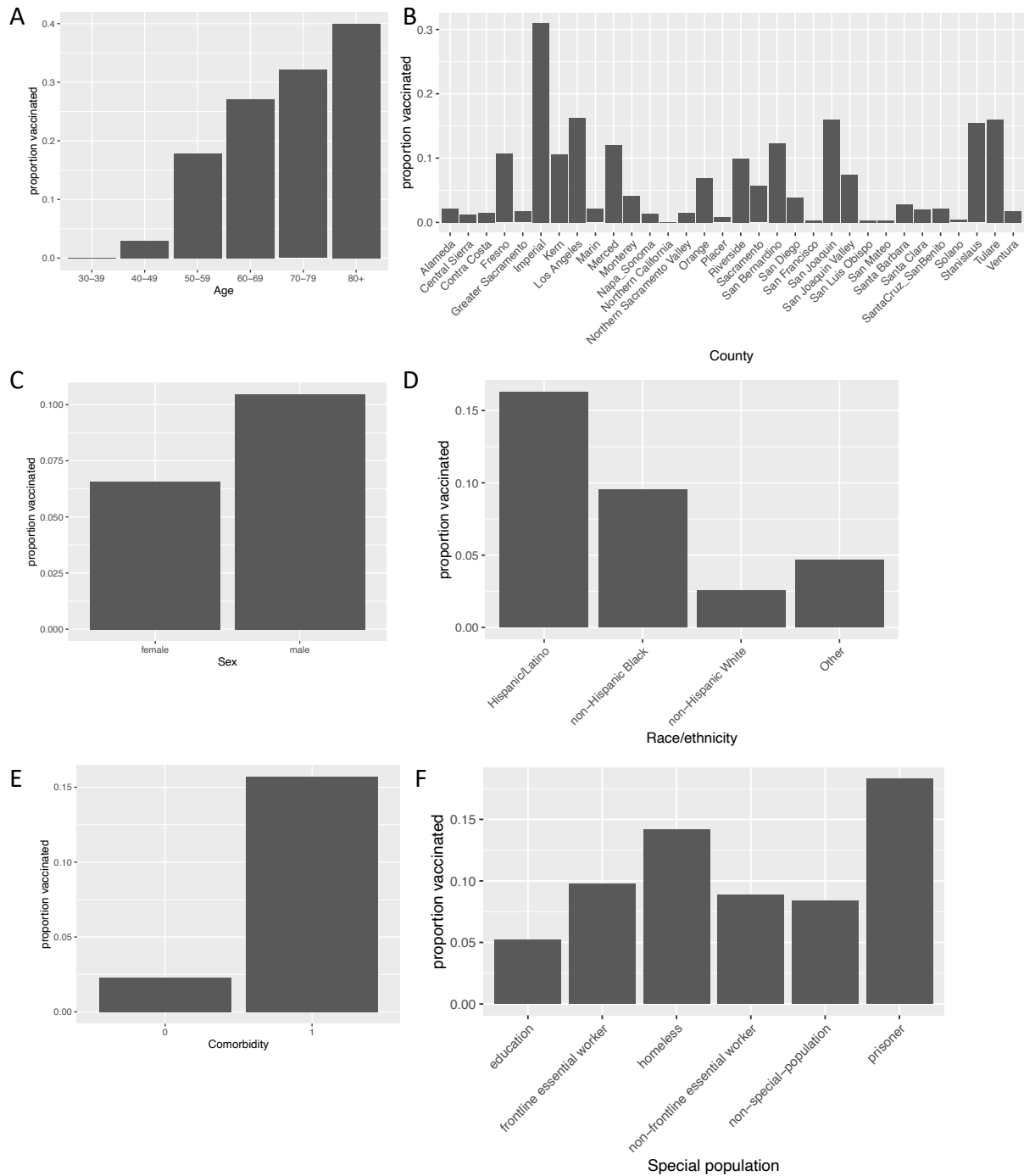


Figure S5. Optimal vaccine allocation under simultaneous targeting by multiple risk factors. Plots show proportion of individuals of each (A) age group, (B) county, (C) sex, (D) race/ethnicity, (E) comorbidity status, and (F) special population vaccinated under optimal distribution of the first 5 million vaccinations in California to avert DALYs. Comorbidity status treated as binary (0 = no comorbidities, 1 = any comorbidities).

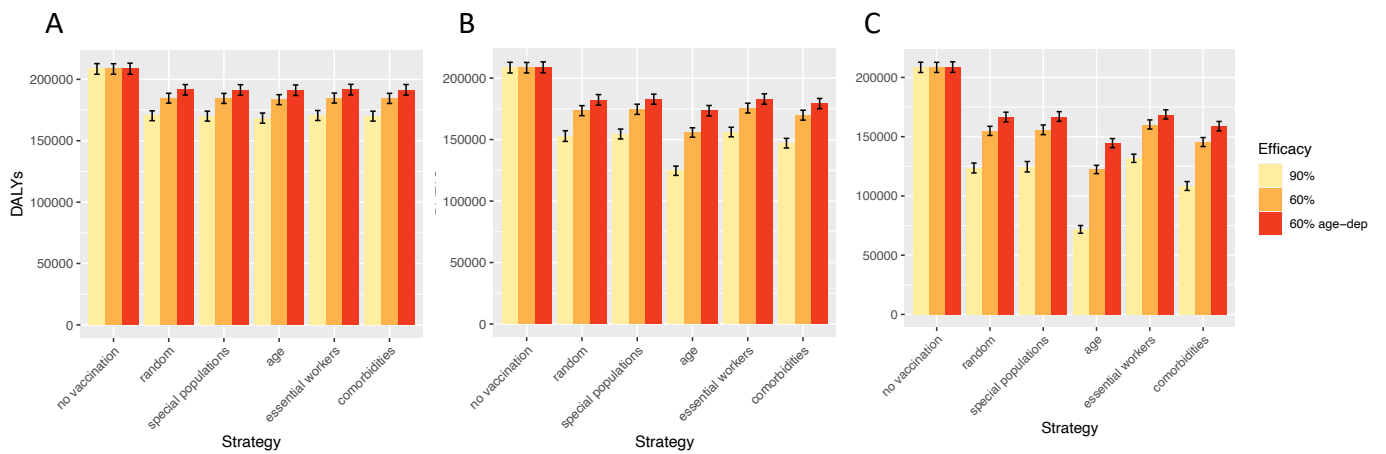


Figure S6. Sensitivity of impact of COVID-19 vaccine prioritization strategies to vaccine efficacy profile. Simulated impact of different prioritization strategies on total DALYs in California over 6 months for different vaccine efficacy profiles (95% age-stable, 60% age-stable, and 60% in <60-year-olds with a 10% decrease per decade for ≥ 60 -year-olds) for (A) 2 million, (B) 5 million and (C) 10 million vaccinations. Black lines show 95% uncertainty intervals from stochastic uncertainty and parameter uncertainty.

Table S8. Infection fatality rate estimates from (9) recalibrated to California

| Age group (years) | Infection fatality rate (%) | |
|--------------------------|------------------------------------|-------------|
| | Female | Male |
| <10 | 0.005 | 0.005 |
| 10-19 | 0.004 | 0.005 |
| 20-29 | 0.01 | 0.03 |
| 30-39 | 0.04 | 0.12 |
| 40-49 | 0.15 | 0.36 |
| 50-59 | 0.42 | 0.98 |
| 60-69 | 1.34 | 2.70 |
| 70-79 | 4.21 | 8.67 |
| 80+ | 15.2 | 28.6 |

Median ensemble infection fatality rate estimates from (9) recalibrated for California to match cumulative number of observed cases in California Department of Public Health data (see *Calibration of infection fatality rate for California*).

References

1. US Census Bureau, American Community Survey. *2018 Am. Community Surv. 5-year Estim.* (2018) Available at: <https://www.census.gov/programs-surveys/acs/data.html> [Accessed January 19, 2021].
2. U.S. Bureau of Labor Statistics, Employed persons by detailed occupation and age (2019) Available at: <https://www.bls.gov/cps/cpsaat11b.htm> [Accessed January 19, 2021].
3. California Assisted Living Association, RCFEs by the Numbers (2020) Available at: <http://caassistedliving.org/about-assisted-living/assisted-living-in-california/assisted-living-by-numbers/> [Accessed January 21, 2021].
4. National Center for Health Statistics, "Long-Term Care Providers and Services Users in the United States--State Estimates Supplement: National Study of Long-Term Care Providers, 2015-2016."
5. U.S. Department of Housing and Urban Development (HUD), 2019 Point in Time Estimates of Homelessness in the U.S. (2019) Available at: <https://www.hud.gov/2019-point-in-time-estimates-of-homelessness-in-US> [Accessed January 19, 2021].
6. Los Angeles Almanac, Homeless in Los Angeles County, California. *Los Angeles Alm.* (2020) Available at: <http://www.laalmanac.com/social/so14.php> [Accessed January 21, 2021].
7. K. Dooling, Phased Allocation of COVID-19 Vaccines ACIP COVID-19 Vaccines Work Group (2020) Available at: <https://www.cdc.gov/vaccines/acip/meetings/downloads/slides-2020-12/slides-12-20/02-COVID-Dooling.pdf> [Accessed January 19, 2021].
8. UCLA Center for Health Policy Research, California Health Interview Survey Available at: <https://healthpolicy.ucla.edu/chis/about/Pages/about.aspx> [Accessed January 19, 2021].
9. M. O'Driscoll, *et al.*, Age-specific mortality and immunity patterns of SARS-CoV-2. *Nature* (2020).
10. Centers for Disease Control and Prevention, Nationwide Commercial Laboratory Seroprevalence Survey (2021) Available at: <https://data.cdc.gov/Laboratory-Surveillance/Nationwide-Commercial-Laboratory-Seroprevalence-Su/d2tw-32xv> [Accessed January 4, 2021].
11. N. G. Davies, *et al.*, Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat. Med.* **26** (2020).
12. S. A. Lauer, *et al.*, The incubation period of coronavirus disease 2019 (CoVID-19) from publicly reported confirmed cases: Estimation and application. *Ann. Intern. Med.* **172**, 577–582 (2020).
13. K. L. Bajema, *et al.*, Estimated SARS-CoV-2 Seroprevalence in the US as of September 2020. *JAMA Intern. Med.* (2020).
14. N. Laird, D. Olivier, Covariance analysis of censored survival data using log-linear analysis techniques. *J. Am. Stat. Assoc.* **76**, 231–240 (1981).
15. M. Aitkin, D. Clayton, The Fitting of Exponential, Weibull and Extreme Value Distributions to Complex Censored Survival Data Using GLIM. *Appl. Stat.* **29**, 156 (1980).
16. R Core Team, R: A Language and Environment for Statistical Computing (2020).

17. Y.-H. Chen, *et al.*, Excess mortality associated with the COVID-19 pandemic among Californians 18–65 years of age, by occupational sector and occupation: March through October 2020. *medRxiv* (2020).
18. O. J. Hardy, *et al.*, A world apart: Levels and factors of excess mortality due to COVID-19 in care homes. The case of Wallonia - Belgium. *medRxiv*, 1–23 (2020).
19. A. Wajnberg, *et al.*, Robust neutralizing antibodies to SARS-CoV-2 infection persist for months. *Science* (80-.). **7728**, eabd7728 (2020).
20. J. M. Dan, *et al.*, Immunological memory to SARS-CoV-2 assessed for greater than six months after infection (2020).
21. C. Gaebler, *et al.*, *Evolution of antibody immunity to SARS-CoV-2* (2021).
22. L. H. Nguyen, *et al.*, Risk of COVID-19 among front-line health-care workers and the general community: a prospective cohort study. *Lancet Public Heal.* **5**, e475–e483 (2020).
23. B. Saloner, K. Parish, J. A. Ward, G. DiLaura, S. Dolovich, COVID-19 Cases and Deaths in Federal and State Prisons Novel. *JAMA* **324**, 602–603 (2020).
24. King County Department of Public Health, Homelessness and COVID-19 (2020) Available at: <https://www.kingcounty.gov/depts/health/covid-19/data/homeless.aspx> [Accessed August 21, 2020].
25. L. A. C. Chapman, *et al.*, Comparison of infection control strategies to reduce COVID-19 outbreaks in homeless shelters in the United States: a simulation study. *medRxiv*, 2020.09.28.20203166 (2020).
26. M. Mutambudzi, *et al.*, Occupation and risk of severe COVID-19: prospective cohort study of 120,075 UK Biobank participants Miriam. *medRxiv*, 1–29 (2020).
27. A. B. Docherty, *et al.*, Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: Prospective observational cohort study. *BMJ* **369**, 1–12 (2020).
28. C. M. Petrilli, *et al.*, Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: Prospective cohort study. *BMJ* **369** (2020).
29. E. J. Williamson, *et al.*, Factors associated with COVID-19-related death using OpenSAFELY. *Nature* **584**, 430–436 (2020).
30. E. Arias, J. Xu, United States Life Tables, 2017. *Natl. Vital Stat. Reports* **68**, 1–66 (2018).
31. J. A. Salomon, *et al.*, Common values in assessing health outcomes from disease and injury: Disability weights measurement study for the Global Burden of Disease Study 2010. *Lancet* **380**, 2129–2143 (2012).
32. J. B. O’Keefe, E. J. Tong, G. A. D. O’Keefe, D. C. Tong, Predictors of disease duration and symptom course of outpatients with acute covid-19: a retrospective cohort study. *medRxiv*, 1–31 (2020).
33. J. A. Lewnard, *et al.*, Incidence, clinical outcomes, and transmission dynamics of severe coronavirus disease 2019 in California and Washington: Prospective cohort study. *BMJ* **369** (2020).
34. D. Wang, *et al.*, Clinical Characteristics of 138 Hospitalized Patients with 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA - J. Am. Med. Assoc.* **323**, 1061–1069 (2020).
35. R. M. Anderson, C. Vegari, R. Baggaley, T. D. D. Hollingsworth, R. Maddren, The Royal Society SET-C Reports. Reproduction number (R) and growth rate (r) of the COVID-19

- epidemic in the UK: methods of estimation, data sources, causes of heterogeneity, and use as a guide in policy formulation [report unpublished]. *R. Soc.* (2020).
36. K. Dooling, *et al.*, The Advisory Committee on Immunization Practices' Ethical Principles for Allocating Initial Supplies of COVID-19 Vaccine—United States, 2020. *Morb. Mortal. Wkly. Rep.* **69**, 1857–1859 (2020).
 37. California Department of Public Health, COVID-19 Vaccine Doses Shipped (2021) Available at: <https://www.cdph.ca.gov/Programs/CID/DCDC/Pages/COVID-19/VaccineDoses.aspx> [Accessed January 17, 2021].
 38. UNICEF Supply Division, COVID-19 Vaccine Market Dashboard (2021) Available at: <https://www.unicef.org/supply/covid-19-vaccine-market-dashboard> [Accessed January 17, 2021].
 39. Duke Global Health Innovation Center, Launch and Scale Speedometer: COVID-19 (2021) Available at: <https://launchandscalefaster.org/COVID-19> [Accessed January 17, 2021].