Journal of Mathematical Biology (2023) 87:24
https://doi.org/10.1007/s00285-023-01948-y

Mathematical Biology

Analysis of metapopulation models of the transmission of SARS-CoV-2 in the United States

MyVan Vo[1] · Zhilan Feng[1,2] · John W. Glasser[3] · Kristie E. N. Clarke[4] · Jefferson N. Jones[3]

Received: 5 October 2022 / Revised: 18 April 2023 / Accepted: 8 June 2023 /
Published online: 8 July 2023

Abstract

During the COVID-19 pandemic, renewal equation estimates of time-varying effective reproduction numbers were useful to policymakers in evaluating the need for and impact of mitigation measures. Our objective here is to illustrate the utility of mechanistic expressions for the basic and effective (or intrinsic and realized) reproduction numbers, $\mathcal{R}_0$, $\mathcal{R}_E$, and related quantities derived from a Susceptible-Exposed-Infectious-Removed (SEIR) model including features of COVID-19 that might affect transmission of SARS-CoV-2, including asymptomatic, pre-symptomatic, and symptomatic infections, with which people may be hospitalized. Expressions from homogeneous host population models can be analyzed to determine the effort needed to reduce $\mathcal{R}_E$ from $\mathcal{R}_0$ to 1 and contributions of modeled mitigation measures. Our model is stratified by age, 0–4, 5–9, …, 75+ years, and location, the 50 states plus District of Columbia. Expressions from such heterogeneous host population models include subpopulation reproduction numbers, contributions from the above-mentioned infectious states, metapopulation numbers, subpopulation contributions, and equilibrium prevalence. While the population-immunity at which $\mathcal{R}_E = 1$ has captured the popular imagination, the metapopulation $\mathcal{R}_E \leq 1$ could be attained in an infinite number of ways even if only one intervention (e.g., vaccination) were capable of reducing $\mathcal{R}_E$. However, gradients of expressions derived from heterogeneous host population models, $\nabla\mathcal{R}_E$, can be evaluated to identify optimal allocations of limited resources among subpopulations. We illustrate the utility of such analytical results by simulating

✉ John W. Glasser
jglasser@cdc.gov

1 Department of Mathematics, Purdue University, West Lafayette, USA

2 Division of Mathematical Sciences, NSF, Alexandria, USA

3 Coronavirus and Other Respiratory Viruses Division, National Center for Immunization and Respiratory Diseases, CDC, 1600 Clifton Road NE, Atlanta, GA 30333, USA

4 Center for Surveillance, Epidemiology, and Laboratory Services, CDC, Atlanta, USA

two hypothetical vaccination strategies, one uniform and other indicated by $\nabla\mathcal{R}_E$, as well as the actual program estimated from one of the CDC's nationwide seroprevalence surveys conducted from mid-summer 2020 through the end of 2021.

Keywords SEIR metapopulation modeling · SARS-CoV-2 transmission · COVID-19 in the United States · Reproduction numbers and related expressions · Optimal vaccination strategies

Mathematics Subject Classification 34C60 · 92B05 · 92D25 · 92D30

1 Introduction

We have developed several metapopulation SEIR models including biological features of COVID-19 that might affect the transmission of SARS-CoV-2. Our objective here is to demonstrate the utility of reproduction numbers and analytical quantities derived from them. While derivation of expressions from mechanistic models typically involves equilibrium assumptions, dynamic analogues related to Hawkes processes in mathematics, together with statistical approximations of the requisite ingredients (Thompson et al. 2019), indicated the need for and impact of public health interventions during the COVID-19 pandemic.

We used information from the CDC's nationwide seroprevalence studies (Bajema et al. 2021; Jones et al. 2021) to parameterize, initialize, and evaluate an age- and location-stratified model of SARS-CoV-2 transmission in the United States (Fig. 1). Serological observations are less timely than conventional surveillance, but much more

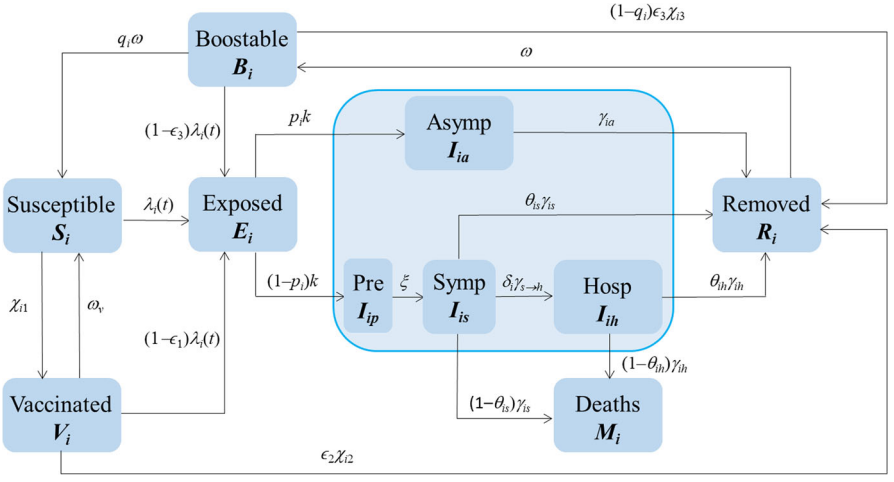

Fig. 1 Diagram of a modified SEIR model. This metapopulation model includes asymptomatic and presymptomatic stages, disease-induced mortality, vaccination, and the waning and boosting of immunity, features of COVID-19 that affect SARS-CoV-2 transmission. It is cross classified by age and location

accurate. Because we used parameter values with which our model reproduces serological observations reasonably well, our reproduction numbers and related analytical quantities are unusually reliable.

Setting the expression for $\mathcal{R}_E$ (average infections per infectious person) from our homogeneous host population model equal to 1 (below which new infections must decline) and solving for equilibrium immunity, we describe how immunization and non-pharmaceutical interventions (NPIs) each contribute to control. Evaluating the gradient of $\mathcal{R}_E$ derived from our age- and location-stratified host population model with respect to vaccination rates, $\nabla\mathcal{R}_E(\chi_i)$, we describe whom to vaccinate to reduce transmission most expeditiously. We compare that hypothetical strategy's impact with (a) immunization rates derived from seroprevalence among blood donors (to which we refer as the actual strategy) and (b) rates determined solely by the doses available and eligible recipients at each location, whereupon age groups would be vaccinated proportionately (i.e., at uniform rates).

Similarly, invasion reproduction numbers (which must exceed one for novel pathogen strains to become established when ancestral ones are at their endemic equilibria) from models in which hosts are stratified by the pathogens to which they have been exposed can help to explain the successive dominance of a few of the many variants to emerge during pandemics. As such models differ fundamentally from those whose hosts are stratified by age or location, we will describe models in which immunity from one strain confers complete or partial immunity to another and illustrate the utility of reproduction numbers derived from them elsewhere.

Because our models are suitable for the respiratory pathogens most likely to cause pandemics, their evaluation may increase the utility of metapopulation modeling in future. We hope that it will also increase our understanding of the transmission and control of those pathogens.

2 Methods

As our age- and location-stratified SARS-CoV-2 transmission model has been described in detail elsewhere (Glasser et al. 2023), we summarize insofar as possible here.

2.1 Model equations

Our models can be represented as systems of ordinary differential equations, where S_i is the number who are susceptible in group i (here i indexes ages, locations, or combinations; there are n such groups); B_i is the number whose immunity has waned and could be boosted by (re)infection or (re)vaccination; E_i is the number who have been infected, but are not yet infectious, V_i is the number who have received one dose of vaccine ($\geq$ 14 days ago), I_{iw} ($w = a,p,s,h$) are the numbers infectious in group i who are asymptomatic, pre-symptomatic, symptomatic, or hospitalized, R_i is the number who are temporarily immune due to recovery from infection, completion of the primary vaccination series or receipt of a booster dose ($\geq$ 14 days ago), and M_i is the number in group i who have died:

$$
\begin{aligned}
S_i' &= q_i\omega B_i + \omega_v V_i - [\lambda_i(t) + \chi_{1i}]S_i,\\
B_i' &= \omega R_i - [q_i\omega + (1-q_i)\varepsilon_3\chi_{3i} + (1-\varepsilon_3)\lambda_i(t)]B_i\\
E_i' &= \lambda_i(t)S_i + (1-\varepsilon_1)\lambda_i(t)V_i + (1-\varepsilon_3)\lambda_i(t)B_i - kE_i,\\
V_i' &= \chi_{1i}S_i - [(1-\varepsilon_1)\lambda_i(t) + \omega_v + \varepsilon_2\chi_{2i}]V_i,\\
I_{ia}' &= kp_iE_i - \gamma_{ia}I_{ia},\\
I_{ip}' &= k(1-p_i)E_i - \xi I_{ip},\\
I_{is}' &= \xi I_{ip} - [\gamma_{is} + \delta_i\gamma_{is\to h}]I_{is},\\
I_{ih}' &= \delta_i\gamma_{is\to h}I_{is} - \gamma_{ih}I_{ih},\\
R_i' &= \gamma_{ia}I_{ia} + \theta_{is}\gamma_{is}I_{is} + \theta_{ih}\gamma_{ih}I_{ih} + \varepsilon_2\chi_{2i}V_i + (1-q_i)\varepsilon_3\chi_{3i}B_i - \omega R_i,\\
M_i' &= (1-\theta_{is})\gamma_{is}I_{is} + (1-\theta_{ih})\gamma_{ih}I_{ih}, \quad i = 1, 2, \ldots, n,
\end{aligned}
$$

where the force or hazard rate of infection among susceptible people in group i (and those having received only one vaccine dose, whose susceptibility to infection is $1 - \epsilon_1$, where ϵ_1 is the efficacy of a single dose) is

$$
\lambda_i(t) = \psi_{iS}a_i\beta_i\sum_{j=1}^{n} c_{ij}\psi_{jI}\frac{\eta_{ja}I_{ja}(t) + \eta_{jp}I_{jp}(t) + I_{js}(t) + \eta_{jh}I_{jh}(t)}{N_j}, \text{ where}
$$

$$
N_j = S_j + B_j + E_j + V_j + I_{ja} + I_{jp} + I_{js} + I_{jh} + R_j + M_j \text{ and } N = \sum_{j=1}^{n} N_j.
$$

In this expression, $\psi_{iS} = 1 - c_ib_{iS}$ and $\psi_{iI} = 1 - c_ib_{iI}$ represent reductions in susceptibility and infectiousness, respectively, due to physical-distancing if possible and mask-wearing otherwise, c_i is the proportion of group i that comply with such recommendations and b_{iS} and b_{iI} are their respective efficacies. The parameter a_i is the *per capita* contact rate, β_i is the probability of transmission per contact between susceptible and infectious people ($a_i \times \beta_i$ is the effective contact rate), and c_{ij} is the proportion of their contacts that members of group i have with those of group j. As a convenience, we include M_j in N_j, so that N is constant.

2.2 Parameter values

We obtained values for most biological parameters, which in this model are either constant or vary only with age (Glasser et al. 2023, Tables 1 and 2), from the primary literature (notably Clark et al. 2020; Hale et al. 2021; He et al. 2020; Jones 2020; Levin et al. 2020; Prem et al. 2017) or subject-matter experts.

We calculated the probabilities of infection on contact, the β_i, from seroprevalence – the humoral (versus cell-mediated) component of the adaptive immune response – as a function of age and time from a study of residual sera from commercial laboratory testing in the United States from mid-2020 through 2021 for purposes other than SARS-CoV-2 infection assessment (Bajema et al. 2021).

Initializing our model from seroprevalence on 7 September 2020, we estimated the efficacies, $b_I = b_S$, required for simulated proportions immune, R_i, to approximate seroprevalence on 25 December 2020. Using those parameter values and vaccination rates derived from reports of weekly doses administered, we simulated through 4 April 2021 and compared R_i to seroprevalence from a study of blood-donors (Jones et al. 2021), which captures vaccination- as well as infection-induced antibodies.

We use a single subscript when possible, but our model of SARS-CoV-2 transmission in the United States, detailed by Glasser et al. (2023) and analyzed here, is stratified by age and location. In the next two sections, we summarize information described in greater detail by those authors.

2.3 Mixing

We multiply the age-specific US mixing matrix of Prem et al. (2017), denoted $c_{a_j a_q}$ below, by ratios of the natural logs of location-specific and overall population densities to obtain $c^{(p)}_{a_j a_q} = c_{a_j a_q} \times \frac{\ln(\rho^{(p)})}{\ln(\rho)},\ 1 \le i,\ p \le m;\ 1 \le j,\ q \le n.$

Consequently, marginal contacts within more (less) densely populated locations are greater (less) than the US average, but most differ by a factor of less than two (Hethcote and van Ark 1987). Between locations, contacts decline exponentially with distance at age-specific rates, b_{a_q} whose calculation Feng et al. (2017) describe. In this model, distances are "as the crow flies" between population centers. Thus, mixing matrix elements are

$$c_{l_i a_j l_p a_q} = \frac{c^{(p)}_{a_j a_q} e^{-b_{a_q} d_{l_i l_p}}}{\sum_{r=1}^{n} \sum_{s=1}^{m} c^{(s)}_{a_j a_r} e^{-b_{a_r} d_{l_i l_s}}}.$$

In these expressions, ρ is population density, the quotient of population and land (i.e., habitable) area, with the indices i and p denoting locations, of which there are $m = 51$ (the 50 states plus District of Columbia), and j and q denoting age groups, of which there are $n = 16$ (0–4, 5–9, …, 75 + years). Thus, the m mixing matrices $c^{(p)}_{a_j a_q}$ and one matrix $c_{l_i a_j l_p a_q}$ have $n^2 = 256$ and $(m \times n)^2 = 665{,}856$ elements, respectively. We retain this compound notation only when describing our mixing functions.

2.4 Reproduction numbers

We derive expressions for the metapopulation reproduction numbers for single and multiple strata (e.g., age or location alone and together) via the next-generation matrix approach of van den Driessche and Watmough (2002) in the appendix. These have the same form as those from other SEIR models without demographic dynamics, but are complicated by our inclusion of vaccination and expansion of the infectious compartment.

2.5 Optimal control

In metapopulation models, we wish to know how to allocate resources (i.e., to which subpopulation or combination of subpopulations) for most expeditious control (i.e., $\mathcal{R}_E \leq 1$). Negative gradients of effective reproduction numbers, $-\nabla\mathcal{R}_E$ with respect to parameters of functions by which mitigation measures are modeled describe how to reduce the metapopulation $\mathcal{R}_E$ most expeditiously (Feng et al. 2015, 2017, 2019).

When there are more than a few groups, explicit expressions for reproduction numbers cannot be derived unless mixing is proportional, when they become traces of the above-mentioned next-generation matrices, explicit expressions whose partial derivatives can be calculated with respect to functions by which mitigation measures are modeled. Thus, gradient calculations require the assumption that mixing is proportional,

$$c^{(p)}_{a_j a_q} = \frac{a_{a_j} N_{l_p a_j}}{\sum_r \sum_k a_{a_k} N_{l_r a_k}} \times \frac{\ln(\rho^{(p)})}{\ln(\rho)}, \quad 1 \leq r,\ p \leq m; 1 \leq k,\ q \leq n,$$

where $a_{a_j} = \sum_{a_q} C_{a_j a_q}$ are marginal contact rates (e.g., from Prem et al. 2017) and $N_{l_p a_j}$ are subpopulation sizes. If mixing is proportional to the n age group contributions to the total number of contacts at each of the p locations, the metapopulation reproduction numbers are traces of the matrix H (appendix), whose multivariate partial derivatives we calculate with respect to the vaccination rates, $\chi_{1(i,j)}$, $\chi_{2(i,j)}$, or $\chi_{3(i,j)}$.

2.6 Non-pharmaceutical interventions

During the COVID-19 pandemic, very effective vaccines and monoclonal antibodies were developed and evaluated with astonishing speed by historical standards. Simultaneously, existing anti-viral medications were screened for efficacy against SARS-CoV-2. Nonetheless, only non-pharmaceutical interventions (e.g., physical-distancing, closing non-essential businesses, limiting gathering sizes, ventilating indoor spaces, and requiring mask-wearing) were available for most of 2020. Because state and local governments imposed and relaxed various combinations independently, modeling particular NPIs throughout the United States was impractical.

Instead, we multiplied age-specific contact rates within jurisdictions by time-varying complements of Oxford Stringency Indices (Hale et al. 2021) expressed as proportions. For compliance, we scored responses to "How often have you worn a face mask outside your home (e.g., when on public transport, going to a supermarket, going to a main road)?" by $n = 33{,}940$ US YouGov survey participants aged 18 to 99 years queried from March to December of 2020 (Jones 2020). Our parameter values are age-group averages from logistic regressions of scored responses as a cubic function of age (Glasser et al. 2023).

2.7 Vaccination

In the United States, vaccination began on 14 December 2020 with healthcare and other essential workers, followed by those at risk of serious illness, hospitalization, or death (i.e., older adults and otherwise immunocompromised people), and then progressively younger and healthier people (Fig. 2a).

For simulations reported by Glasser et al. (2023), we determined age- and location-specific numbers of people eligible for vaccination (i.e., susceptible and vaccinated with one dose) at the beginning of each week by simulation, calculated the proportions vaccinated from the reported first and second doses administered that week, simulated with the corresponding rates, $\chi_{1(i,j)}$ and $\chi_{2(i,j)}$, and repeated the next week.

In the simulations reported here, susceptible people move directly to the removed (temporarily immune) class at weekly immunization rates calculated from a nationwide survey of seroprevalence among blood donors (Jones et al., 2021) or the alternatives described below. We calculate the observed immunization rates the same way that Glasser et al. (2023) calculated the infection rates.

As those authors describe in much greater detail, we fit bivariate logistic regressions to proportions with anti-nucleocapsid (anti-N) and anti-spike (anti-S) antibodies, which indicate previous infection (and possibly vaccination) and to those with anti-S antibodies alone, which indicate vaccination (without a history of infection), as a function of age and time, both modeled as cubic polynomials and, because those variables are related biologically (i.e., subjects aged a at time t are aged $a + 1$ at time $t + 1$), all interactions (Fig. 2b).

Denoting the proportion susceptible among blood donors aged $a_{j\text{-}1} \leq a < a_j$ at time $t_{k\text{-}1} \leq t < t_k$ as $S\ (j,\ k) = 1 - I_V\ (j,\ k)$, where I_v is the proportion with anti-S (i.e., immune by virtue of vaccination), and denoting the proportion immune at time k as a function of that at time $k-1$ as $S(j,\ k) = S(j,\ k-1)exp\{-[\chi(j,\ k)]\}$, we obtain the desired result, $\chi(j,\ k) = -ln\{[1 - I_V(j,\ k)]/[1 - I_V(j,\ k-1)]\}$ by rearrangement and substitution.

We calculated such vaccination rates from surfaces like that for the entire United States (Fig. 2b) for each jurisdiction i, so could write $\chi(i,\ j,\ k)$. In the next section,

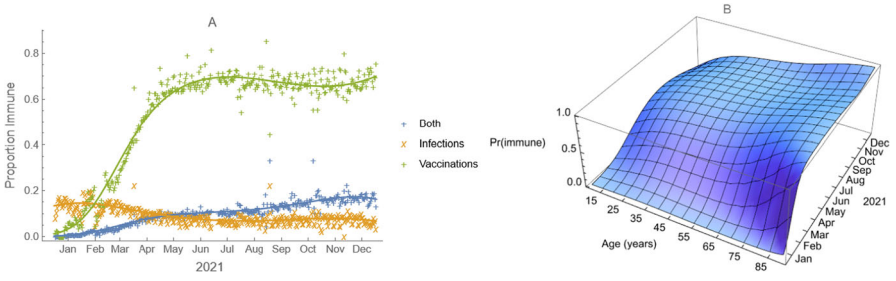

Fig. 2 Proportions immune by virtue of **a**) infection, vaccination and both and **b**) vaccination by age and time during 2021 in the United States. These data are from a subset of blood-donors who were queried about their vaccination status (Busch et al. 2022)

we switch to continuous time, writing $\chi_{(i,j)}(t)$, where i denotes location, j age, and t time in days (since 31 December 2019).

2.8 Alternative strategies

As mentioned, the strategy enacted in the United States initially prioritized vaccination of those whose risk of exposure to SARS-CoV-2 or severe disease if exposed was greatest. Younger and healthier members of the general population were vaccinated subsequently. While vaccinating these high-risk groups was meant to result in protective immunity, some older adults or those with immunocompromising or certain other health conditions might not develop sufficient immunity for protection from severe disease.

As fewer such people might be infected if those who otherwise might infect them were vaccinated instead, we compare observed immunization rates with two alternative strategies: In the first, eligible people at each location were vaccinated at rates determined solely by the number of doses administered. Were our simulations stochastic (Renshaw 1991), we could term this strategy random. The second alternative reduces the average number of secondary infections per primary most expeditiously given available vaccine.

2.9 Experimental conditions

Among the capabilities of transmission models is investigating alternative scenarios, typically in public health interventions to mitigate undesirable outcomes (e.g., infections, hospitalizations, deaths). Insofar as these models reproduce observations under actual conditions, evaluations of actual measures or hypothetical ones being contemplated are reliable. The essence of any experiment is that only one factor is varied at a time to deduce its impact conditional on others. In our experiments, all else equal amounts to NPIs and available doses being the same at each time in each location.

To facilitate identifying the strategy that would minimize $\nabla\mathcal{R}_E$, we reduced the original 16 age groups to the six aged 0–4, 5–19, 20–39, 40–59, 60–74, and 75 + years, which correspond to pre-school and schoolchildren, reproductive-, middle-aged, older, and elderly adults.

We let the first week begin on 13 December 2020 ($w = 1$ that week; https://www.medscape.com/viewarticle/976734#) and consider the doses administered during that and successive weeks at each location. We calculate the number of doses administered as $\sum_{j=1}^{6}\int_{t_{w0}}^{t_w}\chi_{(i,j)}(t)S_{i,j}(t)dt = d_w(i)$ for $i = 1, \ldots, m$, where t_{w0} is the beginning of any given week and t_w is its end. We denote the set of all $\chi_{(i,j,w)}$ values as $\chi_w = \{\chi_{(i,j,w)}\}$ for $i = 1, \ldots, m;\ j = 1, \ldots, 6$, and the set of $\chi_{(i,j,w)}$ values for a specific location i for all age groups as $\chi_{i,w} = (\chi_{(i,1,w)}, \ldots, \chi_{(i,6,w)})$. Finally, we let $\mathcal{D} := \{d_w(i)\}$ be the set of weekly doses used to calculate the location- and age-dependent rates, $\chi_{(i,j,w)}$, where the index i denotes locations, j age groups, and w weeks, to ensure that the $d_w(i) = \sum_{j=1}^{6}\chi_{(i,j,w)}S_{i,j}(t_{w0})$ are the same in these alternatives as in the actual strategy:

1) Uniform rates: the doses available to people at each location during each week are allocated in proportion to the numbers susceptible in each age group. In other words, for a given week w at location i, the immunization rates are the same over all age groups, $k_i = \chi_{(i,j,w)}$ and $\sum_{j=1}^{6} \chi_{(i,j,w)} S_{i,j}(t_{w0}) = k_i \sum_{j=1}^{6} S_{i,j}(t_{w0}) = d_w(i)$.
2) Gradient rates: rates that minimize the metapopulation $\mathcal{R}_E$ accounting for the limited number of doses available. First we note that, because the doses of vaccine available at each location are independent, but the vaccination rates are location-dependent, minimizing $\mathcal{R}_E(\chi_{i,w})$ for each location is equivalent to minimizing $\mathcal{R}_E(\chi_w)$. Then we calculate $\nabla \mathcal{R}_E(\chi_{i,w})$ and use it to solve simultaneously the equations

$\nabla \mathcal{R}_E(\chi_{i,w}) + \lambda\left[S_{i,1}(t_{w0}), \quad S_{i,2}(t_{w0}), \quad \ldots, \quad S_{i,6}(t_{w0})\right] = 0$ and $\sum_{j=1}^{6} \chi_{(i,j,w)} S_{i,j}(t_{w0}) = d_w(i)$, where λ is a Lagrange multiplier (Feng et al. 2015, 2017, 2019), for each location i.

To describe a practical matter, we let $\hat{\chi}_{i,w} = (\hat{\chi}_{(i,1,w)}, \hat{\chi}_{(i,2,w)}, \ldots, \hat{\chi}_{(i,6,w)})$ denote the optimal solution, $\mathcal{R}_E(\hat{\chi}_{i,w}) = \mathcal{R}_{E\,\min}$. When dose availability suffices, a solution exists such that $\hat{\chi}_{i,w} > 0$. When too few doses are available, there does not exist a solution satisfying the equations above with $\hat{\chi}_{i,w} > 0$. In those instances, we checked all boundary conditions where a subset of the $\chi_{i,w}$ were zero. Then, for the set of possible solutions such that $\chi_{i,w} \geq 0$, we evaluated $\mathcal{R}_E(\chi_{i,w})$ and determined which $\chi_{i,w}$ minimized $\mathcal{R}_E(\chi_{i,w})$.

3 Analytical results

3.1 Reproduction numbers

The effective number for a single group i is $\mathcal{R}_{E(i)} = \Psi_{iS} a_i \beta_i [s_i^* + (1-\varepsilon_1) v_i^* + (1-\varepsilon_3) b_i^*] \Psi_{iI} Q_i$, with $s_i^* = \left(1 + \frac{\chi_{1i}}{\omega_v + \varepsilon_2 \chi_{2i}} + \frac{\varepsilon_2 \chi_{1i} \chi_{2i}}{q_j \omega (\omega_v + \varepsilon_2 \chi_{2i})} + \frac{q_j \omega \varepsilon_2 \chi_{2i} \chi_{1i} + (1-q) \varepsilon_3 \chi_{3i} \varepsilon_2 \chi_{1i} \chi_{2i}}{q_j \omega^2 (\omega_v + \varepsilon_2 \chi_{2i})}\right)^{-1}$, $v_i^* = \left(\frac{\chi_{1i}}{\omega_v + \varepsilon_2 \chi_{2i}}\right) s_i^*$, $b_i^* = \left(\frac{\varepsilon_2 \chi_{1i} \chi_{2i}}{q_i \omega (\omega_v + \varepsilon_2 \chi_{2i})}\right) s_i^*$, and $Q_i = \eta_{ia} P_{ia} T_{ia} + \eta_{ip} P_{ip} T_{ip} + P_{ip} T_{is} + \eta_{ih} P_{ip} P_{ih} T_{ih}$.

The expressions for the starred variables are equilibrium proportions susceptible, s_i^* vaccinated with one dose, v_i^* and partially immune, b_i^*. In the expression for s_i^*, the four terms correspond to scaling constants for those susceptible, vaccinated, boosted, and recovered, respectively, in relation to s_i^*. In the expression for Q_i, $P_{ia} = p_i$, $P_{ip} = (1 - p_i)$, and $P_{ih} = \frac{\delta_i \gamma_{is \to h}}{\gamma_{is} + \delta_i \gamma_{is \to h}}$ denote probabilities of transition from E_i to I_{ia}, from E_i to I_{ip}, and from I_{is} to I_{ih}, respectively, while $T_{ia} = \frac{1}{\gamma_{ia}}$, $T_{ip} = \frac{1}{\xi_i}$, $T_{is} = \frac{1}{\gamma_{is} + \delta_i \gamma_{is \to h}}$, and $T_{ih} = \frac{1}{\gamma_{ih}}$ denote mean residence times or sojourns in the I_{ia}, I_{ip}, I_{is} and I_{ih} stages, respectively. The basic reproduction number, $\mathcal{R}_{0(i)}$ is obtained by setting the control parameters (i.e., c, δ, χ) to zero.

The metapopulation numbers are derived in the appendix. Our estimate of $\mathcal{R}_0$ during fall of 2020, when the ancestral variant was circulating, is 5.2; $\mathcal{R}_E$ varies with the Oxford Stringency Indices (i.e., NPIs), but we estimate that it was 1.3 on 25

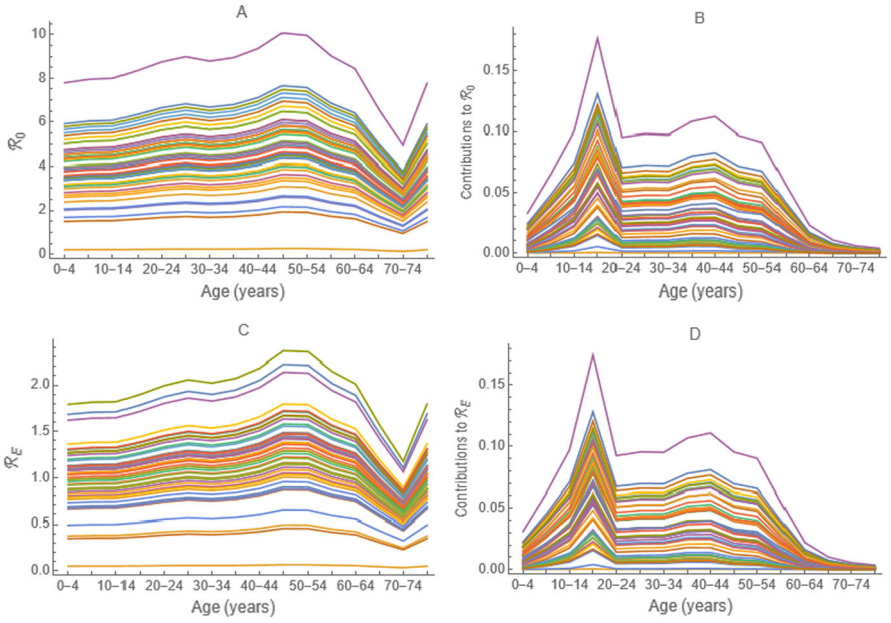

Fig. 3 Subpopulation and contributions to the metapopulation $\mathcal{R}_0$(**a** and **b**), which we estimate to have been 5.2 for the ancestral variant circulating during the fall of 2020, and subpopulation and contributions to the metapopulation $\mathcal{R}_E$(**c** and **d**), which we estimate to have been 1.3 on 25 December 2020. The legends for Figs. 3, 4, 5 are the same as Fig. 7, but Figs. 3, 4, 5 are repeated by HHS Region in the Supplement, where jurisdictions are more easily distinguished

December 2020. It also varies with vaccination, which however had only just begun, so was omitted in that calculation and figures described in the next paragraph.

The subpopulation reproduction numbers and contributions to the metapopulation numbers (left eigenvectors associated with the above-mentioned dominant eigenvalues, Caswell 2001) are juxtaposed in Fig. 3, components of the subpopulation effective numbers, asymptomatic, pre-symptomatic, symptomatic, and hospitalized infections (i.e., $\mathcal{R}_{E(i)} = \mathcal{R}_{A(i)} + \mathcal{R}_{P(i)} + \mathcal{R}_{S(i)} + \mathcal{R}_{H(i)}$) in Fig. 4, and the equilibrium prevalence associated with those metapopulation numbers (right eigenvectors) in Fig. 5.

Because the 51 jurisdictions are impossible to distinguish, Figs. 3, 4, 5 are repeated – with vaccination on 25 December 2020 and 4 April 2021 – by US Department of Health and Human Services (HHS) Region in the on-line Supplement, which includes a discussion of spatial patterns.

3.2 Population-immunity threshold

The population-immunity threshold is the proportion immune at which $\mathcal{R}_{E(i)} = 1$. If $n = 1$ (i.e., the host population is homogeneous), $\mathcal{R}_{E(i)} = \mathcal{R}_E$. Setting $\mathcal{R}_E = 1$, substituting $\mathcal{R}_0$ for $a\beta Q$, and solving for s^*, the equilibrium proportion susceptible, $s^* = \frac{1}{\Psi_S \Psi_I \mathcal{R}_0}\left[1 + (1-\varepsilon_1)\left(\frac{\chi_1}{\omega_v + \varepsilon_2 \chi_2}\right) + (1-\varepsilon_3)\left(\frac{\varepsilon_2 \chi_1 \chi_2}{q\omega(\omega_v + \varepsilon_2 \chi_2)}\right)\right]^{-1}$.

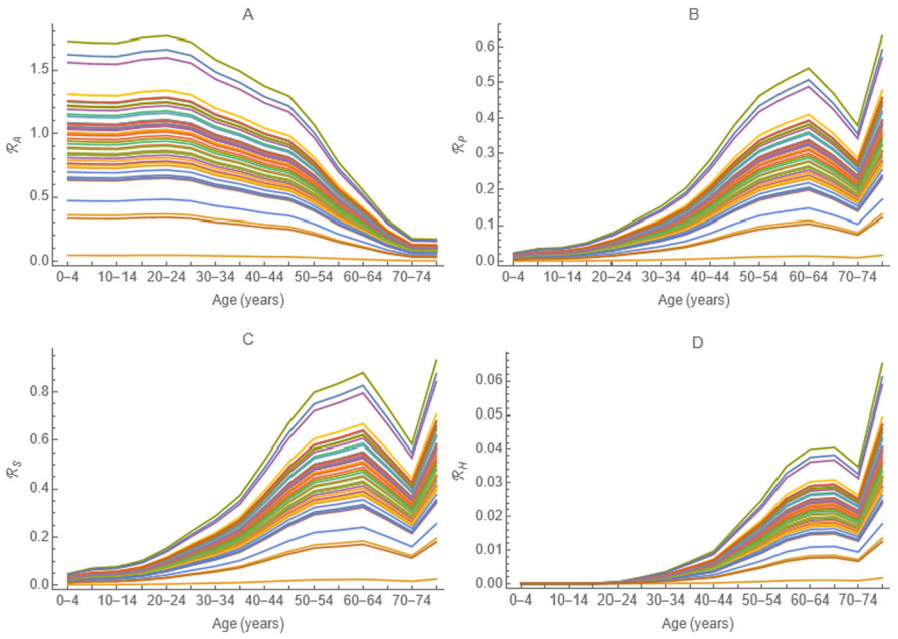

Fig. 4 Infectious state contributions to the $m \times n$ age- and location-specific effective reproduction numbers $\mathcal{R}_{E(i)} = \mathcal{R}_{A(i)} + \mathcal{R}_{P(i)} + \mathcal{R}_{S(i)} + \mathcal{R}_{H(i)}$, where $i = 1, \ldots, 816$, **a**) asymptomatic, **b**) pre-symptomatic, **c**) symptomatic, and **d**) hospitalized (for legend, see Fig. 7b). On 25 December 2020, Florida, New Jersey, and Washington, DC, dominated

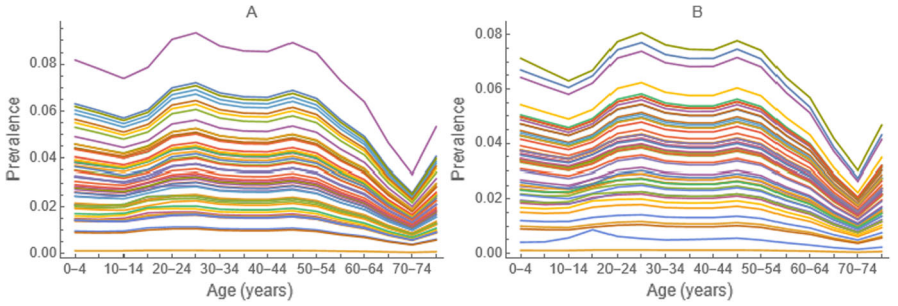

Fig. 5 Prevalence of infection implied by the metapopulation **a**) basic and **b**) effective reproduction numbers. Had interventions remained at their levels on 25 December 2020, the prevalence of infection would have been greater in Florida and New Jersey than the District of Columbia

The population-immunity threshold is $1 - s^*$, which has the same form, $1 - 1/\mathcal{R}_0$, as that from simpler models, but includes terms for vaccination, the waning of immunity, and other mitigation measures that may contribute (Fig. 6). Our estimate varies with $\mathcal{R}_E$, but – again ignoring vaccination – was 0.66 on 25 December 2020.

We do not derive such expressions for our stratified models because an infinite number of combinations of subpopulation immunity would attain $\mathcal{R}_E = 1$ (Feng

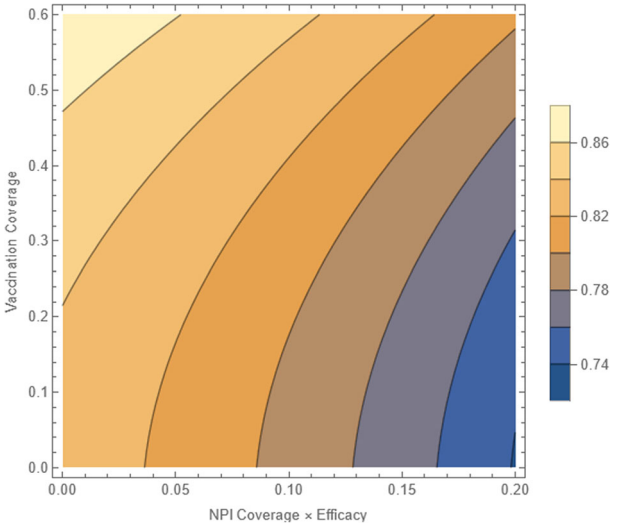

Fig. 6 Contribution of NPIs to the population-immunity threshold. Using our homogeneous population model, whose parameters are US population weighted averages of the age-specific ones in our heterogeneous population models (Glasser et al. 2023, Table 2), we estimate that this threshold would have been 0.83 without NPIs, the lower left corner of this figure, but was 0.66 with them on 25 Dec 2020. Contours curve towards increasing values of NPIs, represented as products of coverage and efficacy

et al. 2015), pairs if there were two subpopulations, triples if there were three, and so on. Thus, the policy-relevant question becomes, "Which is optimal?".

3.3 The gradient

In Fig. 7, we illustrate the magnitude and direction of the gradient of $\mathcal{R}_E$ with respect to $\chi_{(i,j,w)}$ shortly before vaccination began (when $\chi_{(i,j,w)} = 0, i = 1, \ldots, m; j = 1, \ldots, n$). And, in Fig. 8, we illustrate those components during successive weeks of January 2021 (when $\chi_{(i,j,w)} \geq 0$, $i = 1, \ldots, m; j = 1, \ldots, n$) in California, the jurisdiction in which the gradient magnitude is greatest (Fig. 8a), and Washington, DC, the jurisdiction whose age-distribution differs most from the others (Fig. 8b). Utah's age distribution is younger than most, but the difference is not as striking as that of Washington, DC.

4 Numerical results

In this section, we present simulation results in six figures that illustrate outcomes under three vaccination strategies labeled "Actual", "Gradient", and "Uniform," in which the same numbers of doses were administered. To compare them, we consider both absolute differences and percent efficacies relative to results without vaccination. We calculate percent efficacies by subtracting the cumulative outcomes for each scenario

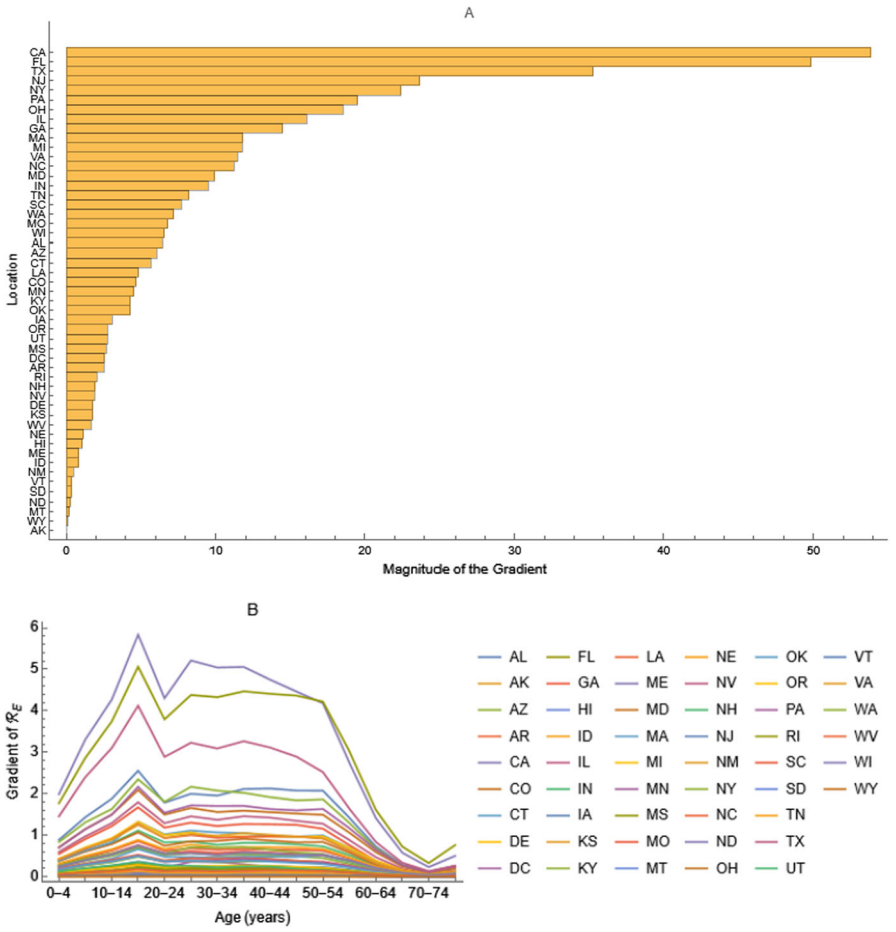

Fig. 7 Components of the gradient of the metapopulation effective reproduction number, $\nabla \mathcal{R}_E$, with respect to χ_{1i}, where $i = 1, \ldots, 816$; **a**) magnitude and **b**) direction. The magnitude is greatest in California, Florida, and Texas and least in Alaska. The directions reflect the marginal contact rates

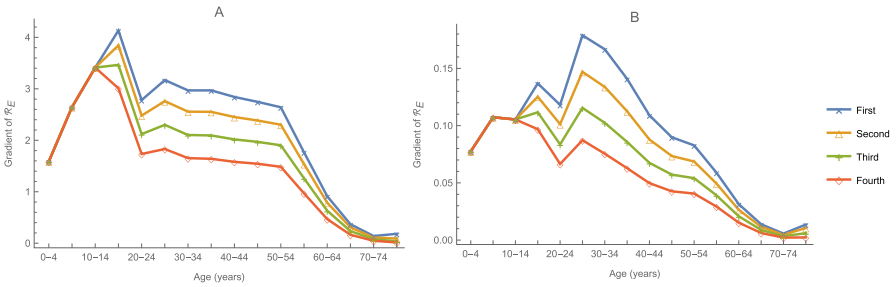

Fig. 8 Gradients for **a**) California and **b**) the District of Columbia, which differ in gradient magnitude and age distribution, respectively, during the first four weeks of January 2021

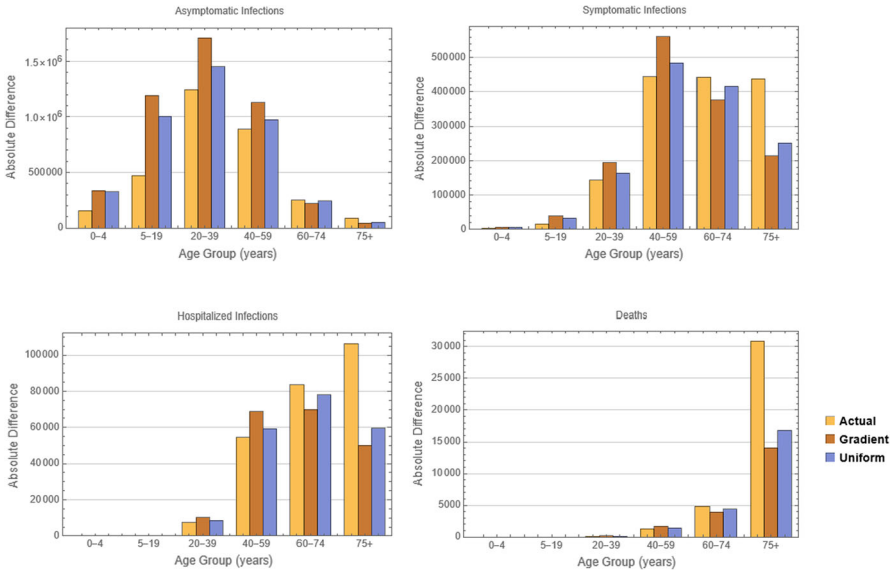

Fig. 9 Absolute differences between cumulative asymptomatic and symptomatic infections, hospitalizations and deaths from 13 December 2020 to 4 April 2021 with the three vaccination strategies and no vaccination

from that without vaccination, dividing by those absent vaccination, and multiplying by one hundred.

Figure 9 illustrates absolute differences between cumulative numbers of asymptomatic and symptomatic infections, hospitalizations and deaths from 13 December 2020 to 4 April 2021 with and without vaccination by scenario and Fig. 10 illustrates percent efficacies. Figures 11, 12, 13, 14 illustrate the prevalence of asymptomatic and symptomatic infections, hospitalizations, and deaths by scenario from 13 December 2020 to 4 April 2021, when incidence was declining, with one panel for each of the six age groups.

5 Discussion

We have developed several metapopulation SEIR models including biological features of COVID-19 that might affect the transmission of SARS-CoV-2. Glasser et al. (2023) describe one in detail, demonstrate its ability to reproduce serological observations in the United States (without having been fit to those observations), and evaluate the initial vaccination program. Our objective here is to demonstrate the utility of analytical quantities derived from that model with and without host population structure (i.e., age groups and spatial locations).

We estimate that less than half of SARS-CoV-2 infections were reported. And under-reporting almost certainly varied with age (Couture et al. 2019) and possibly location. In contrast, serology captures most infections (Petersen et al. 2021). In the calculations and simulations reported here, we use parameter values with which our age- and

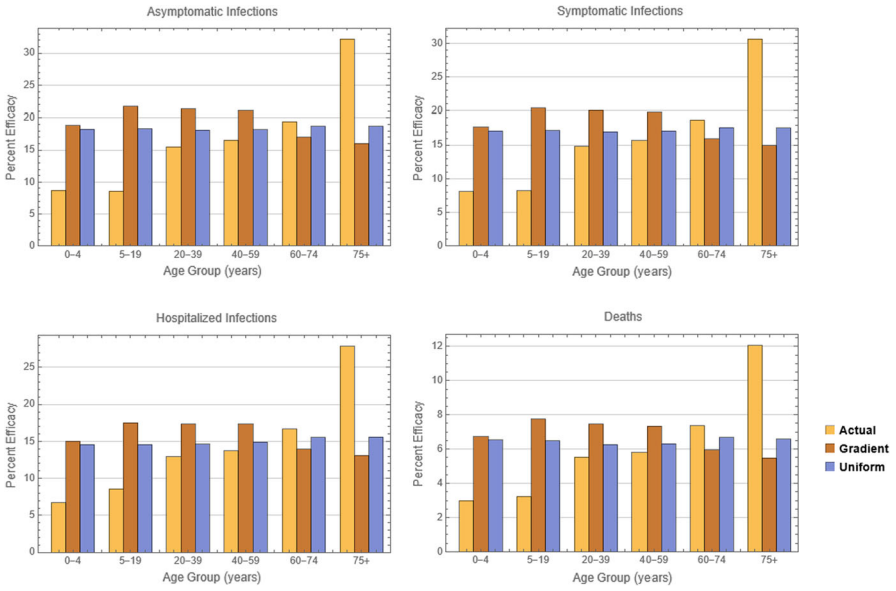

Fig. 10 Efficacy of three vaccination strategies at preventing asymptomatic and symptomatic infections, hospitalizations and deaths from 13 December 2020 to 4 April 2021 calculated as quotients of differences with and without vaccination and those without vaccination times 100

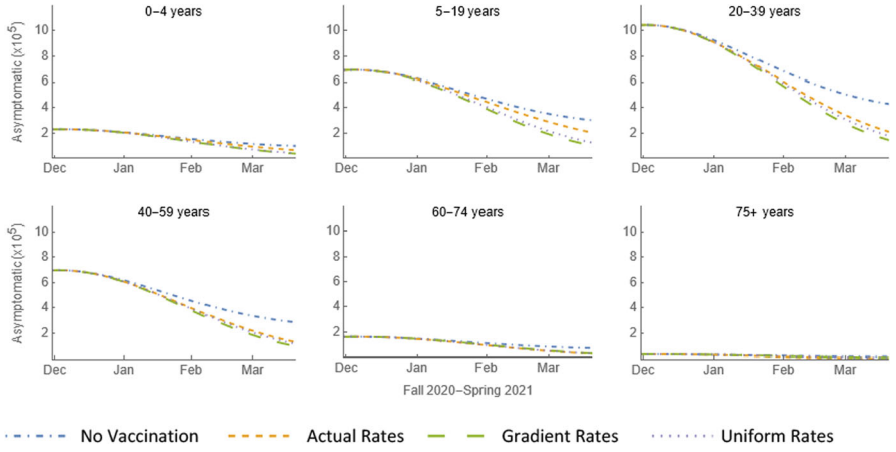

Fig. 11 Prevalence of asymptomatic infections by age in six groups (ages 0–4, 5–19, 20–39, 40–59, 60–74, and 75+ years) from 13 December 2020 through 4 April 2021 without and with three vaccination scenarios

location-stratified model of the transmission of SARS-CoV-2 in the United States either reproduces observations from serial, cross-sectional, nationwide serological surveys or deviates in ways that we believe that we understand (Glasser et al. 2023). Thus, our results are as accurate as is presently possible.

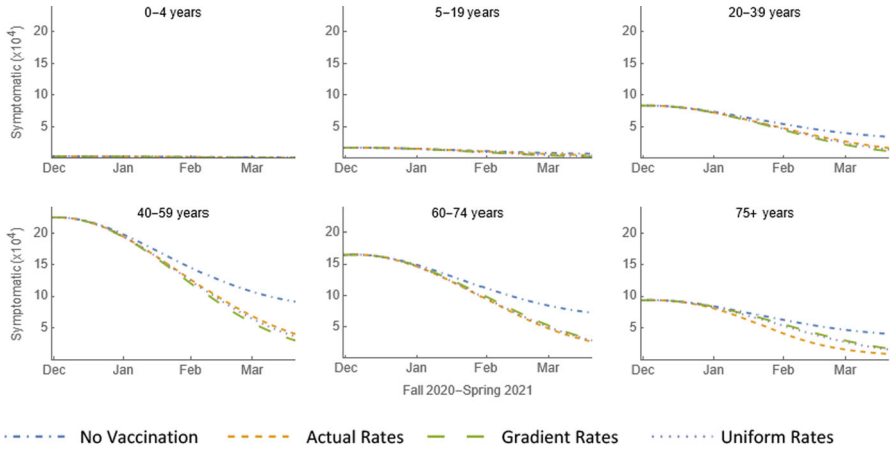

Fig. 12 Prevalence of symptomatic infections by age in six groups (ages 0–4, 5–19, 20–39, 40–59, 60–74, and 75+ years) from 13 December 2020 through 4 April 2021 without and with three vaccination scenarios

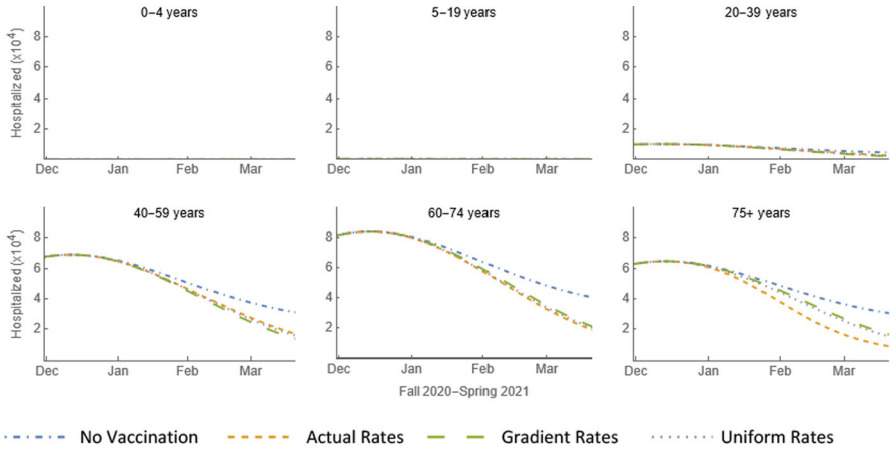

Fig. 13 Prevalence of hospitalizations by age in six groups (ages 0–4, 5–19, 20–39, 40–59, 60–74, and 75 + years) from 13 December 2020 through 4 April 2021 without and with three vaccination scenarios

We use the unstructured model to demonstrate that the population immunity at which $\mathcal{R}_E = 1$, the so-called population-immunity threshold, depends on other mitigation measures. Recalling that the ψ's are reductions in susceptibility and infectiousness by virtue of physical-distancing where possible and mask-wearing otherwise, evidently those NPIs can help to attain $\mathcal{R}_E \leq 1$. On 25 December 2020, we estimate that the population-immunity threshold was 0.66, approximately 0.15 less than the value absent NPIs (i.e., $1 - 1/\mathcal{R}_0$), and vaccination had only just begun. The more effective those measures, the lower the requisite immunity (Fig. 6). This is fortunate insofar as vaccines were not equally accessible throughout the world.

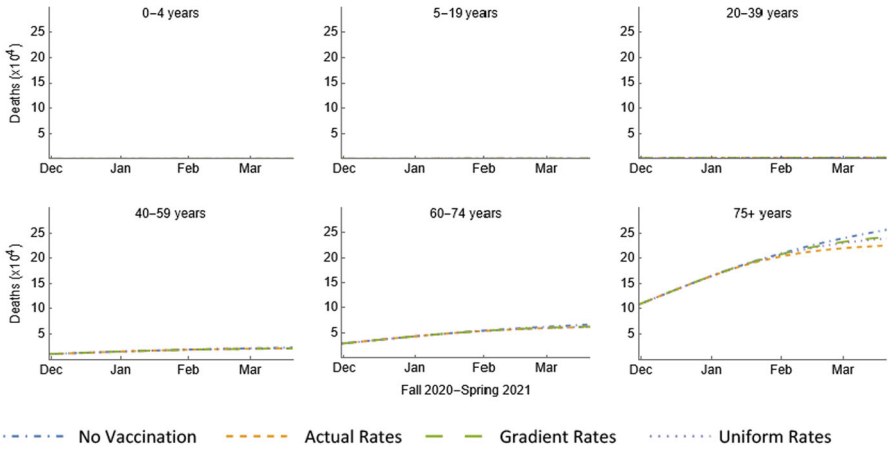

Fig. 14 Cumulative deaths by age in six groups (ages 0–4, 5–19, 20–39, 40–59, 60–74, and 75 + years) from 13 December 2020 through 4 April 2021 without and with three vaccination scenarios

To illustrate the utility of our metapopulation results, we compare subpopulation reproduction numbers (Fig. 3a and c) with their contributions to the metapopulation numbers (Fig. 3b and d). *Per capita*, Washington, DC, has the largest subpopulation $\mathcal{R}_0$ (Fig. 3a), but Florida and New Jersey have larger subpopulation $\mathcal{R}_E$ (Fig. 3c), suggesting that mitigation measures were more effective in Washington on 25 December 2020. As the contributions resemble the marginal contact rates with magnitudes proportional to population density, they illustrate the impact of mixing. Western states generally are larger than eastern ones, so distances between their population centers are greater. Because, in our models, contacts decline with distance faster among older age groups than younger ones, the contributions of older age groups decline east to west (on-line Supplement).

The age-distribution of co-morbidities, together with assumptions about the infectiousness of people in different epidemiological states, determines the *per capita* contributions to subpopulation $\mathcal{R}_{E(i)}$ (Fig. 4). We assume that people with asymptomatic infections are only half as infectious as those with symptomatic ones, who are more infectious the few days before than on average during their symptomatic stages, and that the infectiousness of hospitalized people is limited by infection-control procedures (Glasser et al. 2023, Table 1). These assumptions effectively limit the impact of interventions directed at symptomatic infections (e.g., case investigation and contact tracing). Rather, they emphasize the importance of NPIs and of course vaccination whenever and wherever available.

Then we consider equilibrium prevalence absent and present interventions on 25 December 2020 (Fig. 5). This metric is more evenly distributed, reflecting the age-distributions and infectiousness just described together with mixing, with peaks among adults with young children and possibly grandchildren, but the difference between that associated with the basic and effective reproduction numbers (a and b, respectively) again illustrates the impact of mitigation measures on 25 December 2020. Were such

measures constant at their levels on that date, Florida and New Jersey would have higher equilibrium prevalence than Washington, DC.

Even if vaccination were the only intervention, in metapopulation models, if not the real world, an infinite number of combinations of subpopulation immunity would satisfy the condition that the metapopulation $\mathcal{R}_E = 1$ (Feng et al. 2015). Consequently, the population-immunity threshold is of little practical utility. Moreover, heterogeneity – not only in contact rates, an oft rediscovered result, but any parameter appearing in subpopulation reproduction numbers – increases overall or metapopulation reproduction numbers, as does nonrandom mixing. The conclusion of Britton et al. (2020), that heterogeneity facilitates attaining population immunity thresholds, is not consistent with those results.

Policymakers need to know which combination (i.e., not only of vaccination rates, but compliance with physical-distancing, mask-wearing, and so on) is optimal. To identify that combination, Feng et al. (2015) suggested, and have since employed (Feng et al. 2017, 2020; Hao et al. 2019; Su et al. 2021), the gradient of the effective reproduction number with respect to mitigation measures.

Gradients are vector-valued functions. Their magnitude reflects both the size and density of subpopulations (Fig. 7a) and their direction the marginal contact rates (Fig. 7b). Figure 8 again reflects the marginal contact rates and indicates the need to vaccinate children to reduce the metapopulation $\mathcal{R}_E$. While only two jurisdictions are illustrated, the patterns are similar elsewhere: negative gradient magnitudes among older age groups – whose members were being vaccinated – declined, while those among children – whose vaccination would not begin until fall – remained greater. One cannot help but wonder what would have happened had children been vaccinated earlier and schools remained open.

Figures 9 and 10 indicate that, while the gradient vaccination rates would have reduced infections the most, particularly among those young and middle-aged, the actual rates more effectively reduced hospitalizations and deaths among older people, at least during the first few months of 2021. Thus, whether the gradient rates are optimal or not depends on the policy goal. It may also depend on the situation. If immunity is less protective or wanes more quickly among elderly people (Wang et al. 2021), the gradient rates might well have become the optimal means of protecting them during later pandemic phases.

Figures 11, 12, 13, 14 illustrate the dynamics of these effects during the first few months of 2021, when the pandemic was waning, despite ascendance of the alpha variant (Lambrou et al. 2022), due presumably to infection- and vaccination-induced immunity.

5.1 Limitations

We began modeling the transmission of SARS-CoV-2 long before effective vaccines were developed or approved for use, but – to keep pace with vaccination policy – added one, then two doses, and finally boosting of infection- or vaccine-induced immunity, which however had not yet begun. Glasser et al. (2023) evaluate the actual vaccination strategy. Here, to compare actual and alternative strategies fairly, we assume

100% efficacy (i.e., hypothetical vaccination and observed immunization rates are equivalent).

We reviewed the literature and consulted subject matter experts, but absent information (e.g., compliance with NPIs among those aged < 18 years, infectiousness of people with asymptomatic infections, …) or present conflicting information (e.g., about the efficacy of NPIs), made assumptions (Glasser et al. 2023, Tables 1 and 2).

In the model described by Glasser et al. (2023) and used here, vaccination- and infection-induced immunity are equivalent. We have since learned that the duration and possibly other characteristics of immunity may depend on whether individuals were infected, vaccinated or both, and possibly even the sequence of those events (Goldberg et al. 2022). These developments and the evolution of variants, several successively dominant, motivate us to develop different models for later pandemic periods.

6 Summary

Here we endeavor to illustrate the public health utility of analytical results derived from an age- and location-stratified SEIR model of the transmission of SARS-CoV-2 in the United States that includes features of COVID-19 that might affect transmission. These include the subpopulation and metapopulation $\mathcal{R}_0$ and $\mathcal{R}_E$, subpopulation contributions to the metapopulation $\mathcal{R}_0$ and $\mathcal{R}_E$, subpopulation $\mathcal{R}_{E(i)}$ by infectious state, and $\nabla\mathcal{R}_E$ with respect to all possible vaccination rates before and after vaccination began. To evaluate these expressions, we use parameter values with which this model reproduces seroprevalence reasonably well during the fall of 2020 and spring of 2021 (Glasser et al. 2023). We also compare results of simulated transmission during that period with the actual vaccination strategy and two alternatives.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00285-023-01948-y.

Acknowledgements Contributing to a special issue in memory of Fred Brauer – a pioneering mathematical biologist, colleague, and mentor – is an honor. By modeling the transmission of pathogens among human hosts, we too have obtained epidemiological insights, shared them with others, and influenced public policy. We are grateful to David Bell, Aaron Curns, Betsy Gunnels, Cindy Weinbaum, and several anonymous others for constructive reviews of earlier drafts of this manuscript. We use Mathematica (Wolfram Research, Champaign, IL); the current version is 13.3.

Funding The seroprevalence studies were supported by the Centers for Disease Control and Prevention [via multiple contracts] and our modeling in part by the National Science Foundation [via DMS-1814545 to ZF].

Declarations

Conflict of interest The authors declare that they have no competing interests that are relevant to the content of this article.

Disclaimer The findings and conclusions in this report are those of the authors and do not necessarily represent the official positions of the Centers for Disease Control and Prevention, National Science Foundation, or Purdue University.

Appendix

Reproduction numbers for the metapopulation models

Let **v** denote the vector of infected state variables in the following order:

$$v = \begin{pmatrix} E_{(1,1)}, I_{a(1,1)}, I_{p(1,1)}, I_{s(1,1)}, I_{h(1,1)}, E_{(1,2)}, I_{a(1,2)}, I_{p(1,2)}, I_{s(1,2)}, I_{h(1,2)}, \ldots, E_{(1,n)}, I_{a(1,n)}, I_{p(1,n)}, I_{s(1,n)}, I_{h(1,n)}, \\ E_{(2,1)}, I_{a(2,1)}, I_{p(2,1)}, I_{s(2,1)}, I_{h(2,1)}, E_{(2,2)}, I_{a(2,2)}, I_{p(2,2)}, I_{s(2,2)}, I_{h(2,2)}, \ldots, E_{(2,n)}, I_{a(2,n)}, I_{p(2,n)}, I_{s(2,n)}, I_{h(2,n)}, \ldots, \\ E_{(m,1)}, I_{a(m,1)}, I_{p(m,1)}, I_{s(m,1)}, I_{h(m,1)}, E_{(m,2)}, I_{a(m,2)}, I_{p(m,2)}, I_{s(m,2)}, I_{h(m,2)}, \ldots, E_{(m,n)}, I_{a(m,n)}, I_{p(m,n)}, I_{s(m,n)}, I_{h(m,n)} \end{pmatrix}$$

and consider the system composed of fractions of these variables within each group; i.e.,

$$\begin{aligned} s_{(i,j)} &= \frac{S_{(i,j)}}{N_{(i,j)}}, \quad b_{(i,j)} = \frac{B_{(i,j)}}{N_{(i,j)}}, \quad v_{(i,j)} = \frac{V_{(i,j)}}{N_{(i,j)}}, \quad r_{(i,j)} = \frac{R_{(i,j)}}{N_{(i,j)}}, \\ m_{(i,j)} &= \frac{M_{(i,j)}}{N_{(i,j)}}, \quad x_{(i,j)} = \frac{E_{(i,j)}}{N_{(i,j)}}, \quad y_{a(i,j)} = \frac{I_{a(i,j)}}{N_i}, \quad y_{p(i,j)} = \frac{I_{p(i,j)}}{N_{(i,j)}}, \\ y_{s(i,j)} &= \frac{I_{s(i,j)}}{N_{(i,j)}}, \quad y_{h(i,j)} = \frac{I_{h(i,j)}}{N_{(i,j)}}, \quad i = 1, 2, \cdots, m, \quad j = 1, 2, \cdots, n. \end{aligned}$$

At the Disease-Free Equilibrium (DFE), we have the following from the system of equations in the main text:

$$\begin{aligned} 0 &= q_j \omega b^*_{(i,j)} + \omega_v v^*_{(i,j)} - \chi_{1(i,j)} s^*_{(i,j)}, \\ 0 &= \omega r^*_{(i,j)} - \left[q_j \omega + (1 - q_j)\varepsilon_3 \chi_{3(i,j)}\right] b^*_{(i,j)}, \\ 0 &= \chi_{1(i,j)} s^*_{(i,j)} - \left[\omega_v + \int_2 \chi_{2(i,j)}\right] v^*_{(i,j)}, \\ 0 &= \varepsilon_2 \chi_{2(i,j)} v^*_{(i,j)} + (1 - q_j)\varepsilon_3 \chi_{3(i,j)} b^*_{(i,j)} - \omega r^*_{(i,j)}, \quad i = 1, 2, \ldots, m, \quad j = 1, 2, \ldots, n. \end{aligned}$$

We can solve for $v^*_{(i,j)}$ in terms of $s^*_{(i,j)}$ using the third equation immediately above,

$$v^*_{(i,j)} = \frac{\chi_{1(i,j)}}{\omega_v + \varepsilon_2 \chi_{2(i,j)}} s^*_{(i,j)}.$$

We can solve for $b^*_{(i,j)}$ in terms of $s^*_{(i,j)}$ using the first DFE equation and that immediately above,

$$b^*_{(i,j)} = \left(\frac{\varepsilon_2 \chi_{1(i,j)} \chi_{2(i,j)}}{q_j \omega(\omega_v + \varepsilon_2 \chi_{2(i,j)})}\right) s^*_{(i,j)}.$$

And we can solve for $r^*_{(i,\,j)}$ in terms of $s^*_{(i,\,j)}$ using last DFE equation and two immediately above,

$$r^*_{(i,\,j)} = \left(\frac{q_j \omega \varepsilon_2 \chi_{2(i,\,j)} \chi_{1(i,\,j)} + \left(1 - q_j\right) \varepsilon_3 \chi_{3(i,\,j)} \varepsilon_2 \chi_{1(i,\,j)} \chi_{2(i,\,j)}}{q_j \omega^2 (\omega_v + \varepsilon_2 \chi_{2(i,\,j)})} \right) s^*_{(i,\,j)}.$$

At the DFE, we have

$$1 = s^*_{(i,\,j)} + v^*_{(i,\,j)} + b^*_{(i,\,j)} + r^*_{(i,\,j)}.$$

So, together with the previous three equations, we can solve for $s^*_{(i,\,j)}$,

$$s^*_{(i,\,j)} = \left(1 + \frac{\chi_{1(i,\,j)}}{\omega_v + \varepsilon_2 \chi_{2(i,\,j)}} + \frac{\varepsilon_2 \chi_{1(i,\,j)} \chi_{2(i,\,j)}}{q_j \omega (\omega_v + \varepsilon_2 \chi_{2(i,\,j)})} + \frac{q_j \omega \varepsilon_2 \chi_{2(i,\,j)} \chi_{1(i,\,j)} + \left(1 - q_j\right) \varepsilon_3 \chi_{3(i,\,j)} \varepsilon_2 \chi_{1(i,\,j)} \chi_{2(i,\,j)}}{q_j \omega^2 (\omega_v + \varepsilon_2 \chi_{2(i,\,j)})} \right)^{-1}.$$

The effective reproduction number for group $(i,\ j)$ is
$\mathcal{R}_{E(i,\,j)} = \Psi_{jS} a_{(i,\,j)} \beta_j c_{l_i a_j l_i a_j} \left[s^*_{(i,\,j)} + (1-\varepsilon_1) v^*_{(i,\,j)} + (1-\varepsilon_3) b^*_{(i,\,j)} \right] \Psi_{jI} Q_j$, where
$Q_j = \eta_a P_{ja} T_{ja} + \eta_p P_{jp} T_{jp} + P_{jp} T_{js} + \eta_h P_{jp} P_{jh} T_{jh}$, with
$P_{ja} = p_j$, $P_{jp} = \left(1 - p_j\right)$, and $P_{jh} = \frac{\delta_j \gamma_{js \to h}}{\gamma_{js} + \delta_j \gamma_{js \to h}}$ denoting probabilities of transition from $E_{(i,\,j)}$ to $I_{a(i,\,j)}$, from $E_{(i,\,j)}$ to $I_{p(i,\,j)}$, and from $I_{s(i,\,j)}$ to $I_{h(i,\,j)}$, respectively, and $T_{ja} = \frac{1}{\gamma_{ja}}$, $T_{jp} = \frac{1}{\xi_j}$, $T_{js} = \frac{1}{\gamma_{js} + \delta_j \gamma_{js \to h}}$, $T_{jh} = \frac{1}{\gamma_{jh}}$ denoting mean residence times in the $I_{a(i,\,j)}$, $I_{p(i,\,j)}$, $I_{s(i,\,j)}$, and $I_{h(i,\,j)}$ states, respectively. The effective reproduction number for group i is
$\mathcal{R}_{E(i)} = \sum_{j=1}^{n} \Psi_{jS} a_{(i,\,j)} \beta_j c_{l_i a_j l_i a_j} \left[s^*_{(i,\,j)} + (1-\varepsilon_1) v^*_{(i,\,j)} + (1-\varepsilon_3) b^*_{(i,\,j)} \right] \Psi_{jI} Q_j$.
For the age- and location-stratified models, we follow the method outlined by van den Driessche and Watmough (2002). The next-generation matrix is an $mn \times mn$ block matrix $K = (K_{rs})$ with

$$K_{rs} = \begin{pmatrix} \Psi_{\pi(r)S} a_{(\phi(r),\ \pi(r))} \beta_{\pi(r)} z^*_{(\phi(r),\ \pi(r))} \Psi_{\pi(s)I} Q_{\pi(s)} c_{l_{\phi(r)} a_{\pi(r)} l_{\phi(s)} a_{\pi(s)}} & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

for $r,\ s = 1,\ 2,\ \ldots,\ mn$, where $z^*_{(i,\,j)} = s^*_{(i,\,j)} + (1-\varepsilon_1) v^*_{(i,\,j)} + (1-\varepsilon_3) b^*_{(i,\,j)}$, $\phi(x) = \lceil x/n \rceil$, and $\rho(x) = (x - 1 \mod n) + 1$; the "*" entries do not affect the result. Because only the first row of the K_{ij} matrix is non-zero, the next-generation matrix K has $4mn$ rows of zeros and mn rows with non-zero elements. Consequently, K has $4mn$ zero eigenvalues and mn non-zero ones, given by the following matrix (whose $(i,\ j)^{\text{th}}$ element is the $(1,\ 1)$ element of matrix K_{ij}):

$$\begin{aligned} \mathbf{H} &= \left(\Psi_{\pi(r)S} a_{(\phi(r),\ \pi(r))} \beta_{\pi(r)} z^*_{(\phi(r),\ \pi(r))} \Psi_{\pi(s)I} Q_{\pi(s)} c_{l_{\phi(r)} a_{\pi(r)} l_{\phi(s)} a_{\pi(s)}} \right) \\ &= \mathbf{Z}^* \boldsymbol{\Psi}_S \mathbf{A} \boldsymbol{\Psi}_I, \end{aligned}$$

where

$$\mathbf{A} = \begin{pmatrix} a_{(1,1)}\beta_1 Q_1 c_{l_1 a_1 l_1 a_1} & a_{(1,1)}\beta_1 Q_2 c_{l_1 a_1 l_1 a_2} & \dots & a_{(1,1)}\beta_1 Q_n c_{l_1 a_1 l_m a_n} \\ a_{(1,2)}\beta_2 Q_1 c_{l_1 a_2 l_1 a_1} & a_{(1,2)}\beta_2 Q_2 c_{l_1 a_2 l_1 a_2} & \dots & a_{(1,2)}\beta_2 Q_n c_{l_1 a_2 l_m a_n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{(1,n)}\beta_n Q_1 c_{l_1 a_n l_1 a_1} & a_{(1,n)}\beta_n Q_2 c_{l_1 a_n l_1 a_2} & \dots & a_{(1,n)}\beta_n Q_n c_{l_1 a_n l_m a_n} \\ a_{(2,1)}\beta_1 Q_1 c_{l_2 a_1 l_1 a_1} & a_{(2,1)}\beta_1 Q_2 c_{l_2 a_1 l_1 a_2} & \dots & a_{(2,1)}\beta_1 Q_n c_{l_2 a_1 l_m a_n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{(m,n)}\beta_n Q_1 c_{l_m a_n l_1 a_1} & a_{(m,n)}\beta_n Q_2 c_{l_m a_n l_1 a_2} & \dots & a_{(m,n)}\beta_n Q_n c_{l_m a_n l_m a_n} \end{pmatrix},$$

$$= \left(a_{(\phi(r),\ \pi(r))}\beta_{\pi(r)} Q_{\pi(s)} c_{l_{\phi(r)} a_{\pi(r)} l_{\phi(s)} a_{\pi(s)}} \right)$$

$$\mathbf{Z}^* = \operatorname{diag}\left\{ z^*_{(1,1)}, \dots, z^*_{(1,n)}, z^*_{(2,1)}, \dots, z^*_{(2,n)}, \dots, z^*_{(m,1)}, \dots, z^*_{(m,n)} \right\},$$

$\mathbf{\Psi}_S = \operatorname{diag}\{\Psi_{1S}, \dots, \Psi_{nS}, \Psi_{1S}, \dots, \Psi_{1S}, \dots, \Psi_{1S}, \dots, \Psi_{nS}\}$, and $\mathbf{\Psi}_I = \operatorname{diag}\{\Psi_{1I}, \dots, \Psi_{nI}, \Psi_{1I}, \dots, \Psi_{nI}, \dots, \Psi_{1I}, \dots, \Psi_{nI}\}$.

Letting $\rho(\mathbf{H})$ denote the dominant eigenvalue of $\mathbf{H}$, the effective reproduction number for the metapopulation is given by $\mathcal{R}_E = \rho(\mathbf{H})$. Note that matrix $\mathbf{A}$ does not depend on control parameters, and that matrices $\mathbf{Z}^*$, $\mathbf{\Psi}_S$ and $\mathbf{\Psi}_I$ represent the effects of immunity, reduction in susceptibility, and reduction in infectivity, respectively. Thus, the basic reproduction number is given by $\mathcal{R}_0 = \rho(\mathbf{A})$.

Matrix $\mathbf{A}$ can also be written as a product of matrices representing transmission, mixing, and weighted sojourns; i.e., $\mathbf{A} = \mathbf{BCQ}$, where

$\mathbf{C} = \left(c_{l_{\phi(r)} a_{\pi(r)} l_{\phi(s)} a_{\pi(s)}}\right)$, $r, s = 1, 2, \dots, mn$ is the mixing matrix,

$\mathbf{B} = \operatorname{diag}\left\{a_{(1,1)}\beta_1, \dots, a_{(1,n)}\beta_n, \dots, a_{(m,1)}\beta_1, \dots, a_{(m,n)}\beta_n\right\}$ consists of the transmission rates, and $\mathbf{Q} = \operatorname{diag}\{Q_1, \dots, Q_n, \dots, Q_1, \dots, Q_n\}$ of the weighted sojourns.

Matrix $\mathbf{Q}$ can be further decomposed by writing $Q_j = \mathbf{P}_j \mathbf{T}_j$, where $\boldsymbol{\eta} = \left(\eta_a, \eta_p, 1, \eta_h\right)$, the relative infectiousness of the several infectious states, $\mathbf{P}_j = \operatorname{diag}\left\{P_{ja}, P_{jp}, P_{js}, P_{jh}\right\}$, the transition probabilities, and $\mathbf{T}_j = \left(T_{aj}, T_{jp}, T_{js}, T_{jh}\right)$, the mean sojourns.

If we let $\mathcal{R}_{A(i,j)}$, $\mathcal{R}_{P(i,j)}$, $\mathcal{R}_{S(i,j)}$ and $\mathcal{R}_{H(i,j)}$ denote the number of secondary infections per infected individual in the $I_{a(i,j)}$, $I_{p(i,j)}$, $I_{s(i,j)}$ and $I_{h(i,j)}$ stages, respectively, $\mathcal{R}_{E(i,j)} = \mathcal{R}_{A(i,j)} + \mathcal{R}_{P(i,j)} + \mathcal{R}_{S(i,j)} + \mathcal{R}_{H(i,j)}$, where

$$\begin{aligned} \mathcal{R}_{A(i,j)} &= \Psi_{jS} a_{(i,j)} \beta_j z^*_{(i,j)} \Psi_{jI} \eta_{ja} P_{ja} T_{ja} \\ \mathcal{R}_{P(i,j)} &= \Psi_{jS} a_{(i,j)} \beta_j z^*_{(i,j)} \Psi_{jI} \eta_{jp} P_{jp} T_{jp} \\ \mathcal{R}_{S(i,j)} &= \Psi_{jS} a_{(i,j)} \beta_j z^*_{(i,j)} \Psi_{jI} P_{jp} T_{js} \\ \mathcal{R}_{H(i,j)} &= \Psi_{jS} a_{(i,j)} \beta_j z^*_{(i,j)} \Psi_{jI} \eta_{jh} P_{jp} P_{jh} T_{jh}, \quad i = 1, 2, \dots, m, \quad j = 1, 2, \dots, n, \end{aligned}$$

are the contributions of the infectious stages to the metapopulation reproduction number.

Finally, the reproduction number expression under proportionate mixing is given by the trace of the next-generation matrix. When mixing is proportional, $c_{l_i a_j l_p a_q} = a_{(i,j)} N_{(i,j)} / \sum_{r=1}^{m} \sum_{s=1}^{n} a_{(r,s)} N_{(r,s)}$. So, $\mathcal{R}_E = \sum_{i=1}^{m} \sum_{j=1}^{n} \Psi_{jS} a_{(i,j)} \beta_j z^*_{(i,j)} \Psi_{jI} Q_i c_{l_i a_j l_i a_j}$ when mixing is proportional.

References

Bajema KL, Wiegand RE, Cuffe K et al (2021) Estimated SARS-CoV-2 seroprevalence in the US as of September 2020. JAMA Intern Med 181:450–460

Britton T, Ball F, Trapman P (2020) A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. Science 369:846–849

Busch MP, Stramer SL, Stone M et al (2022) Population-weighted seroprevalence from SARS-CoV-2 infection, vaccination, and hybrid immunity among US blood donations from January-December 2021. Clin Inf Dis. 75:S254–S263. https://doi.org/10.1093/cid/ciac470

Caswell H (2001) Matrix population models: construction, analysis and interpretation, 2nd edn. Sinauer Associates, Sunderland, MA

Clark A, Jit M, Warren-Gash C et al (2020) Global, regional, and national estimates of the population at increased risk of severe COVID-19 due to underlying health conditions in 2020: a modelling study. Lancet Glob Health 8:E1003–E1017

Couture A, Lyons BC, Mehrotra ML et al (2022) Severe acute respiratory syndrome coronavirus 2 seroprevalence and reported Coronavirus Disease 2019 cases in US children. Open Forum Infect Dis 9:ofac5044. https://doi.org/10.1093/ofid/ofac044

Feng Z, Glasser JW (2019) Mixing in metapopulation models. In: Bianchi A, Hillen T, Lewis M, Yi Y (eds) The dynamics of biological systems, Chapter 3. Springer, Berlin

Feng Z, Hill AN, Smith PJ, Glasser JW (2015) An elaboration of theory about preventing outbreaks in homogeneous populations to include heterogeneity or preferential mixing. J Theor Biol 386:177–187

Feng Z, Hill AN, Curns AT, Glasser JW (2017) Evaluating targeted interventions via metapopulation models with multi-level mixing. Math Biosci 287:93–104

Feng Z, Feng Y, Glasser JW (2020) Influence of demographically realistic mortality schedules on vaccination strategies in age-structured models. Theor Pop Biol 132:24–32

Glasser JW, Feng Z, Vo MV, Jones JN, Clarke KEN (2023) Analysis of serological surveys of antibodies to SARS-CoV-2 in the United States to estimate parameters needed for transmission modeling and to evaluate and improve the accuracy of predictions. J Theor Biol 556:111296

Goldberg Y, Mandel M, Bar-On YM et al (2022) Protection and waning of natural and hybrid immunity to SARS-CoV-2. N Engl J Med 386:2201–2212. https://doi.org/10.1056/NEJMoa2118946

Hale T, Angrist N, Goldszmidt R et al (2021) A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). Nat Hum Behav 5:529–538

Hao L, Glasser JW, Su Q et al (2019) Evaluating vaccination policies to accelerate measles elimination in China: a metapopulation modeling study. Int J Epidemiol 48:1240–1251

He X, Lau EHY, Wu P et al (2020) Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med 26:672–675

Hethcote HW, van Ark JW (1987) Epidemiological models for heterogeneous populations: proportionate mixing, parameter estimation and immunization programs. Math Biosci 84:85–118

Jones SP (2020) Covid 19 Behaviour Tracker. Imperial College London YouGov Covid Data Hub, v1.0, YouGov Plc, April 2020 (https://github.com/YouGov-Data/covid-19-tracker/)

Jones JM, Stone M, Sulaeman H et al (2021) Estimated US infection- and vaccine-induced SARS-CoV-2 seroprevalence based on blood donations, July 2020-May 2021. JAMA 326:1400–1409

Lambrou AS, Shirk P, Steele MK, et al. (2022) Genomic surveillance for SARS-CoV-2 variants: predominance of the Delta (B.1.617.2) and Omicron (B.1.1.529) variants — United States, June 2021–January 2022. MMWR Morb Mortal Wkly Rep 2022, 71:206–211. https://doi.org/10.15585/mmwr.mm7106a4

Levin AT, Hanage WP, Owusu-Boaitey N, Cochran KB, Walsh SP, Meyerowitz-Katz G (2020) Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications. Eur J Epidemiol 35:1123–1138

Petersen LR, Sami S, Vuong N et al (2021) Lack of antibodies to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in a large cohort of previously infected persons. Clin Inf Dis 73:e3066–e3073. https://doi.org/10.1093/cid/ciaa1685

Prem K, Cook AR, Jit M (2017) Projecting social contact matrices in 152 countries using contact surveys and demographic data. PLoS Comput Biol 13:e1005697

Renshaw E (1991) Modeling biological populations in space and time. Cambridge Univ Press, Cambridge

Su Q, Feng Z, Hao L et al (2021) Assessing the burden of congenital rubella syndrome in China and evaluating mitigation strategies: a metapopulation modeling study. Lancet Inf Dis 21:1004–1013

Thompson RN, Stockwind JE, van Gaalene RD et al (2019) Improved inference of time-varying reproduction numbers during infectious disease outbreaks. Epidemics 29:100356

van den Driessche P, Watmough J (2002) Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math Biosci 180:29–48

Wang J, Tong Y, Li D, Li J, Li Y (2021) The impact of age difference on the efficacy and safety of COVID-19 vaccines: a systematic review and meta-analysis. Front Immunol 12:758294